Consequences of General Relativity: The Schwarzschild and Robertson-Walker Spacetimes

Owen Bechtel
Advisor: Floyd Williams

Abstract

Certain properties of gravity in general relativity — that it can bend the trajectories of light rays, cause time to slow down, and in extreme cases form strange objects known as "black holes" — are widely known. Also widely known is the fact that the universe is expanding, and that a force or substance known as "dark energy" may be causing its expansion to accelerate. This work is an exposition of the mathematics behind such statements, focusing on two relativistic spacetimes: the Schwarzschild and FLRW (Friedmann-Lemaître-Robinson-Walker) universes.

Contents

1 Background

1.1 History of Relativity

Before the 20th century, physicists believed that light was a wave traveling through a medium known as the luminiferous ether. Light was said to always move at the same speed relative to this ether. However, a number of experiments during the 19th century proved to be problematic for this theory. For example, an 1887 experiment by American physicists Albert Michelson and Edward Morley found that light always travels at the same speed relative to the ground, regardless of its direction of motion and regardless of the time of year. This would imply that the Earth is always stationary in the ether, despite revolving around the sun. On its own, this result could be explained by the complete ether drag hypothesis, which states that the ether is "dragged" by nearby matter, such as the Earth. But the 1851 Fizeau experiment, which measured the speed of light through moving water, had already disproven complete ether drag.

After the Michelson-Morley experiment, no contemporary theory could explain the speed of light. Hendrik Lorentz began to fix this problem in the 1890s, developing what is now known as Lorentz ether theory. This theory was a precursor to Einstein's special relativity, the primary differences being that Lorentz held on to the notion of the ether, which constituted a single preferred reference frame, and made assumptions (such as length contraction) that are now known to be superfluous. Finally, in his 1905 paper *On the Electrodynamics of Moving Bodies*, Einstein showed how to derive the Lorentz transformation from the single principle that light always moves at the same speed c in the reference frame of any inertial (constant-velocity) observer. He also modified the laws of mechanics to make them invariant under Lorentz transformations, resolving certain asymmetries in classical electromagentism such as the moving magnet and conductor problem. In Einstein's theory, it could no longer be said that electromagentism possesses any notion of absolute rest, making the ether obsolete. The theory developed by Lorentz, Einstein, Poincare, and Minkowski is now known as "special relativity," in contrast to the more general theory of general relativity, which incorporates gravity.

General relativity was developed by Einstein over the course of eight years, ending with the publication of what is now known as the Einstein field equation in 1915. The utility of differential geometry — the mathematical basis of general relativity — was pointed out to Einstein by mathematician Marcel Grossmann. The fundamental realization behind general relativity is that gravity is exclusively a large-scale phenomenon; at (spatially and temporally) small scales, it cannot be detected. For example, the experience of someone standing on the Earth is the same as that of some of someone standing on an upwardly-accelerating world in a universe without gravity. The only difference is that, in our universe, objects whose velocities are intially equal can grow closer together or farther apart over time. But this effect is only detectable on large scales — objects dropped on opposite sides of the Earth obviously grow closer over time, but this is not apparent for objects dropped a few meters apart. We can compare gravity to the curvature of a sphere: up close, the sphere looks like a flat plane, but on larger scales, its curvature is apparent in the fact that objects tracing "straight lines" (geodesics) can grow closer together over time, even if they start out parallel. General relativity takes this comparison to its ultimate conclusion and states that *gravity is curvature* (of 4-dimensional spacetime rather than a 2-dimensional surface).

Since its inception, general relativity has been very successful. Among other things, it accounted for a discrepancy between the observed perihelion precession of Mercury and that predicted by Newtonian gravity, and correctly predicted phenomena such as gravitational redshift, bending of light, and gravitational waves.

1.2 Mathematics of Relativity

The spacetime of special relativity is a 4-dimensional vector space (known as Minkowski space) equipped with a bilinear form η of signature $(-,+,+,+)$.[1] This means that there exist bases (a_0, a_1, a_2, a_3) such that $\eta(a_0) = -1$, $\eta(a_1) = \eta(a_2) = \eta(a_3) = 1$, and $\eta(a_i, a_j) = 0$ for all $i \neq j$. Such a basis is known in the context of special relativity as an inertial reference frame.

An inertial reference frame $A = (a_0, a_1, a_2, a_3)$ measures a point (or "event") $x = x_0a_0 + x_1a_1 + x_2a_2 + x_3a_3$ to occur at time x_0/c and position (x_1, x_2, x_3).[2] Events which are multiples of a_0 have position $(0, 0, 0)$, so this reference frame corresponds to an observer moving in the direction of a_0.

Depending on the value of η, a spacetime vector v is called *timelike* ($\eta(v) < 0$), *lightlike* ($\eta(v) = 0$), or *spacelike* ($\eta(v) > 0$). Curves can be classified likewise; for example, a lightlike curve is a curve all of whose tangent vectors are lightlike. Massless objects such as light travel along lightlike curves (hence the name), and are measured to have a speed of c in any inertial reference frame, while massive objects travel along timelike curves.

[1] More accurately, spacetime is the affine space modeled on this vector space, since it has no inherent "origin."

[2] The conversion factor c allows one to use different units for time and distance, but is unimportant from a purely mathematical perspective. It is not uncommon to use $c = 1$, which amounts to measuring time in (say) seconds and distance in light-seconds.

A linear map that converts one inertial reference frame into another is called a Lorentz transformation. From the Lorenz transformation, one can derive a number of (often counterintuitive) phenomena, such as length contraction, time dilation, and the relativity of order of events. I could go into more detail, but this paper is primarily concerned with the consequences of general relativity, not special relativity.

General relativity replaces the flat Minkowski spacetime of special relativity with an arbitrary Lorentzian manifold; that is, a 4-dimensional manifold M equipped with a pseudo-Riemannian metric g of signature $(-,+,+,+)$. The tangent spaces of such a manifold are Minkowski spaces, so general relativity is indistinguishable from special relativity at sufficiently small scales.

As in special relativity, tangent vectors and some curves can be classified as timelike, lightlike, or spacelike, with slower-than-c particles traveling along timelike curves, and free-falling particles traveling along geodesics. The amount of time experienced by a particle traveling along timelike curve C is

$$\frac{1}{c}\int_a^b \sqrt{-g(\gamma'(\lambda))}\,\mathrm{d}\lambda,$$

where $\gamma : [a,b] \to M$ is a parametrization of C (the parametrization used does not matter). Similarly, the proper length of a spacelike curve is

$$\int_a^b \sqrt{g(\gamma'(\lambda))}\,\mathrm{d}\lambda,$$

the only difference between the two formulas being the unit conversion factor c, and a change in sign to keep the number inside the square root positive.

Given a particle traveling along a timelike curve with parametrization γ, its *4-velocity* is defined as

$$\frac{c\gamma'(\lambda)}{\sqrt{-g(\gamma'(\lambda))}},$$

its *4-momentum* is its 4-velocity times its mass, and its *4-acceleration* is the covariant derivative of 4-velocity (viewed as a vector field along a curve) with respect to itself. The 4-acceleration of an object in free fall is 0, by the definition of a geodesic.

The metric is constrained by the distribution of matter in the universe by the equation

$$\mathrm{Ric} - \frac{1}{2}Rg + \Lambda g = \frac{8\pi G}{c^4}T$$

where Ric is the Ricci curvature tensor, R is the Ricci curvature scalar, Λ is a constant known as the cosmological constant, G is Newton's gravitational constant, and T is the stress-energy tensor. In a vacuum with zero cosmological constant ($T = 0$, $\Lambda = 0$), the equation above becomes equivalent to $\mathrm{Ric} = 0$.[3] That is, the vacuum solutions of general relativity are those that are Ricci flat.

2 The Schwarzschild Spacetime

In 1916, one year after Einstein published his field equation, Karl Schwarzschild, a physicist and officer in the German army, found a solution to Einstein's equation that models spacetime around a non-rotating spherical mass. This solution, known as the Schwarzschild metric, predicts a number of interesting phenomena, including gravitational time dilation, gravitational lensing (bending of light), and the potential existence of black holes.

2.1 Derivation

The Schwarzschild metric is most commonly given in spherical coordinates $(x_0, x_1, x_2, x_3) = (ct, r, \theta, \phi)$. The most general formula for a metric in such a coordinate system is

$$g = \sum_{i,j=0}^{3} g_{ij}\, \mathrm{d}x_i \otimes \mathrm{d}x_j,$$

where the g_{ij} are functions such that $g_{ij} = g_{ji}$ for all i and j.

We can place more restrictions on the components of g by making some assumptions. First, we assume that g does not change over time, which means that the g_{ij} do not depend on t, but only on r, θ, and ϕ. We also assume that g is invariant under time reversal; mathematically, this means that replacing $\mathrm{d}x_0$ with $-\mathrm{d}x_0$ does not affect g. Negating $\mathrm{d}x_0$ is equivalent to negating the g_{0i} components for $i \neq 0$,[4] so these components must be 0. Physically, invariance under time reversal corresponds to the mass at the center not rotating.

g should also be invariant under spatial reflection. Reflection across the horizontal plane (sending θ to $\pi - \theta$) causes $\mathrm{d}x_2 = \mathrm{d}\theta$ to be negated, but leaves the other $\mathrm{d}x_i$s fixed. Similarly, reflection across a vertical plane (sending ϕ to $-\phi$) causes $\mathrm{d}x_3 = \mathrm{d}\phi$ to be negated. Thus g_{2i} for $i \neq 2$ and g_{3i} for $i \neq 3$ are all 0. At this point, we know that all of the off-diagonal components of g are 0.

Finally, g should be invariant under spatial rotation. This implies two things: that g_{00} and g_{11} do not depend on θ or ϕ, and that g, when restricted to a sphere of constant t and r, is simply the standard metric on the sphere, possibly scaled by some number that may depend on r. The standard metric on a sphere of radius r, written in spherical coordinates, is

$$r^2\, \mathrm{d}\theta^2 + r^2 (\sin\theta)^2\, \mathrm{d}\phi^2.$$

[3] To prove this, we take the "trace" (contraction with the inverse metric) on both sides of $\mathrm{Ric} = \frac{1}{2} R g$, getting $R = 2R$ and thus $R = 0$.

[4] g_{00} is not affected, since the two negatives cancel.

So g is of the form

$$-A(r)\,\mathrm{d}x_0^2 + B(r)\,\mathrm{d}r^2 + C(r)r^2\,\mathrm{d}\theta^2 + C(r)r^2(\sin\theta)^2\,\mathrm{d}\phi^2.$$

We can get rid of C by replacing r with a new coordinate defined as $\sqrt{C(r)}\cdot r$. Then g becomes

$$-A(r)\,\mathrm{d}x_0^2 + B(r)\,\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2 + r^2(\sin\theta)^2\,\mathrm{d}\phi^2,$$

where A and B have now been reparametrized.

All that remains is to solve for the functions A and B. We do this by imposing the Einstein vacuum equation on g, which is equivalent to the Ricci tensor of g being 0. Three components of the Ricci tensor are

$$\begin{aligned} R_{00} &= 2rABA'' - rAA'B' + 4ABA' - rB(A')^2 \\ R_{11} &= -2rABA'' + rAA'B' + 4A^2B' + rB(A')^2 \\ R_{22} &= -2AB + 2AB^2 - rA'B + rAB'. \end{aligned}$$

The equation $R_{00} + R_{11} = 0$ simplifies to $A'B + AB' = 0$, which implies that $AB = k$ for some constant k. If we assume that g approaches the flat Minkowski metric as $r \to \infty$, then A and B both approach 1 as $r \to \infty$, which implies that $k = 1$. Therefore, $B = 1/A$, so the equation $R_{22} = 0$ becomes

$$-2 + 2/A - 2rA'/A = 0.$$

Multiplying by $A/2$ on both sides and solving for A', this becomes

$$A' = \frac{1-A}{r},$$

a differential equation whose solutions are of the form

$$A(r) = 1 - \frac{r_\mathrm{s}}{r}$$

for some constant r_s.

So the Schwarzschild metric is

$$g = -\left(1 - \frac{r_\mathrm{s}}{r}\right)\mathrm{d}x_0^2 + \left(1 - \frac{r_\mathrm{s}}{r}\right)^{-1}\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2 + r^2(\sin\theta)^2\,\mathrm{d}\phi^2.$$

In summary, the Schwarzschild metric g is the only metric on (3+1)-dimensional spacetime that is spherically symmetric, invariant under time translation and reversal, approaches the Minkowski metric as $r \to \infty$, and satisfies the Einstein vacuum equation.

