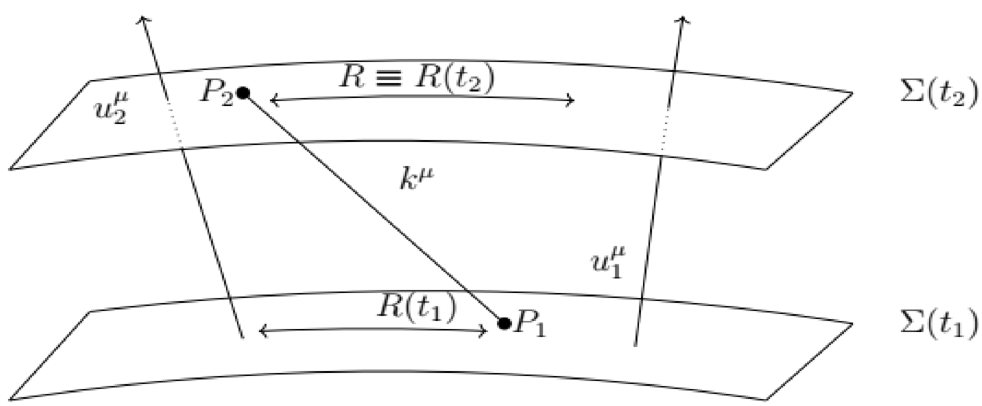
Redshift in cosmological spacetimes

A simple, but very important effect in an expanding universe is the phenomenon of redshift. Let us consider a light ray, following an affinely parameterized null geodesic with 'wave vector' k^a. The frequency ω measured by

a comoving (with the "Hubble flow" described by u^a) observer following the curve γ with velocity $u^a = \dot{\gamma}^a = (\partial/\partial t)^a$ is

$$\omega = -k_a u^a.$$

We now wish to compute the change in frequency when two observers measure the same light ray.



This calculation is aided by the presence of Killing vector fields in our examples. If ξ^a is a spacelike Killing vector field, then one can show that $k_a\xi^a = f$ is constant along the light ray (exercises). Assuming that k^a has a projection into Σ tangent to the Killing field ξ^a, we find (using the shorthand $|\xi|^2 = \xi^a \xi_a$)

$$\begin{aligned} \frac{a(t_1)^2}{a(t_2)^2} &= \frac{|\xi|^2_{p_1}}{|\xi|^2_{p_2}} \\ k_a u_1^a &= -\frac{k_a \xi^a}{|\xi|}\bigg|_{p_1} \\ k_a u_2^a &= -\frac{k_a \xi^a}{|\xi|}\bigg|_{p2} \\ \Rightarrow \quad \frac{\omega_1}{\omega_2} &= \frac{a(t_2)}{a(t_1)}. \end{aligned}$$

The redshift factor is

$$\begin{aligned} z &= \frac{\lambda_2 - \lambda_1}{\lambda_1} = \frac{\omega_1}{\omega_2} - 1 = \frac{a_2}{a_1} - 1 \\ &\approx HR \end{aligned} \tag{34}$$

where R is the distance which is $\approx t_2 - t_1$ for nearby galaxies. For a flat universe ($k = 0$) this is

$$R = a(t)\sqrt{x^2 + y^2 + z^2} = \text{ geodesic distance.}$$

Another way to state (34) is

$$\frac{a_1}{a_2} = \frac{a(\text{emitted})}{a(\text{observed})} = \frac{1}{1+z}.$$

The 'physical'= geodesic distance R on a given slice between two points is of little empirical interest, since we can only observe objects in our past lightcone. A more useful notion of distance is the "luminosity distance", d_L, defined as

$$d_L^2 = \frac{L}{4\pi F}. \tag{35}$$

Here L is the absolute luminosity and F is the flux seen by the observer. For Minkowski space, $d_L = R$ since in that case the area of a sphere of the physical radius R is $4\pi R^2 = A$, that is $\frac{F}{L} = \frac{1}{A}$ in a flat space ($a(t) \equiv 1, k = 0$). On the other hand, in an expanding ($k = 0$) Friedmann-Lemaître-Robertson-Walker universe, we have instead

$$\frac{F}{L} = \frac{1}{(1+z)^2 A}$$

since the photons arriving at t_2 are $(1+z)^{-1}$ times less energetic ($E_{\text{ph}} = h\nu = \hbar\omega$) and the rate of emission also goes down by the same factor. Thus,

$$d_L = (1+z)R.$$

Because $(k^\mu) = (1, \dot{R}, 0, 0)$ is null we must have $1 = a^2\dot{R}^2$ and hence $\dot{R} = \frac{1}{a}$. Then,

$$\begin{aligned} R = R(t_2) &= a(t)\sqrt{x^2 + y^2 + z^2} \\ &= \int_{t_1}^{t_2} \dot{R}(t)dt \\ &= \int_{t_1}^{t_2} \frac{dt}{a(t)} \\ &= \int_{a_1}^{a_2} \frac{da}{a^2 H(a)}, \end{aligned}$$

or, with $dz = \frac{-da}{a^2}$

$$\begin{aligned} R &= \int_{z_2}^{z_1} \frac{dz}{H(z)} \\ &= \int_0^z \frac{dz'}{H(z')} \\ \Rightarrow \quad d_L &= (1+z) \int_0^z \frac{dz'}{H(z')}. \end{aligned}$$