Note that g is undefined if $r = 0$ or if $r = r_\mathrm{s}$, as both of these cases result in a division by 0. So the Schwarzschild metric is defined on two disconnected regions: the region where $r \in (0, r_\mathrm{s})$ and the region where $r \in (r_\mathrm{s}, \infty)$. In the outer region, t is timelike and r is spacelike (as one would expect), but in the inner region, t is spacelike and r is timelike. At the moment, these two regions are separate Lorentzian manifolds, but as we will see later, they can both be embedded into a larger Lorentzian manifold, a fact that becomes clear after a change of coordinates.

The quantity r_s is known as the Schwarzschild radius.

2.2 Physical Interpretation

If we imagine a universe containing only a single mass, which is spherically symmetric and non-rotating, then the vacuum around the mass should have all of the properties that imply the Schwarzschild metric. Therefore, if general relativity is correct, this vacuum is described by the Schwarzschild metric. We should expect objects traveling along geodesics in the Schwarzschild spacetime to be drawn toward the center, and to behave the same as in Newtonian physics in the limit of low velocity and low gravity, for if not, general relativity would not be a viable theory of gravity.

The 4-velocity of an observer with constant r, θ, and ϕ (so stationary relative to the mass at the center) is

$$\frac{c}{\sqrt{-g_{00}}}e_0 = \frac{c}{\sqrt{1-r_\mathrm{s}/r}}e_0.$$

The proper acceleration is the covariant derivative of the 4-velocity with respect to itself, that is

$$\begin{aligned}\frac{c}{\sqrt{1-r_\mathrm{s}/r}}\nabla_{e_0}\left(\frac{c}{\sqrt{1-r_\mathrm{s}/r}}e_0\right) &= \frac{c^2}{1-r_\mathrm{s}/r}\sum_{i=0}^{3}\Gamma^i_{00}e_i \\ &= \frac{c^2 r}{r-r_\mathrm{s}}\frac{1}{2}\frac{r_\mathrm{s}(r-r_\mathrm{s})}{r^3}e_r = \frac{c^2 r_\mathrm{s}}{2r^2}e_r.\end{aligned}$$

So, to remain stationary, an object must have a constant outward proper acceleration. It follows that if it were following a geodesic, it would fall inward. If we ignore the fact that the magnitude of e_r (i.e. $\sqrt{g_{11}}$) is dependent on r (a fact that is only of great significance if r is close to r_s), we see that the acceleration an object must have to remain stationary is proportional to r^{-2}, just as in Newtonian physics. Specifically, the acceleration predicted by Newton is GM/r^2, where G is the gravitational constant and M is the mass of the central object. So, if general relativity becomes Newtonian gravity for a stationary object in the limit as $r \to \infty$ (and thus $|e_r| \to 1$), we must have

$$\frac{c^2 r_\mathrm{s}}{2r^2} = \frac{GM}{r^2}, \quad \text{i.e.} \quad r_\mathrm{s} = \frac{2GM}{c^2}.$$

That is, the Schwarzschild radius is proportional to the mass of the central object. The proportionality constant $2G/c^2$ is about 1.5×10^{-27} when measured in SI units of meters per kilogram, and about 2.95 when measured in kilometers per solar mass. So the Schwarzschild radius of the Sun is about 2.95 kilometers. Since this is much less than the radius of the Sun, and the Schwarzschild metric is only valid in the vacuum outside of the Sun,[5] the Schwarzschild metric stops being accurate long before one reaches the Schwarzschild radius. The same is true for most of the objects in our universe, which is why Newtonian gravity is a good approximation in most situations.

If we do take into account the dependence of $|e_r|$ on r, we find that the magnitude of the proper acceleration of a stationary object is

$$\frac{c^2 r_\mathrm{s}}{2r^2}|e_r| = \frac{GM}{r^2}\left(1 - \frac{r_\mathrm{s}}{r}\right)^{-1/2}.$$

The factor $(1 - r_\mathrm{s}/r)^{-1/2}$ approaches ∞ as $r \to r_\mathrm{s}$. So if r is close to r_s, the gravitational field is much stronger than Newtonian gravity would predict.

The sphere at $r = r_\mathrm{s}$ is known as the event horizon. For reasons explained above, most objects in nature do not have an event horizon, but in this paper we are considering the pure Schwarzschild metric, which amounts to treating the central object as a point mass. (If we were modeling the Sun, we might use the Schwarzschild metric for $r > r_g$ and another (non-vacuum) metric for $r < r_g$, where r_g is the radius of the Sun.)

2.3 Gravitational Time Dilation

The length of a stationary curve beginning at $(0, r, \theta, \phi)$ and ending at (ct, r, θ, ϕ) is

$$ct|e_0| = ct\sqrt{-g_{00}} = ct\sqrt{1 - \frac{r_\mathrm{s}}{r}}.$$

The elapsed proper time is therefore related to the elapsed coordinate time by a factor of $\sqrt{1 - r_\mathrm{s}/r}$. This approaches 1 as $r \to \infty$, but approaches 0 as $r \to r_\mathrm{s}$. As a result, a clock closer to the central mass will turn slower than a clock farther from the central mass from the perspective of an outside observer. This effect is known as gravitational time dilation. Since proper time approaches coordinate time as $r \to \infty$, we say that the time coordinate $t = x_0/c$ is time as measured by an observer at infinity.

Since the ratio of proper time to coordinate time approaches 0 as $r \to r_\mathrm{s}$, a falling object will never actually reach the event horizon from the perspective of an outside observer. (In the object's own reference frame, it does reach the event horizon in finite time, as we will see in Section 2.6.)

Gravitational time dilation is taken into account in the engineering of the atomic clocks on GPS satellites. The gravitational time dilation experienced by these satellites relative to the Earth's surface amounts to about 46 microseconds per

[5] Approximately valid, that is. The Sun is not a perfect sphere of stationary matter (and the space around it is not a perfect vacuum), but for our purposes it can be modeled as such.

day (although it is partially countered by the kinematic time dilation of 7 µs per day in the opposite direction). To account for this, the atomic clocks are calibrated to run slightly slower than a typical atomic clock [3][1].

2.4 Spatial Curvature

The distance between (ct, r_0, θ, ϕ) and (ct, r_1, θ, ϕ) is not simply $r_1 - r_0$. In other words, the r coordinate is not an accurate reflection of radial distances. Instead, since g_{11} is larger closer to r_s, equal changes in r correspond to greater proper distances closer to r_s. For example, the proper distance between $2r_\text{s}$ and $3r_\text{s}$ is greater than that between $3r_\text{s}$ and $4r_\text{s}$.

More precisely, the length of a radial line segment from $(ct, r_\text{s}, \theta, \phi)$ to (ct, R, θ, ϕ) is

$$\begin{aligned} D(R) &= \int_{r_\text{s}}^{R} \sqrt{g_{11}(r)}\, \mathrm{d}r = \int_{r_\text{s}}^{R} \left(1 - \frac{r_\text{s}}{r}\right)^{-1/2} \mathrm{d}r \\ &= R\sqrt{1 - \frac{r_\text{s}}{R}} + r_\text{s} \tanh^{-1}\left(\sqrt{1 - \frac{r_\text{s}}{R}}\right), \end{aligned}$$

and in general, the distance from (ct, r_0, θ, ϕ) to (ct, r_1, θ, ϕ) is $D(r_1) - D(r_0)$.

Graphed, the function D looks like this:

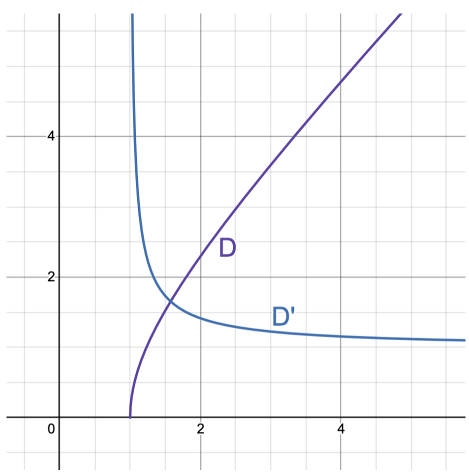

Figure 1: $D(r)$ = proper distance from r_s to r. $D'(r) = \sqrt{g_{11}(r)}$ = ratio of proper radial distance to coordinate radial distance in a small neighborhood.

Although the r coordinate does not accurately measure radial distance, it does have a physical meaning: the circumference of a circle with constant r-coordinate centered at the origin is $2\pi r$. To prove this, consider the curve $\gamma : [0, 2\pi] \to M$,

$\gamma(\theta) = (ct, r, \theta, \phi)$. The proper length of this curve is

$$\int_0^{2\pi} \sqrt{g(\gamma'(\theta), \gamma'(\theta))}\, \mathrm{d}\theta = \int_0^{2\pi} \sqrt{g_{\theta\theta}}\, \mathrm{d}\theta = \int_0^{2\pi} r\, \mathrm{d}\theta = 2\pi r.$$

This is not a coincidence. Recall that in Section 2.1, I got rid of the scaling function C by redefining the r coordinate. The result of this was that r accurately describes the sizes of spheres around the center. So we have a situation where spheres have more space inside of them than their surface area would predict in a Euclidean setting. (This is analogous to the situation on a sphere, in which circles have more area than their circumferences would predict.) Space is curved by gravity.

If we restrict the Schwarzschild metric to the equatorial plane with $\theta = \pi/2$ and t constant, then it can be isometrically embedded into 3D Euclidean space. In cylindrical coordinates, one such embedding is given by

$$(r, \phi) \mapsto (r, \phi, 2\sqrt{r_{\mathrm{s}}(r - r_{\mathrm{s}})}).$$

The image of this embedding is known as Flamm's paraboloid (despite not being a paraboloid).

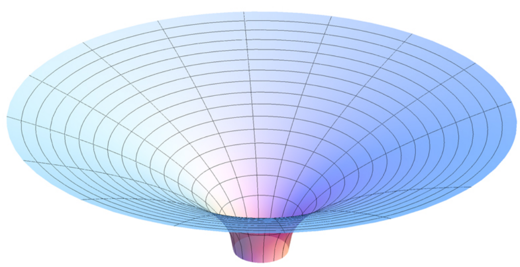

Figure 2: Flamm's paraboloid. Radial distances in the Schwarzschild metric are equal to radial distances along the "paraboloid".

2.5 Radial Lightlike Geodesics

We would like to find a vector field in the (x_0, r) plane that describes the trajectories of light beams. Such a vector field will be of the form $v = ue_0 + e_1$ for some scalar field u, and will satisfy $g(v, v) = 0$. Solving this equation:

$$\begin{aligned} u^2 g_{00} + g_{11} &= 0 \\ -u^2 \left(1 - \frac{r_{\mathrm{s}}}{r}\right) + \left(1 - \frac{r_{\mathrm{s}}}{r}\right)^{-1} &= 0 \\ u^2 = \left(1 - \frac{r_{\mathrm{s}}}{r}\right)^{-2} &= \left(\frac{r}{r - r_{\mathrm{s}}}\right)^2 \\ u &= \pm\frac{r}{r - r_{\mathrm{s}}}. \end{aligned}$$

A lightlike curve γ in the (x_0, r) plane must satisfy

$$\gamma' = v = \pm \frac{r}{r - r_\mathrm{s}} e_0 + e_1. \tag{1}$$

(For an ingoing light beam, the $\pm$ is $-$; for an outgoing light beam, the $\pm$ is $+$.) If we think of γ as a function of r, we can integrate equation (1) to get

$$\gamma(r) = (\pm(r + r_\mathrm{s} \log |r - r_\mathrm{s}|) + C,\ r)$$

where C is some constant.[6] Plotted for various values of C, with x_0 as the vertical axis and r as the horizontal axis, the lightlike curves look like this:

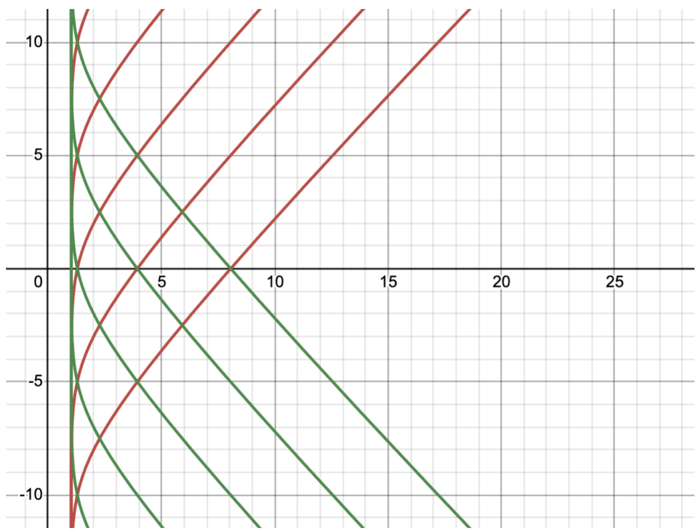

Figure 3: Ingoing light beams are in green; outgoing light beams are in red.

Strangely, the ingoing light beams never actually cross the event horizon (in this coordinate system), instead "slowing down" asymptotically. This does not violate the constancy of the speed of light, because the t and r coordinates do not accurately reflect proper time and distance when r is near r_s.

Although we have thus far focused on the region where $r > r_\mathrm{s}$, we can also plot lightlike curves in the region where $r < r_\mathrm{s}$.

At first glance, it may appear that the blue curves in Figure 4 are outgoing, while the black curves are ingoing. But recall from Section 2.1 that r, not t, is the timelike coordinate inside the event horizon. To interpret trajectories inside the event horizon, we need to make a choice: should decreasing r mean moving forward in time, or should increasing r mean moving forward in time? Under the former convention, all lightlike and timelike trajectories are ingoing (in the

[6] log denotes the natural logarithm.

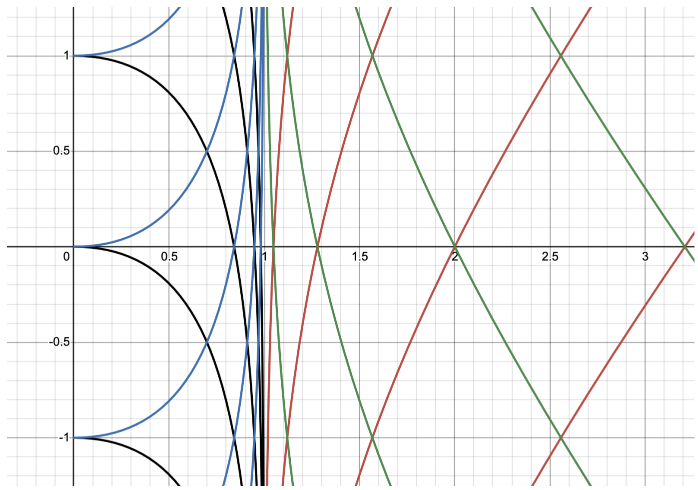

Figure 4: Lightlike curves inside and outside the event horizon.

sense that r decreases), and the spacetime is known as a "black hole". Under the latter convention, all lightlike and timelike trajectores are outgoing, and the spacetime is known as a "white hole". Nothing can escape from a black hole, including light, hence the name. Conversely, nothing can stay inside of a white hole. White holes are named such not because they necessarily appear white, but simply because white is the opposite of black.

2.6 Crossing the Event Horizon

Although we noted earlier that nothing ever crosses the event horizon from the perspective of an outside observer, things can cross the event horizon in finite proper time. Consider, for instance, the timelike trajectory

$$\gamma(r) = (-2r - 2r_{\mathrm{s}} \log(r - r_{\mathrm{s}}),\ r)$$

in the (x_0, r) plane. We know this curve is timelike because it is one of the ingoing lightlike curves from the previous section but with the time coordinate scaled by a factor of 2. To calculate the proper time along γ between, say,

$r = 2r_{\mathrm{s}}$ and $r = r_{\mathrm{s}}$, we evaluate the integral

$$\begin{aligned}
\frac{1}{c}\int_{r_{\mathrm{s}}}^{2r_{\mathrm{s}}} \sqrt{-g(\gamma'(r))}\,\mathrm{d}r &= \frac{1}{c}\int_{r_{\mathrm{s}}}^{2r_{\mathrm{s}}} \sqrt{-g\left(\frac{-2r_{\mathrm{s}}}{r - r_{\mathrm{s}}}e_0 + e_1\right)}\,\mathrm{d}r \\
&= \frac{1}{c}\int_{r_{\mathrm{s}}}^{2r_{\mathrm{s}}} \sqrt{-\left(\frac{2r_{\mathrm{s}}}{r - r_{\mathrm{s}}}\right)^2 g_{00} - g_{11}}\,\mathrm{d}r \\
&= \frac{1}{c}\int_{r_{\mathrm{s}}}^{2r_{\mathrm{s}}} \sqrt{\frac{4r_{\mathrm{s}}^2 - r^2}{r(r - r_{\mathrm{s}})}}\,\mathrm{d}r \\
&= \frac{r_{\mathrm{s}}}{c}\int_{1}^{2} \sqrt{\frac{4 - u^2}{u(u-1)}}du \\
&\approx 2.56\frac{r_{\mathrm{s}}}{c}.
\end{aligned}$$

Since this is finite, an object traveling along γ will cross the event horizon in finite proper time, even while an outside observer sees it slow down and never quite reach the event horizon.

At the moment, we cannot describe what happens to an object after it crosses the event horizon, since the object's t coordinate goes to infinity at $r = r_{\mathrm{s}}$. In Section 2.8, we will define a coordinate transformation that joins up paths inside the event horizon with paths on the outside.

2.7 Photon Sphere

Suppose that a light beam is traveling in circles around the center. The motion of such a light beam is described by a curve $\gamma(t) = (ct, r, \pi/2, kt)$, where r and k are constants. Note that $\gamma'(t) = ce_0 + ke_3$, so for γ to be lightlike, we must have $c^2 g_{00} + k^2 g_{33} = 0$, that is

$$-c^2\left(1 - \frac{r_{\mathrm{s}}}{r}\right) + k^2 r^2 = 0.$$

Solving for k, we find that

$$k = \frac{c}{r}\sqrt{1 - \frac{r_{\mathrm{s}}}{r}}.$$

For γ to be a possible trajectory of a light beam in a vacuum, it must be a geodesic in addition to being lightlike. The geodesic equation states that

$$\frac{\mathrm{d}^2 x_i}{\mathrm{d}t^2} + \sum_{j,k} \Gamma^i_{jk} \frac{\mathrm{d}x_j}{\mathrm{d}t}\frac{\mathrm{d}x_k}{\mathrm{d}t} = 0.$$

In our case, all of the second derivatives are 0, and the only nonzero first derivatives are $\mathrm{d}x_0/\mathrm{d}t = c$ and $\mathrm{d}x_3/\mathrm{d}t = k$, so we only need to care about Christoffel symbols where both of the lower indices are 0 or 3. The only nonzero Christoffel symbols of this form are

$$\Gamma^1_{00} = \frac{r_\mathrm{s}(r - r_\mathrm{s})}{2r^3} \quad \text{and} \quad \Gamma^1_{33} = r_\mathrm{s} - r.$$

Therefore, the only equation we need to check is that where $i = 1$, namely

$$\frac{r_\mathrm{s}(r - r_\mathrm{s})}{2r^3}c^2 + (r_\mathrm{s} - r)k^2 = 0.$$

Substituting in the value of k we found earlier, we get a quadratic equation in r with solution

$$r = r_\mathrm{s} \quad \text{or} \quad r = \frac{3}{2}r_\mathrm{s}.$$

The first solution, $r = r_\mathrm{s}$, does not correspond to a lightlike geodesic because g is undefined at $r = r_\mathrm{s}$. However, the solution $r = \frac{3}{2}r_\mathrm{s}$ does. The sphere of radius $\frac{3}{2}r_\mathrm{s}$ is known as the photon sphere, because light can orbit around the center at this radius. If you were at $\frac{3}{2}r_\mathrm{s}$, you could look forward and see the back of your head. However, this orbit is unstable in the sense that light at $\frac{3}{2}r_\mathrm{s}$ must have no motion in the r direction in order to orbit forever; if r is decreasing at all, the light will fall toward the event horizon, and if r is increasing at all, the light will escape [4, § 6.2.4].

The photon sphere is an extreme case of the more general phenomenon of gravitational lensing, in which gravity causes light to deviate from a straight path through space.

2.8 Kruskal-Szekeres Coordinates

I have hinted at an alternative coordinate system that allows the two regions $r < r_\mathrm{s}$ and $r > r_\mathrm{s}$ to be united into a single Lorentzian manifold. One such coordinate system is Kruskal-Szekeres coordinates, first described by Martin Kruskal and George Szekeres in 1960. The idea behind Kruskal-Szekeres coordinates is to replace x_0 and r with two new coordinates such that radial light beams travel along straight diagonal lines, rather than the curved lines seen in Figure 4.

To derive Kruskal-Szekeres coordinates, we begin by defining two coordinates u and v such that u is constant along outgoing light beams and v is constant along ingoing light beams. In the region where $r < r_\mathrm{s}$ (which we will interpret as a black hole, i.e. decreasing r means forward in time), the distinction between outgoing and ingoing light beams is replaced by a distinction between light beams that increase in x_0 and light beams that decrease in x_0.

Recall from Section 2.5 that radial lightlike curves are given by the formula

$$x_0 \pm (r + r_\mathrm{s} \log |r - r_\mathrm{s}|) = C$$

where C is some constant, and the $\pm$ is $+$ for ingoing curves and $-$ for outgoing curves. We can modify this formula by replacing C with $C \pm r_\mathrm{s} \log r_\mathrm{s}$ and then incorporating the new term into the logarithm on the left, resulting in

$$x_0 \pm \left(r + r_\mathrm{s} \log \left| \frac{r}{r_\mathrm{s}} - 1 \right| \right) = C.$$

The purpose of this change is to make the quantity in parentheses, which we shall denote r^*, approach 0 as r approaches 0. Without the shift, we would still get a valid coordinate system in the end, but it would differ from the standard Kruskal-Szekeres coordinates by a scale factor of $\sqrt{r_\mathrm{s}}$.

We define u and v as

$$u = x_0 - r^* \qquad v = x_0 + r^*,$$

so that the coordinate lines are radial lightlike curves.

To compute e_u and e_v in terms of e_0 and e_r, we will use the following facts:

$$\begin{aligned} x_0 &= \frac{u+v}{2} \qquad & r^* &= \frac{-u+v}{2} \\ \frac{\mathrm{d}r^*}{\mathrm{d}r} &= \frac{r}{r - r_\mathrm{s}} \qquad & \frac{\mathrm{d}r}{\mathrm{d}r^*} &= \frac{r - r_\mathrm{s}}{r} = 1 - \frac{r_\mathrm{s}}{r}. \end{aligned}$$

Now we can compute e_u and e_v (also denoted $\partial/\partial u$ and $\partial/\partial v$) as follows:

$$\begin{aligned} \frac{\partial}{\partial u} &= \frac{\partial x_0}{\partial u}\frac{\partial}{\partial x_0} + \frac{\partial r^*}{\partial u}\frac{\mathrm{d}r}{\mathrm{d}r^*}\frac{\partial}{\partial r} = \frac{1}{2}\frac{\partial}{\partial x_0} - \frac{1}{2}\left(1 - \frac{r_\mathrm{s}}{r}\right)\frac{\partial}{\partial r}. \\ \frac{\partial}{\partial v} &= \frac{\partial x_0}{\partial v}\frac{\partial}{\partial x_0} + \frac{\partial r^*}{\partial v}\frac{\mathrm{d}r}{\mathrm{d}r^*}\frac{\partial}{\partial r} = \frac{1}{2}\frac{\partial}{\partial x_0} + \frac{1}{2}\left(1 - \frac{r_\mathrm{s}}{r}\right)\frac{\partial}{\partial r}. \end{aligned}$$

Both of these vector fields are lightlike everywhere; e_u points along ingoing light beams, while e_v points along out outgoing light beams.

Inside the event horizon, r^* ranges from 0 to $-\infty$. Therefore, u and v both range from $-\infty$ to ∞, but with the restriction that $(-u+v)/2 < 0$, i.e. $u > v$. Outside the event horizon, r^* ranges from $-\infty$ to ∞, so (u, v) can be any pair of numbers. Note that it is possible for a point in the inner region and a point in the outer region to be described by the same coordinates (u, v, θ, ϕ) (as seen with points A and B in Figure 5), so although (u, v, θ, ϕ) constitutes a coordinate system on the outer region, and also on the inner region, it cannot cover both regions at once while remaining injective. Moreover, u and v are not even defined at $r = r_\mathrm{s}$ due to the logarithm in their definitions.

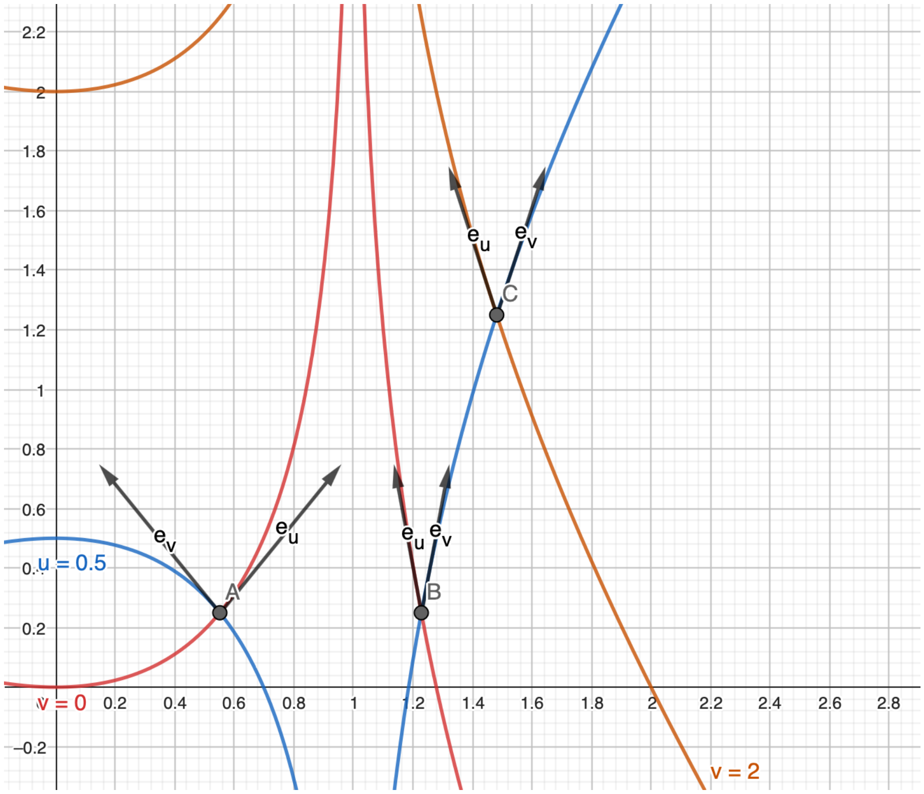

Figure 5: Tangent vectors e_u and e_v at various points on the (x_0, r) plane (with r the horizontal axis and $r_s = 1$).

To get rid of this logarithm, we define new coordinates[7]

$$U = \operatorname{sgn}(r_\mathrm{s} - r)\exp\left(\frac{-u}{2r_\mathrm{s}}\right) = \operatorname{sgn}(r_\mathrm{s} - r)\exp\left(\frac{-x_0}{2r_\mathrm{s}}\right)\exp\left(\frac{r}{2r_\mathrm{s}}\right)\left|\frac{r}{r_\mathrm{s}} - 1\right|^{1/2}$$
$$V = \exp\left(\frac{v}{2r_\mathrm{s}}\right) = \exp\left(\frac{x_0}{2r_\mathrm{s}}\right)\exp\left(\frac{r}{2r_\mathrm{s}}\right)\left|\frac{r}{r_\mathrm{s}} - 1\right|^{1/2}$$

with inverse transformation given by

$$u = -2r_\mathrm{s}\log|U| \qquad v = 2r_\mathrm{s}\log|V|.$$

The basis vectors e_U and e_V are parallel to e_u and e_v respectively, so they are still lightlike. If we want light to travel along diagonal lines, then the last step

[7]sgn here is the sign function, which returns -1 for negative inputs and $+1$ for positive inputs.

is to define new coordinates

$$T = \frac{U+V}{2} = \begin{cases} \sinh\left(\frac{x_0}{2r_\mathrm{s}}\right)\exp\left(\frac{r}{2r_\mathrm{s}}\right)\left(\frac{r}{r_\mathrm{s}}-1\right)^{1/2} & r > r_\mathrm{s} \\ \cosh\left(\frac{x_0}{2r_\mathrm{s}}\right)\exp\left(\frac{r}{2r_\mathrm{s}}\right)\left(1-\frac{r}{r_\mathrm{s}}\right)^{1/2} & r < r_\mathrm{s} \end{cases}$$

$$X = \frac{-U+V}{2} = \begin{cases} \cosh\left(\frac{x_0}{2r_\mathrm{s}}\right)\exp\left(\frac{r}{2r_\mathrm{s}}\right)\left(\frac{r}{r_\mathrm{s}}-1\right)^{1/2} & r > r_\mathrm{s} \\ \sinh\left(\frac{x_0}{2r_\mathrm{s}}\right)\exp\left(\frac{r}{2r_\mathrm{s}}\right)\left(1-\frac{r}{r_\mathrm{s}}\right)^{1/2} & r < r_\mathrm{s} \end{cases}$$

with inverse transformation

$$U = T - X \qquad V = T + X.$$

In the outer region where $r > r_\mathrm{s}$, U is negative and V is positive, which is equivalent to $-X < T < X$. In the inner region where $r < r_\mathrm{s}$, U and V are both positive, or equivalently, $-T < X < T$. These two inequalities are mutually exclusive, so unlike the (u, v, θ, ϕ) coordinate system, the (T, X, θ, ϕ) coordinate system can cover both regions simultaneously. Moreover, the inequality $u > v$ which is satisfied at all points in the inner region translates to $UV < 1$, or $T^2 - X^2 < 1$. The domain of (T, X) is shown in Figure 6.

To write the Schwarzschild metric in Kruskal-Szekeres coordinates, we first note that

$$\mathrm{d}r^2 = \left(1 - \frac{r_s}{r}\right)^2 \mathrm{d}r^{*2},$$

and therefore

$$g = \left(1 - \frac{r_\mathrm{s}}{r}\right)(-\mathrm{d}x_0^2 + \mathrm{d}r^{*2}) + r^2\mathrm{d}\theta^2 + r^2(\sin\theta)^2\mathrm{d}\phi^2. \tag{2}$$

The next step is to write $\mathrm{d}x_0$ and $\mathrm{d}r^*$ in terms of $\mathrm{d}T$ and $\mathrm{d}X$. By composing transformations given above, we find that

$$x_0 = r_s \log\left|\frac{T+X}{T-X}\right| \qquad r^* = r_s \log\left|T^2 - X^2\right|$$

and therefore

$$\mathrm{d}x_0 = \frac{\partial x_0}{\partial T}\mathrm{d}T + \frac{\partial x_0}{\partial X}\mathrm{d}X = \frac{-2r_\mathrm{s}X}{UV}\mathrm{d}T + \frac{2r_\mathrm{s}T}{UV}\mathrm{d}X.$$
$$\mathrm{d}r^* = \frac{\partial r^*}{\partial T}\mathrm{d}T + \frac{\partial r^*}{\partial X}\mathrm{d}X = \frac{2r_\mathrm{s}T}{UV}\mathrm{d}T + \frac{-2r_\mathrm{s}X}{UV}\mathrm{d}X.$$

We now compute

$$-\mathrm{d}x_0^2 + \mathrm{d}r^{*2} = \frac{-4r_\mathrm{s}^2}{UV}(-\mathrm{d}T^2 + \mathrm{d}X^2)$$

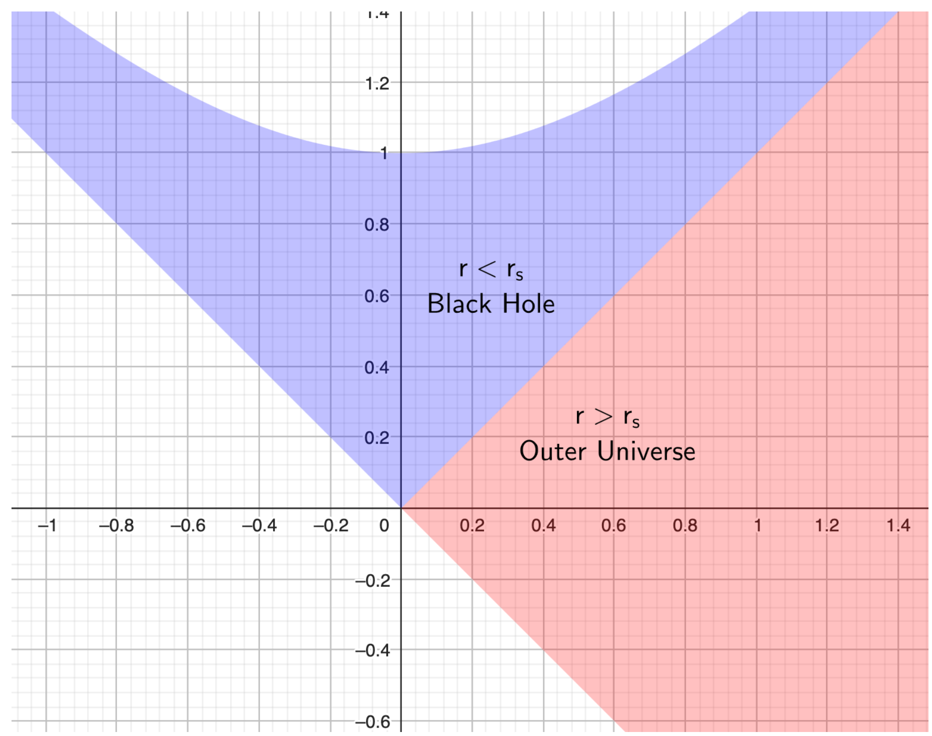

Figure 6: The black hole and outer universe in Kruskal-Szekeres coordinates, with T the vertical axis and X the horizontal axis.

and

$$UV = \operatorname{sgn}(r_\mathrm{s} - r) \exp\left(\frac{r^*}{r_\mathrm{s}}\right) = \left(1 - \frac{r}{r_\mathrm{s}}\right) \exp\left(\frac{r}{r_\mathrm{s}}\right). \tag{3}$$

Substituting these formulas into equation (2), we get

$$g = \frac{4r_\mathrm{s}^2}{r} \exp\left(\frac{-r}{r_\mathrm{s}}\right) (-\mathrm{d}T^2 + \mathrm{d}X^2) + r^2 \mathrm{d}\theta^2 + r^2 (\sin\theta)^2 \mathrm{d}\phi^2.$$

Note that there is no longer any singularity at $r = r_\mathrm{s}$; the metric remains well-defined and nondegenerate. Note also that T is timelike everywhere, while X is spacelike everywhere.

Although this is the formula typically given for the Schwarzschild metric in Kruskal-Szekeres coordinates, it is not fully in Kruskal-Szekeres coordinates due to the appearance of r in the coefficients. We have already written x_0 and r^* in terms of T and X, but have not yet written r in terms of T and X. To do this, we can solve equation (3) using the Lambert W function,[8] resulting in

$$r = r_{\mathrm{s}}\left(1 + W\left(\frac{-UV}{e}\right)\right) = r_{\mathrm{s}}\left(1 + W\left(\frac{X^2 - T^2}{e}\right)\right).$$

2.9 Maximal Extension, Black Hole and White Hole

If we view the Kruskal-Szekeres form of the Schwarzschild metric on its own terms, there is no reason why it must be restricted to the region $T + X > 0$, as in Figure 6. The only necessary restriction placed on T and X is the inequality $T^2 - X^2 < 1$ arising from the use of the Lambert W function, which is only defined for numbers greater than $-1/e$.

If we extend the Kruskal-Szekeres Schwarzschild metric as far as possible, we get the spacetime with 4 regions shown in Figure 7. It is easy to see that the functions $(T, X, \theta, \phi) \mapsto (T, -X, \theta, \phi)$ and $(T, X, \theta, \phi) \mapsto (-T, X, \theta, \phi)$ are isometries, since T and X (and their differentials) only appear squared in the definitions of g and r. So region III is isometric to region I, and region IV is isometric to region II. However, if we care about the orientation of time, then regions II and IV are distinct, since the isometry that relates them reverses time. Specifically, whereas region II is a black hole, region IV is a white hole, under the usual convention of increasing T meaning forward in time. So, whereas the original coordinates contain an ambiguous region that is either a black hole or a white hole, as seen in Section 2.5, Kruskal-Szekeres coordinates contain both a black hole and a white hole.

As we showed in the previous section, light beams travel along diagonal lines in the (T, X) plane, just as in the Minkowski diagrams of special relativity. Objects traveling slower than c have worldlines that are more vertical than horizontal. Assuming nothing can travel faster than c, it is impossible to travel out of region II or into region IV, or to travel between regions I and III. Everything in region II eventually hits the singularity at $T = \sqrt{1 + X^2}$, while everything in region IV originated from the singularity at $T = -\sqrt{1 + X^2}$.

Region III is a "parallel universe" isometric to but causally disconnected from region 1. While it is impossible to travel between regions I and III, an object from region I could travel into the black hole and encounter an object that originated from region III.

The existence of region III is already hinted at in the original coordinate system. If we interpret the left side of Figure 4 as a black hole, then the blue curves represent light beams that originated from region I, while the black curves represent light beams that originated from region III.

Kruskal-Szekeres coordinates are not the only alternative coordinate system for the Schwarzschild universe. Earlier coordinate systems, such as Lemaître coordinates and Eddington-Finkelstein coordinates, demonstrated that the coordinate

[8] W is the inverse of the restriction of $x \mapsto xe^x$ to numbers greater than -1, and is a strictly increasing function from $(-1/e, \infty)$ onto $(-1, \infty)$.

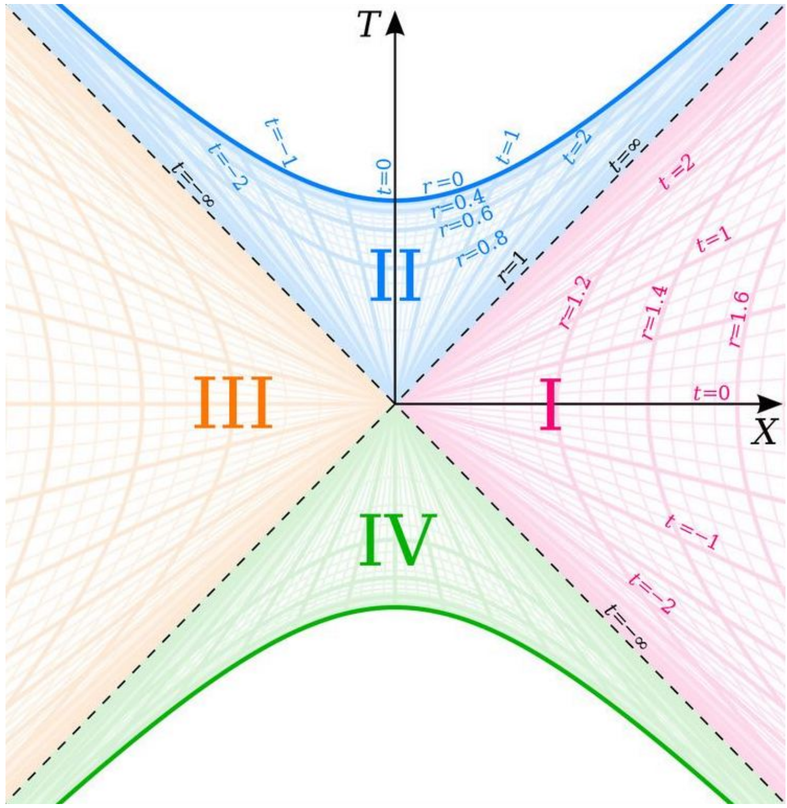

Figure 7: Maximal extension of the Schwarzschild metric in Kruskal-Szekeres coordinates, along with lines of constant t and r. In this image, $r_{\mathrm{s}} = 1$ and natural units are used, so x_0 is identified with t [SOURCE].

singularity at $r = r_{\mathrm{s}}$ is not a physical singularity while still only covering two of the four spacetime regions.

For a long time after the Schwarzschild metric was discovered, both black holes and white holes were widely thought not to exist. Although it was known that an object could collapse into a black hole if it were dense enough, many physicists (including Einstein) predicted that various forces would prevent this density from being reached. However, it is now believed that black holes do form after the collapse of large stars, and astronomers have identified many objects that are best explained as black holes. White holes, on the other hand, are thought not to exist, as unlike the maximally extended Schwarzschild spacetime, the spacetime of collapsing star contains only a black hole, not a white hole. Likewise, parallel universes like the region III described above are also predicted not to exist.

2.10 Graviational Singularity

A (pseudo-)Riemannian manifold is said to be non-singular if the domain of any geodesic can be extended to all of $\mathbb{R}$. If this is not the case, the manifold is singular, and is said to have a "singularity" wherever geodesics fail to admit extensions. These "singularities" are not actually points on the manifold, but they may have coordinates in some coordinate systems. For example, the Schwarzschild metric has a singularity at $r = 0$ in radial coordinates, and at $T^2 - X^2 = 1$ in Kruskal-Szekeres coordinates.

Of particular interest in general relativity are timelike singularities (failures of timelike geodesics to be extended), and those singularities that cannot be resolved by embedding the manifold into a larger one (often accomplished via a change of coordinates, as is the case with the singularity at $r = r_{\mathrm{s}}$ in the Schwarzschild metric). These "gravitational singularities" are important because they constitute failures of general relativity to make predictions. Nothing can be said about what happens to an object after it reaches a singularity, because the manifold of spacetime ends there.

Although the singularity at $r = 0$ in the Schwarzschild metric is often said to lie in the "center" of a black hole, this is misleading, because inside the event horizon, r is a time coordinate, not a spatial coordinate. So the singularity is better understood as a boundary where time ends.

Understanding what happens at and near gravitational singularities is a goal of modern theoretical physics. Many physicists believe that singularities do not actually exist in our universe, and that the long-sought quantum theory of gravity will not contain singularities.

2.11 Gravitational Redshift

Recall that the exterior derivative of a 1-form is defined in coordinates as

$$\mathrm{d}\left(\sum_i f_i \, \mathrm{d}x_i\right) = \sum_{i,j} \frac{\partial f_i}{\partial x_j} \, \mathrm{d}x_j \wedge \mathrm{d}x_i.$$

In general relativity, a wave can be represented as a 1-form whose exterior derivative is 0. The frequency of the wave as measured by an observer is found by evaluating the 1-form on that observer's 4-velocity vector. In particular, an observer moving along with a wave measures its frequency to be 0, so the 1-form evaluates to 0 on any multiple of such an observer's 4-velocity. For a wave propagating at a speed of c, then, the wave's 1-form should vanish on lightlike vectors in the direction the wave is propagating in. Using this condition, and the formula for radial lightlike vectors derived in Section 2.5 (namely equation (1)) we can begin to derive a formula for the 1-form corresponding to a radial light wave:

$$\omega = \alpha(r)\left(\mathrm{d}x_0 \pm \frac{r}{r - r_{\mathrm{s}}} \mathrm{d}r\right)$$

where α is some function, and the $\pm$ is $+$ for an ingoing light wave and $-$ for an outgoing light wave. The function α can be found using the condition that $\mathrm{d}(\omega) = 0$. Using the formula for ω above, we have

$$\mathrm{d}(\omega) = \frac{\mathrm{d}\alpha}{\mathrm{d}r}\,\mathrm{d}r \wedge \mathrm{d}x_0,$$

which is zero only when α is constant. This leaves one degree of freedom in the description of radial light waves (namely the constant value of α) which is proportional to the frequency of the light.

Recall that an observer with constant r coordinate has a 4-velocity of

$$v(r) = \frac{c}{\sqrt{1 - \frac{r_\mathrm{s}}{r}}} e_0.$$

This observer will measure the light wave described by ω to have a frequency of

$$f(r) = \omega(v(r)) = \frac{c\alpha}{\sqrt{1 - \frac{r_\mathrm{s}}{r}}}.$$

As r increases, $f(r)$ approaches $c\alpha$. Thus α has a physical interpretation as the wave number (number of wavelengths per unit distance) of the light wave as measured by an observer at infinity. However, for all finite values of r, $f(r)$ is greater than $c\alpha$, and it is especially high (approaching ∞) for r close to r_s. Therefore, a yellow light source close to a black hole may appear blue to an observer even closer to the black hole, and red to an observer farther out. In fact, the light emitted by any star in our universe is slightly redshifted on its way outward.

3 The FLRW Spacetime

While the Schwarzschild metric (and the more general Kerr metric, which describes space around a rotating mass) can be useful for modeling small[9] objects in our universe, the FLRW metric models our universe at the largest scales. Named after its early investigators Alexander Friedmann, Georges Lemaître, Howard Robinson, and Arthur Walker, it describes an expanding or contracting universe with positive, zero, or negative spatial curvature.

[9] From a cosmological point of view.

3.1 Derivation

The most commonly encountered parametrization of the 3-sphere as a subset of 4-dimensional Euclidean space is as follows:

$$\begin{aligned}
x_1 &= \cos\chi \\
x_2 &= \sin\chi\cos\theta \\
x_3 &= \sin\chi\sin\theta\cos\phi \\
x_4 &= \sin\chi\sin\theta\sin\phi.
\end{aligned}$$

The coordinates χ and θ range from 0 to π, while ϕ is 2π-periodic. The corresponding basis vectors are defined by

$$\frac{\partial}{\partial\chi} = \sum_i \frac{\partial x_i}{\partial\chi}\frac{\partial}{\partial x_i},$$

and likewise for θ and ϕ. By computing the dot products of these basis vectors, one arrives at an expression for the pullback of the Euclidean metric onto the 3-sphere:

$$\mathrm{d}\chi^2 + (\sin\chi)^2\,\mathrm{d}\theta^2 + (\sin\chi)^2(\sin\theta)^2\,\mathrm{d}\phi^2.$$

The resulting Riemannian manifold (which we will henceforth regard as a manifold in its own right, and not as a subset of 4-dimensional space) has constant positive Ricci scalar curvature. To obtain a manifold with constant negative curvature, we consider the hypersurface

$$H = \{(x_1, x_2, x_3, x_4) \in \mathbb{R}^4 \mid x_1^2 - x_2^2 - x_3^2 - x_4^2 = 1 \text{ and } x_1 > 0\}$$

and pull back not the Euclidean metric, but rather the $(-,+,+,+)$ Minkowski metric. H can be parametrized as follows:

$$\begin{aligned}
x_1 &= \cosh\chi \\
x_2 &= \sinh\chi\cos\theta \\
x_3 &= \sinh\chi\sin\theta\cos\phi \\
x_4 &= \sinh\chi\sin\theta\sin\phi,
\end{aligned}$$

with χ now ranging from 0 to ∞. The components of the pullback metric are the results of evaluating the Minkowski metric on each pair of tangent vectors:

$$\mathrm{d}\chi^2 + (\sinh\chi)^2\,\mathrm{d}\theta^2 + (\sinh\chi)^2(\sin\theta)^2\,\mathrm{d}\phi^2.$$

Note that the two metrics we have obtained are identical except for the alternation of $\sin\chi$ with $\sinh\chi$. This motivates the introduction of a new coordinate, r, defined as $\sin\chi$ in the positive-curvature case and $\sinh\chi$ in the negative-curvature case. It is straightforward to obtain an expression for $\mathrm{d}\chi$ in terms of r and $\mathrm{d}r$:

$$\begin{aligned}
\chi &= \sin^{-1} r & \qquad \chi &= \sinh^{-1} r \\
\mathrm{d}\chi &= \frac{1}{\sqrt{1-r^2}}\,\mathrm{d}r & \qquad \mathrm{d}\chi &= \frac{1}{\sqrt{1+r^2}}\,\mathrm{d}r
\end{aligned}$$

This coordinate change allows us to write both metrics in the same form:

$$\Sigma = \frac{1}{1 - kr^2}\,\mathrm{d}r^2 + r^2\,\mathrm{d}\theta^2 + r^2(\sin\theta)^2\,\mathrm{d}\phi^2.$$

If $k = +1$, we obtain the positive-curvature metric, and if $k = -1$, we obtain the negative-curvature metric. Moreover, setting k to 0 results in the Euclidean (zero-curvature) metric in standard spherical coordinates. Note that the coordinate r ranges from 0 to 1 in the positive-curvature case, and from 0 to ∞ in the other two cases. Also note that this coordinate system, in the positive-curvature case, only covers half of the 3-sphere due to the sine function (which we used to define r) being non-injective on $[0, \pi]$.

To get the FLRW metric, we add a time coordinate $x_0 = ct$ and scale the spatial part of the metric by a time-varying factor $a(t)$:

$$\begin{aligned} g &= -\mathrm{d}x_0^2 + \frac{a(t)^2}{1 - kr^2}\,\mathrm{d}r^2 + a(t)^2 r^2\,\mathrm{d}\theta^2 + a(t)^2 r^2(\sin\theta)^2\,\mathrm{d}\phi^2 \\ &= -\mathrm{d}x_0^2 + a(t)^2\,\Sigma. \end{aligned}$$

When $a(t)$ increases, spacelike vectors such as e_r become longer — that is, the same change in r constitutes a greater proper distance — and the universe is said to be expanding. When $a(t)$ decreases, spacelike vectors become shorter, and the universe is said to be contracting. This terminology is justified by the fact that paths of constant r, θ, and ϕ are geodesics,[10] so that test particles tend to get farther apart in an expanding universe, and closer together in a contracting universe. An object with constant r, θ, and ϕ is said to be "comoving," and the coordinate system we are using is known as "comoving coordinates."

Note that the r coordinate is dimensionless, while $a(t)$ has units of distance. If $k = 1$, then $a(t)$ is the radius of the 3-spherical universe at time t (because we began with a unit 3-sphere and then scaled it by $a(t)$). Similarly, if $k = -1$, then $a(t)$ is radius of curvature of the hyperbolic universe. However, if $k = 0$, then the value of $a(t)$ at some fixed time t has no physical meaning; it is only changes in $a(t)$ that matter. This is seen in the fact that replacing $a(t)$ with $Ba(t)$ and r with r/B, where B is any positive constant, leaves the metric unchanged.[11]

The Ricci scalar curvature of the spatial slice at time t is $6ka(t)^{-2}$. This affirms the claims made earlier about the sign and constancy of the spatial curvature, and demonstrates the important fact that curvature is inversely proportional to the square of the scaling factor.

3.2 Friedmann Equations

Unlike the Schwarzschild metric, the FLRW metric does not satisfy the Einstein vacuum equation, outside of a few specific choices of k and a. The Ricci tensor

[10] The Christoffel symbols Γ^i_{00} are all 0.

[11] This is not the case if $k \neq 0$ due to the kr^2.

and scalar of the FLRW metric are

$$\begin{aligned}\mathrm{Ric} &= \frac{-3\ddot{a}}{c^2 a}\,\mathrm{d}x_0^2 + \frac{2kc^2 + 2\dot{a}^2 + a\ddot{a}}{c^2}\,\Sigma \\ R &= \frac{6kc^2 + 6\dot{a}^2 + 6a\ddot{a}}{c^2 a^2}.\end{aligned}$$

Using the Einstein field equation, the covariant stress-energy tensor can be computed:

$$\begin{aligned}T &= \frac{c^4}{8\pi G}\left(\mathrm{Ric} - \frac{1}{2}Rg + \Lambda g\right) \\ &= -\frac{\Lambda c^4 a^2 - 3kc^4 - 3c^2\dot{a}^2}{8\pi G a^2}\,\mathrm{d}x_0^2 + \frac{\Lambda c^4 a^2 - kc^4 - c^2\dot{a}^2 - 2c^2 a\ddot{a}}{8\pi G}\,\Sigma. \qquad (4)\end{aligned}$$

This turns out to be the stress-energy tensor of a comoving "perfect fluid"; that is, a fluid with no shear stresses or viscosity, characterized entirely by its density ρ and its pressure q,[12] which may vary over time but not over space. Specifically, the covariant stress-energy tensor of a perfect fluid is

$$T = \left(\rho + \frac{q}{c^2}\right) U \otimes U + qg,$$

where U is the covariant 4-velocity of the fluid, derived from its (contravariant) 4-velocity by contraction with g. If the fluid is comoving, then its 4-velocity is $u = ce_0$ and its covariant 4-velocity is $g(u, -) = -c\,\mathrm{d}x_0$. So the stress-energy tensor is

$$T = (\rho c^2 + q)\,\mathrm{d}x_0^2 + qg = \rho c^2\,\mathrm{d}x_0^2 + qa^2\,\Sigma. \qquad (5)$$

At this point, we have two expressions for the stress-energy tensor: equation (4), derived from the curvature of spacetime via the Einstein field equation, and equation (5), derived from the density and pressure. Equating these expressions will yield two equations — the Friedmann equations — which relate the density and pressure of the universe to its curvature, cosmological constant, and rate of expansion. The first Friedmann equation results from taking the $\mathrm{d}x_0^2$ component of each expression:

$$-\frac{\Lambda c^4 a^2 - 3kc^4 - 3c^2\dot{a}^2}{8\pi G a^2} = \rho c^2.$$

This equation attains a more commonly-cited form after some algebraic manipulation:

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G\rho}{3} + \frac{\Lambda c^2}{3} - \frac{kc^2}{a^2}. \qquad (6)$$

The second Friedmann equation results from contracting each expression against the contravariant metric $g^{-1} = -e_0^2 + a^{-2}\Sigma^{-1}$. For a tensor expressed as a

[12] It is usual to represent pressure with the symbol p, but I will use q to avoid confusion with the similar-looking letter ρ.

linear combination of $\mathrm{d}x_0^2$ and Σ, say $f\mathrm{d}x_0^2 + h\Sigma$, this results in the scalar field $-f + 3a^{-2}h$. Applying this to the expressions in (4) and (5) gives

$$\frac{2\Lambda c^4 a^2 - 3kc^4 - 3c^2\dot{a}^2 - 3c^2 a\ddot{a}}{4\pi G a^2} = -\rho c^2 + 3q,$$

which after some algebraic manipulation becomes

$$\frac{\ddot{a}}{a} + \frac{\dot{a}^2 + kc^2}{a^2} = \frac{2\Lambda c^2}{3} + \frac{4\pi G}{3}\left(\rho - \frac{3q}{c^2}\right).$$

The second term on the left hand side matches two terms of the first Friedmann equation (6). Applying that equation yields

$$\frac{\ddot{a}}{a} + \frac{8\pi G\rho}{3} + \frac{\Lambda c^2}{3} = \frac{2\Lambda c^2}{3} + \frac{4\pi G}{3}\left(\rho - \frac{3q}{c^2}\right),$$

which, after some more algebraic manipulation, results in the second Friedmann equation:

$$\frac{\ddot{a}}{a} = \frac{\Lambda c^2}{3} - \frac{4\pi G}{3}\left(\rho + \frac{3q}{c^2}\right). \tag{7}$$

The first Friedmann equation relates $\dot{a}$ (the rate of expansion of the universe) to k, Λ and ρ, while the second relates $\ddot{a}$ (the acceleration of the universe's expansion) to Λ, ρ, and q.

3.3 Change in Density

The relationship between the two Friedmann equations deserves some clarification. The first involves $\dot{a}$ and the second involves $\ddot{a}$, but the second equation is not simply the derivative of the first equation. It is natural, then, to ask what restrictions must hold for both equations to be true. It turns out that, if we take the derivative of the first Friedmann equation and compare it with the second Friedmann equation, we get an equation for $\dot{\rho}$, the rate of change of mass-energy density.

The exact derivation is as follows: we first multiply both sides of (6) by a^2 to isolate the $\dot{a}^2$, resulting in

$$\dot{a}^2 = \frac{8\pi G\rho a^2}{3} + \frac{\Lambda c^2 a^2}{3} - kc^2.$$

We then take the derivative of both sides and solve for $\frac{\ddot{a}}{a}$:

$$\frac{\ddot{a}}{a} = \frac{4\pi G\dot{\rho}a}{3\dot{a}} + \frac{8\pi G\rho}{3} + \frac{\Lambda c^2}{3}. \tag{8}$$

Equating the right hand side of (8) with that of (7) and solving for $\dot{\rho}$, we get

$$\dot{\rho} = -3\,\frac{\dot{a}}{a}\left(\rho + \frac{q}{c^2}\right). \tag{9}$$

One consequence of this equation is that, if $\rho > 0$ (and $q \geq 0$), then $\dot{\rho}$ and $\dot{a}$ have opposite signs; the density decreases as the universe expands, and increases as the universe contracts.

We can integrate equation (9) if we assume that $q = wc^2\rho$ for some proportionality constant w. For a universe filled with "dust-like" matter that exerts no pressure, $w = 0$; for a universe filled with electromagnetic radiation, $w = \frac{1}{3}$.[13] With this assumption, equation (9) becomes

$$\frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{a}}{a}.$$

Integrating both sides with respect to t results in

$$\log(\rho) = -3(1+w)\log(a) + C,$$

which we can solve for ρ to get

$$\rho = Da^{-3(1+w)}. \tag{10}$$

(Here, $D = \exp(C)$ is some constant resulting from the integration.)

In the case of dust-like matter, equation (10) states that $\rho = Da^{-3}$. This makes sense, because if one imagines the universe as a cloud of dust particles, then the distance between any two particles should be proportional to a, making the average number of particles per unit volume proportional to a^{-3}. In the case of radiation, however, equation (10) becomes $\rho = Da^{-4}$. The additional factor of a^{-1} is due to the fact that, as the universe expands, photons are redshifted in addition to growing farther apart, and the energy of a photon is proportional to its frequency.

For a hypothetical substance with $w = -1$, and thus with negative pressure equal to $-c^2\rho$, equation (9) becomes $\dot{\rho} = 0$ and equation (10) becomes $\rho = D$. In other words, the density of such a substance does not change over time. In Section 3.7, we will see that the cosmological constant Λ has the same effect as such a substance, meaning one can incorporate Λ into ρ and q, increasing the density and decreasing the pressure if Λ is positive, and the reverse if Λ is negative. However, for the sake of classifying FLRW universes, I will treat Λ as a separate parameter, and will assume that $\rho, q \geq 0$.

3.4 Universes with $\Lambda < 0$

If $\Lambda < 0$, then the right side of the second Friedmann equation (7) is always negative, so $\ddot{a}$ is always negative, meaning $\dot{a}$ is always decreasing. In fact, $\dot{a}$ necessarily decreases below 0, as opposed to approaching some limiting value.

[13] This can be derived from the fact that the stress-energy tensor of an electromagnetic field always has zero trace.

To prove this, we choose a positive constant a_1 such that $a > a_1$ at some time. Then the second Friedmann equation gives

$$\ddot{a} \leq \frac{\Lambda c^2 a}{3} < \frac{\Lambda c^2 a_1}{3}.$$

Therefore, as long as a is greater than a_1, $\ddot{a}$ is bounded above by a negative constant, so $\dot{a}$ decreases at a rate that is at least linear. Something decreasing at such a rate must eventually become negative, and the condition $a > a_1$ can only cease to hold if $\dot{a}$ has become negative.

Since $\dot{a}$ is eventually negative and never increases, a is bound to reach 0, causing the universe to end in a singularity — a "big crunch".

If the universe is empty (that is, $\rho = q = 0$), then k must be -1; if not, the right hand side of the first Friedmann equation (6) would be negative, and $(\dot{a}/a)^2$ cannot be negative. Such a universe (an empty universe with negative cosmological constant and thus negative spatial curvature) is known as anti-De Sitter space. In anti-De Sitter space, the first Friedmann equation states that

$$\dot{a}^2 = c^2 + \frac{\Lambda c^2}{3} a^2. \tag{11}$$

This differential equation is solvable. Since the function $y(t) = \sin(t)$ satisfies $\dot{y}^2 = 1 - y^2$, one might expect the solution to (11) to be some kind of sinusoid. Indeed, the solution is

$$a(t) = \sqrt{\frac{-3}{\Lambda}} \sin\left(t\sqrt{\frac{-\Lambda c^2}{3}} + C\right).$$

So, over the course of the universe, the distance between two comoving test particles in anti-De Sitter space increases and then decreases sinusoidally.

3.5 Universes with $\Lambda = 0$

If $\Lambda = 0$, the fate of the universe depends on its curvature. The first Friedmann equation (6) can be written as

$$\dot{a}^2 = \frac{8\pi G \rho a^2}{3} - kc^2. \tag{12}$$

If $k = -1$, then, since the first term on the right is always nonnegative, we find that $|\dot{a}| \geq c$. So the universe either begins in a singularity and expands without bound, or constantly contracts before ending in a "big crunch". (The latter is simply the time-reversal of the former.) In the expanding case, ρa^2 will approach 0, so $\dot{a}$ approaches the constant c as t increases.

If $k = 0$, then either the universe is empty and a is constant (this is just the familiar Minkowski space), or $\rho > 0$ and $\dot{a}$ is never zero. In the latter case, the universe must either be always expanding or always contracting. Focusing again on the expanding case, we can rewrite equation (12) as

$$\dot{a} = \sqrt{\frac{8\pi G\rho a^2}{3}}.$$

To determine the asymptotic behavior of a, we consider the two extremes where the universe is entirely full of matter or entirely full of radiation. In these situations, the equation becomes

$$\dot{a} = Pa^d,$$

where P is some constant and d is -0.5 for matter or -1 for radiation. The solution to this separable differential equation is

$$a(t) = (P(1-d)(t+C))^{\frac{1}{1-d}}.$$

In other words, a is proportional to either $t^{2/3}$ or $t^{1/2}$, after shifting the time scale by some constant C. So the universe expands without bound, but at a rate that approaches 0, as opposed to the c we saw in the $k=-1$ case.

If $k=1$, then the universe must contain matter; otherwise, equation (12) would give a negative value for $\dot{a}^2$. More precisely, ρa^2 must be at least $3c^2/8\pi G$. The dynamics prevent this inequality from being violated: if ρa^2 is ever equal to this critical value, then $\dot{a}=0$; by the second Friedmann equation (7), $\ddot{a}$ is negative, so $\dot{a}$ will become negative, causing a to decrease and ρa^2 to increase. In fact, the universe is destined for a big crunch after this point, because $\dot{a}$ can never increase again.

If the universe is initially expanding, then ρa^2 must eventually reach the critical value. The only other possibility would be for a to approach some limit without ever reaching it, but then $\ddot{a}$ would approach 0 without ρ approaching 0, violating the second Friedmann equation. So the universe always expands until ρa^2 reaches the critical value, then begins to contract, before finally ending in a big crunch.

In this section, we have seen that an FLRW universe with zero cosmological constant cannot be static, unless it is the empty Minkowski space. This fact is what led Einstein to introduce the cosmological constant in the first place. At the time, it was not known that our universe is expanding, and the cosmological constant was a way of recovering a static universe from general relativity.[14] After work by Edwin Hubble convinced many (including Einstein) that the universe is expanding, it appeared that the cosmological constant was a mistake, and that the non-static universe predicted by Einstein's original equations was an advantage of general relativity rather than a problem to be solved. It was discovered much later (in 1998) that the expansion of the universe is accelerating, which, as we have seen, is not possible with zero cosmological constant.

[14] The static universe proposed by Einstein, with positive cosmological constant, will appear in the next section.

This led to the reintroduction of the cosmological constant, with a small but positive value.

3.6 Universes with $\Lambda > 0$

If $\Lambda > 0$ and $k \neq 1$, then, according to the first Friedmann equation (6),

$$|\dot{a}| \geq a\sqrt{\frac{\Lambda c^2}{3}}.$$

In the expanding case, we find that $\dot{a}$ is greater than some minimum value that does not decrease. It follows that the universe expands without bound. As it does so, ρ and k/a^2 both approach 0. So the right hand side of the Friedmann equation is eventually dominated by the Λ term, giving us

$$\dot{a} \approx a\sqrt{\frac{\Lambda c^2}{3}}, \tag{13}$$

which integrates to

$$a(t) = C\exp\left(t\sqrt{\frac{\Lambda c^2}{3}}\right).$$

So, even if the universe does not expand exponentially at first, it will eventually expand exponentially once ρ and k/a^2 become negligible. (In the flat and empty case $\rho, k = 0$, the equation above is exact for all time. This solution is known as De Sitter space.)

If $k = 1$, then the situation is somewhat different, as the first Friedmann equation contains both positive and negative terms. Explicitly, the equation is

$$\dot{a}^2 = \frac{8\pi G}{3}\rho a^2 + \frac{\Lambda c^2}{3}a^2 - c^2. \tag{14}$$

There is nothing preventing the right hand side of this equation from being 0 at some point in time. In fact, as Einstein discovered, it is possible to make a, ρ, and q constant by finely tuning the parameters. If we set the right hand sides of equations (14) and (7) to 0 (making $\dot{a}$ and $\ddot{a}$ both 0) and write $q = wc^2\rho$ for some constant w, we can solve for Λ and ρ to obtain

$$\Lambda = \frac{3w+1}{w+1}a^{-2} \qquad\qquad \rho = \frac{c^2}{4\pi G(w+1)}a^{-2}.$$

(Special cases are $w = 0$ for a static universe full of dust-like matter, and $w = 1/3$ for a static universe full of radiation.) Note that if a (the radius of the 3-spherical universe) is higher, the values of Λ and ρ necessary to maintain a static universe are lower, decreasing proportionally to a^{-2}.

However, Einstein's static universe is a very special case. Other universes with $\Lambda > 0$ and $k = 1$ tend to either collapse in a singularity (as in the $\Lambda \leq 0$, $k = 1$

case) or expand exponentially (as in the $\Lambda > 0$, $k \neq 1$ case). If we make the simplifying assumption that $q = 0$, then

$$\rho = \frac{M}{2\pi^2} a^{-3}$$

for some constant M. (I chose to add a factor of $2\pi^2$ because it makes the total mass of the universe equal to M. Since the volume of a 3-sphere with radius a is $2\pi^2 a^3$, the total mass is $2\pi^2 a^3 \rho = M$.) With this substitution, equation (14) becomes

$$\dot{a} = \pm\sqrt{\frac{4GM}{3\pi} a^{-1} + \frac{\Lambda c^2}{3} a^2 - c^2},$$

which is a time-independent ordinary differential equation. Writing

$$P(a) = \frac{4GM}{3\pi} + \frac{\Lambda c^2}{3} a^3 - c^2 a,$$

we know that $\dot{a}$ is nonzero when $P(a)$ is positive, zero when $P(a)$ is zero, and undefined when $P(a)$ is negative (i.e. such universes cannot exist). Note that P is a cubic polynomial with $P(0) \geq 0$ (equalling 0 only when the universe is empty), $P'(0) < 0$, and $P(a) \to \infty$ as $a \to \infty$. P thus has exactly one local minimum, at $a_{\mathrm{m}} > 0$, and the sign of $P(a_{\mathrm{m}})$ is of great importance to the dynamics of the universe. By setting $P' = 0$ one finds

$$a_{\mathrm{m}} = \frac{1}{\sqrt{\Lambda}} \qquad P(a_{\mathrm{m}}) = \frac{4GM}{3\pi} - \frac{2c^2}{3\sqrt{\Lambda}}$$

the condition $P(a_{\mathrm{m}}) = 0$ can be rewritten in the dimensionless form

$$\frac{2GM\sqrt{\Lambda}}{\pi c^2} = 1.$$

So, defining $f(M, \Lambda)$ to be the value on the left, there are four possibilities:

- $f(M, \Lambda) > 1$. In this case, $P(a)$ is positive for all $a \geq 0$ and the universe expands without bound. Eventually, the curvature will be negligible and the expansion will be exponential as in the $k \neq 1$ case. (Of course, the time-reversed version, with the universe ending in a singularity, is also mathematically possible.)
- $0 < f(M, \Lambda) < 1$. In this case, $P(a)$ has two positive roots, and there are two possibilites: a is always less than or equal to the lower root, in which case the universe begins in a big bang, reaches the lower root, and then ends in a big crunch, or a is always greater than or equal to the higher root, in which case the universe contracts, "bounces" at the higher root, and then expands without bound. Note that a cannot approach one of the roots without ever reaching it, or reach it and then stay there, because by the second Friedmann equation, $\ddot{a}$ is only 0 at a single value of a, which lies in the unreachable region strictly between the two roots.

- $M = 0$ and thus $f(M, \Lambda) = 0$. This case is similar to the previous one, except that the lower root of P is at $a = 0$, so the big bang-big crunch scenario is impossible, leaving only the bouncing scenario. Since $\rho = 0$, equation (14) can actually be integrated to find that $a(t)$ is a scaled and shifted cosh function. This universe — as well as its analogue with $k = -1$, in which $a(t)$ is a sinh function — is actually isometric to the De Sitter space with the same cosmological constant [6, § IX.10]. In other words, for a fixed positive cosmological constant, the empty FLRW universes with $k = -1$, $k = 0$, and $k = +1$ are just different coordinate systems on the same spacetime. Since the De Sitter universe is empty, there is natural notion of whether something is "comoving" and thus no single comoving coordinate system.

- $f(M, \Lambda) = 1$. In this case, $P(a)$ has a single root at $a = \Lambda^{-1/2}$. a can stay at this root for all time, resulting in the Einstein static universe. To understand the behavior for other initial values of a, we can linearize the differential equation

  $$\dot{a} = \pm\sqrt{\frac{P(a)}{a}} = \pm\sqrt{\frac{2c^2\Lambda^{-1/2}}{3}a^{-1} + \frac{c^2\Lambda}{3}a^2 - c^2}$$

  at the root $a = \Lambda^{-1/2}$. By applying L'Hopital's rule twice, we find that

  $$\lim_{a \to \Lambda^{-1/2}} \left(\frac{\mathrm{d}}{\mathrm{d}a} \sqrt{\frac{P(a)}{a}} \right)^2 = c^2\Lambda,$$

  so the linearization is

  $$\dot{a} = \pm c\Lambda^{1/2}(a - \Lambda^{-1/2}),$$

  with solution

  $$a(t) = \Lambda^{-1/2} + C \exp\left(\pm c\Lambda^{1/2} t \right).$$

  So the universe approaches Einstein's static universe (either as $t \to -\infty$ or as $t \to +\infty$) at an approximately exponential rate. Essentially, there are four possibilities: the universe can begin in a big bang and approach Einstein's universe, begin arbitrary close to Einstein's universe and then contract, ending in a big crunch, contract forever while approaching Einstein's universe, or begin arbitrarily close to Einstein's universe and then expand without bound. The last of these was proposed by Eddington, early in the history of expanding-universe theories, as a model for our own universe [2].

Although I have been assuming that $q = 0$, the case of a $\Lambda > 0$, $k = 1$ universe full of radiation ($w = \frac{1}{3}$) can be analyzed in much the same way. The primary difference is that the mass of the universe is no longer constant; you can instead use a constant equal to the mass times a.

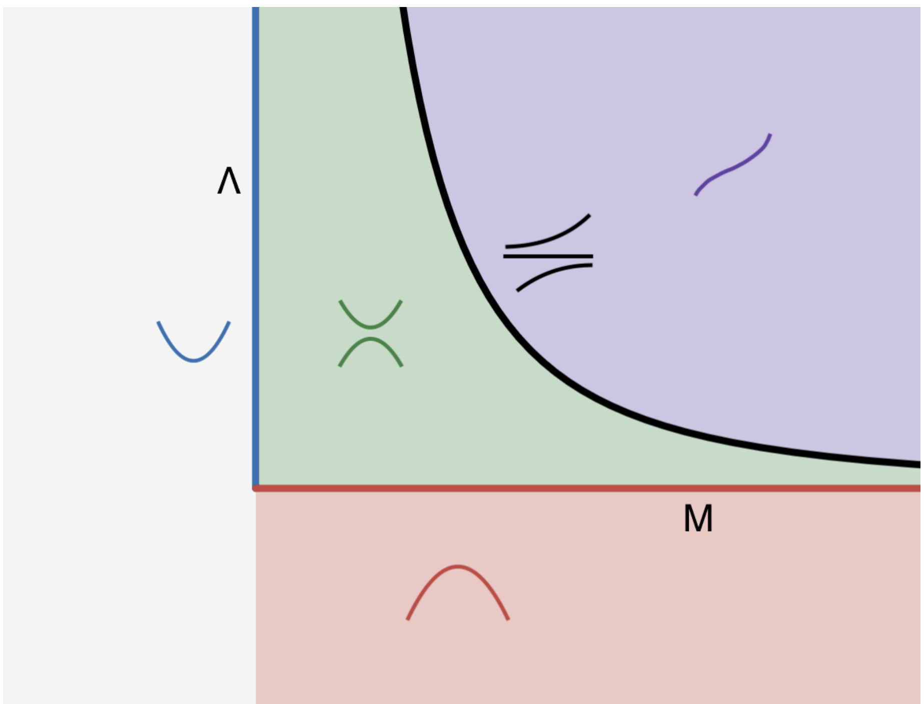

Figure 8: Summary of 3-spherical FLRW universes with $q = 0$. The horizontal axis is the constant mass M of the universe, and the vertical axis is the cosmological constant Λ. Universes with $M = 0, \Lambda \leq 0$ are invalid. The small color-coded graphs show how the radius $a(t)$ changes over time; the expanding universes in the purple and black regions have contracting analogues that are not shown.

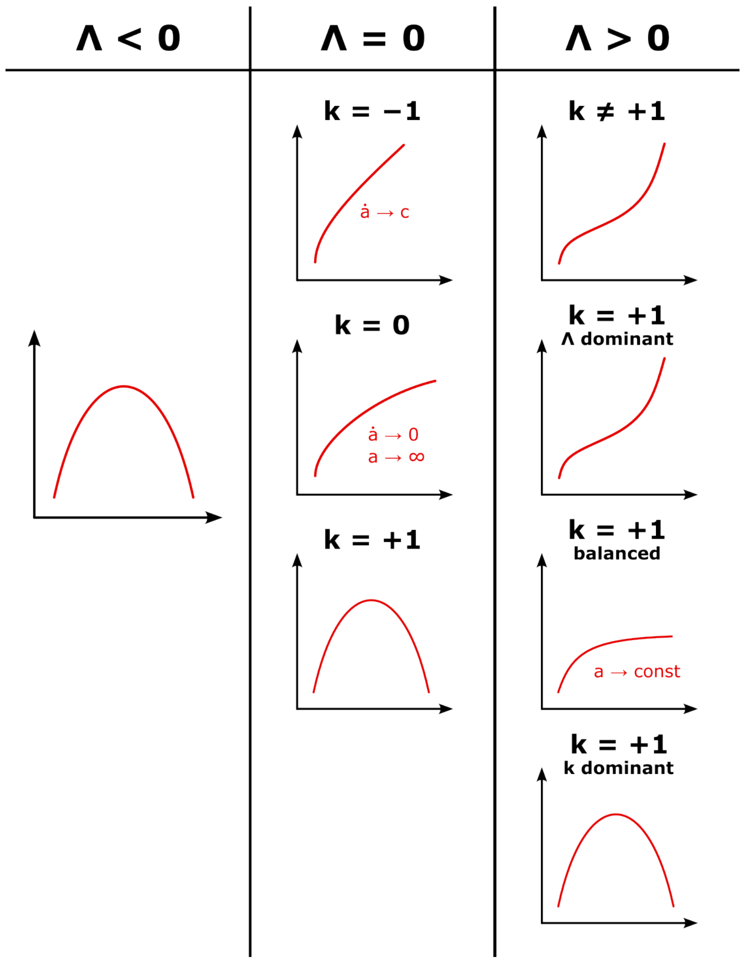

Figure 9: Summary of big bang models.

3.7 Redundance of Λ

Recall the Einstein field equation:

$$\text{Ric} - \frac{1}{2}Rg + \Lambda g = \frac{8\pi G}{c^4}T.$$

This can be rewritten as

$$\text{Ric} - \frac{1}{2}Rg = \frac{8\pi G}{c^4}\left(T - \frac{\Lambda c^4}{8\pi G}g\right).$$

In other words, adding a cosmological constant is equivalent to adjusting the stress-energy tensor by a term proportional to g. For an FLRW universe, where the stress-energy tensor is given by equation (5), this means replacing ρ with

$$\rho + \frac{\Lambda c^2}{8\pi G}$$

and q with

$$q - \frac{\Lambda c^4}{8\pi G}.$$

Indeed, it is easy to check that performing this substitution while removing the separate Λ terms preserves the Friedmann equations. So Λ can be treated either as a parameter separate from ρ and q, or as a substance which contributes to ρ and q and whose pressure is the negative of its energy density.

Interestingly, spatial curvature has the same effect on the expansion of the universe as a hypothetical substance with $w = -\frac{1}{3}$ (that is, $q = -\frac{1}{3}c^2\rho$) and whose density is therefore proportional to a^{-2}. Specifically, adding a term to ρ equal to

$$\frac{-3kc^2}{8\pi G a^2}$$

and a term to q equal to

$$\frac{kc^4}{8\pi G a^2},$$

while removing the separate k term in the first Friedmann equation, leaves the Friedmann equations unchanged. This means that spatial curvature can be said to have an "equivalent density" and an "equivalent pressure", with positive curvature having negative equivalent density. Curvature is physically distinguishable from its equivalent substance, however, as it appears in the metric tensor g.

3.8 Our Universe

Observations indicate that the universe is expanding at an accelerating rate (that is, $\ddot{a} > 0$), which suggests a positive cosmological constant: a universe with $\Lambda \leq 0$ and $\rho, q \geq 0$ cannot exhibit accelerating expansion by the second

Friedmann equation (7). The results of Section 3.6 suggest that the universe will continue to expand forever at a rate that is asymptotically exponential, a fate known as the "Big Freeze". The estimated value for Λ, based on estimates of $\ddot{a}$ as well as the universe's density, is on the order of $\Lambda = 10^{-52}\text{m}^{-2}$. Since Λ can be viewed as contributing a density and pressure (Section 3.7), we can compare its density to that of ordinary matter. Currently, about 70% of the universe's total density comes from Λ, with only 30% coming from matter with $q \geq 0$. The vast majority of this 30% is non-relativistic matter with $q \approx 0$; only a tiny proportion is radiation. However, since the density of ordinary matter decreases as the universe expands while Λ does not, these proportions were different in the past. For the first 50,000 years or so of the universe's existence, radiation dominated, and for a period of billions of years after that, non-relativistic matter dominated. Around 4 billion years ago (or 10 billion years after the Big Bang), the density contributed by Λ reached 1/3 of the total, causing the universe's expansion to stop decelerating and begin to accelerate.

The acceleration of the universe's expansion can be explained by factors other than a cosmological constant. Recalling that Λ can be viewed as a substance with $w = -1$, we can also imagine substances with w close to -1 or with a w that changes over time, being close to -1 in the present day but greater in the past. Such models are known as "quintessence", and are not included in the classification in sections 3.4–3.6, in which it was assumed that $\rho, q \geq 0$ (with Λ treated as a separate parameter). Note that if $w < -1/3$, quintessence contributes to acceleration, while if $w > -1/3$, it contributes to deceleration. In general, a substance that causes the expansion of the universe to accelerate is known as "dark energy", with Λ and the more general quintessence model being two types of dark energy.

Some cosmologists have argued that dark energy does not exist at all, and that the inhomogeneity of the universe — which is ignored in the FLRW model — is large enough to create the appearance of expansion acceleration when there is none. The "timescape cosmology" proposed by David Wiltshire takes into account general-relativistic time dilation, with time passing slower in galaxies than in cosmic voids (supposedly as much as 35% slower) [5].

Nevertheless, the most widely-accepted cosmological model (as of 2025) is the "Λ-CDM" model, which includes a positive cosmological constant as described at the beginning of this section.

As for the curvature type k, measurements are consistent with the universe being perfectly flat ($k = 0$). Of course, it could be the case that $k = \pm 1$ with a very large, since any Riemannian manifold appears arbitrarily flat at small scales. A value of $k = 0$ (or $k = -1$) would seem to suggest that the universe is of infinite volume, with spatial slices homeomorphic to $\mathbb{R}^3$, but this is not actually required, as other 3-dimensional topologies (including some compact ones) can be given a flat metric. For example, the universe could be a 3-torus, repeating periodically in three independent directions. But this is purely speculative, and in the absence of any evidence to the contrary, it is simplest to model the universe as infinite. (Contrary to popular perception, the Big Bang does not imply that the universe is finite.)

If we assume that $k = 0$ and $q = 0$ — the latter of which is a fine approximation for most of the universe's history — then the first Friedmann equation (6) can actually be integrated to get a closed-form expression for the scale factor of the universe as a function of time. (This fact was pointed out to me by Dr. Williams.) Writing $\rho = Da^{-3}$ for some constant D, we get

$$\frac{\mathrm{d}a}{\mathrm{d}t} = \sqrt{\frac{8\pi GD}{3}a^{-1} + \frac{c^2\Lambda}{3}a^2}.$$

Separating and integrating:

$$t = \int \left(\frac{8\pi GD}{3}a^{-1} + \frac{c^2\Lambda}{3}a^2\right)^{-1/2} \mathrm{d}a = \int \left(\frac{8\pi GD}{3} + \frac{c^2\Lambda}{3}a^3\right)^{-1/2} a^{1/2}\,\mathrm{d}a.$$

We can get rid of the $a^{1/2}$ by substituting $u = a^{3/2}$, so that $a^{1/2}\,\mathrm{d}a = \frac{2}{3}\,\mathrm{d}u$. The result is

$$t = \frac{2}{3}\int \left(\frac{8\pi GD}{3} + \frac{c^2\Lambda}{3}u^2\right)^{-1/2} \mathrm{d}u = \frac{1}{\sqrt{6\pi GD}}\int \left(1 + \frac{c^2\Lambda}{8\pi GD}u^2\right)^{-1/2} \mathrm{d}u.$$

Recalling that the derivative of $\sinh^{-1}(u)$ is $(1+u^2)^{-1/2}$, we can write the integral as

$$\begin{aligned} t &= \frac{1}{\sqrt{6\pi GD}}\sqrt{\frac{8\pi GD}{c^2\Lambda}}\sinh^{-1}\left(u\sqrt{\frac{c^2\Lambda}{8\pi GD}}\right) \\ &= \frac{2}{\sqrt{3c^2\Lambda}}\sinh^{-1}\left(a^{3/2}\sqrt{\frac{c^2\Lambda}{8\pi GD}}\right). \end{aligned}$$

(I am leaving out the constant of integration, because it is not needed if we use the convention that $t = 0$ is the time when $a = 0$.) Solving for a, we get

$$a(t) = \left(\frac{8\pi GD}{c^2\Lambda}\right)^{1/3}\sinh\left(t\frac{\sqrt{3c^2\Lambda}}{2}\right)^{2/3}. \tag{15}$$

Essentially, the universe grows something like $\sinh(t)^{2/3}$.[15] If we'd like, we can write $D = \rho_0 a_0^3$, where ρ_0 and a_0 are the universe's present density and scale factor respectively.[16] Then equation (15) becomes

$$\frac{a(t)}{a_0} = \left(\frac{8\pi G\rho_0}{c^2\Lambda}\right)^{1/3}\sinh\left(t\frac{\sqrt{3c^2\Lambda}}{2}\right)^{2/3},$$

which gives the scale factor of the universe at any time t (with $t = 0$ being the big bang) relative to that at the present time.

[15] In the early period of radiation dominance, it is more like $t^{1/2}$, by the results of Section 3.5.

[16] ρ_0 is approximately $2.7 \times 10^{-27}\,\mathrm{kg\,m^{-3}}$, most of which is contributed by invisible "dark matter."

References

[1] Neil Ashby. "Relativity in the Global Positioning System". In: *Living Reviews in Relativity* 6.1 (Jan. 2003). DOI: 10.12942/lrr-2003-1. URL: https://doi.org/10.12942/lrr-2003-1.

[2] A. S. Eddington. "On the Instability of Einstein's Spherical World". In: *Monthly Notices of the Royal Astronomical Society* 90.7 (May 1930), pp. 668–678. ISSN: 0035-8711. DOI: 10.1093/mnras/90.7.668. eprint: https://academic.oup.com/mnras/article-pdf/90/7/668/2901975/mnras90-0668.pdf. URL: https://doi.org/10.1093/mnras/90.7.668.

[3] Matthias Kreier. *Relative time for GPS satellites*. URL: https://kreier.github.io/gps-relativity/. (accessed: 2025-08-26).

[4] Robert J. A. Lambourne. *Relativity, Gravitation, and Cosmology*. Cambridge University Press, 2010. ISBN: 9780521761192.

[5] Antonia Seifert et al. "Supernovae evidence for foundational change to cosmological models". In: *Monthly Notices of the Royal Astronomical Society: Letters* 537.1 (Dec. 2024), pp. L55–L60. ISSN: 1745-3925. DOI: 10.1093/mnrasl/slae112. eprint: https://academic.oup.com/mnrasl/article-pdf/537/1/L55/61215381/slae112.pdf. URL: https://doi.org/10.1093/mnrasl/slae112.

[6] Anthony Zee. *Einstein Gravity in a Nutshell*. Princeton University Press, 2013. ISBN: 9780691145587.