It is common to use the first Friedmann equation – or rather its obvious generalization to several species of "particles" – in order to replace H^2 by

$$H^2 = \frac{8\pi G_N}{3} \sum_{\text{species } i} \rho_i.$$

Assuming that each species evolves according to a power law, we find

$$\rho_i = \rho_{i,\text{today}} a^{-n_i}$$

where $\rho_{i,\text{today}}$ is the matter density at t_2 (today). Assuming without loss of generality that $a_2 = 1$ we find

$$\begin{aligned} H(z) &= \sqrt{\frac{8\pi G_N}{3}} \left(\sum_i \rho_{i,\text{today}} (1+z)^{n_i} \right)^{\frac{1}{2}} \\ \Rightarrow \quad \frac{H(z)}{H_{\text{today}}} &= \left(\sum_i \Omega_{i,\text{today}} (1+z)^{n_i} \right)^{\frac{1}{2}} \end{aligned}$$

where $\Omega_i(t) \equiv \frac{8\pi G_N}{3} \frac{\rho_i(t)}{H(t)^2}$. Then we get

$$\boxed{d_L(z) = \frac{1+z}{H_{\text{today}}} \int_0^z \frac{dz'}{\left(\sum_i \Omega_{i,\text{today}} (1+z')^{n_i}\right)^{\frac{1}{2}}}}$$

In practice, we measure $d_L(z)$ for large redshifts z and extract $\Omega_{i,\text{today}}$ and H_{today}. For that, we need objects with known intrinsic luminosity L, such as type IIa supernovae.

Particle Horizons

Another important feature of expanding universes is the possible existence of "particle horizons". This is most easily demonstrated for flat universes, $k = 0$, where the metric is

$$g = -dt^2 + a(t)^2(dx^2 + dy^2 + dz^2). \tag{36}$$

Let us introduce again the *conformal time parameter* η by

$$\eta = \int_{t_0}^{t} \frac{dt'}{a(t')} \qquad \left(\frac{d\eta}{dt} = \frac{1}{a(t)} \Leftrightarrow a d\eta = dt\right) \ . \tag{37}$$

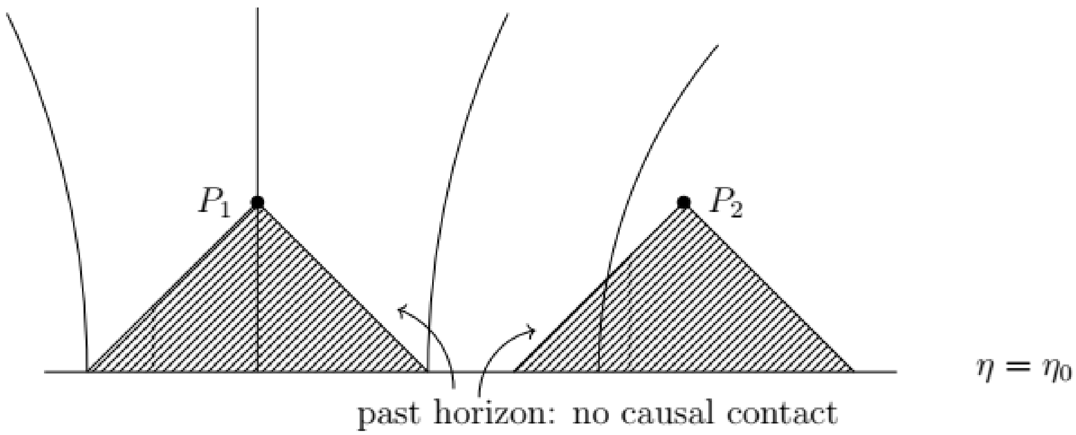
from which it follows that

$$ds^2 = a(\eta)^2 \Big\{ \underbrace{-d\eta^2 + dx^2 + dy^2 + dz^2}_{\text{Minkowski-space}} \Big\} \tag{38}$$

We see that writing the metric in terms of η has the advantage that its relationship with Minkowski space becomes manifes: It is "conformal" to Minkowski spacetime, or possible a subset thereof.

Due to the conformal factor $a(\eta)^2$, geodesics in Minkowski space in general do not in general coincide with the geodesics of g_{ab}. However, the conformal factor preserves the causal character of a curve, and therefore the causal relationships in spacetime. In particular, we may ask whether it is possible for two points p, p' in spacetime to be such that their causal pasts (i.e. the set of all points that can be reached by past directed timelike or null curves) are disjoint. It is clear form the following picture that this will be the case if and only if the parameter η has a finite range for negative values, which in turn will be the case if and only if $\int_t^{t_1} \frac{dt'}{a(t')}$ converges to a finite value for $t \to t_0$.

Whether or not this is the case therefore depends on the behavior of $a(t)$ near $t = t_0$, which is in turn determined by the equation of state. Indeed, recall that for $P = w\rho$, we had $a(t) \propto t^{\frac{2}{3(w+1)}}$ (choosing $t_0 = 0$) so the integral is finite in particular for all $w \geq 0$ (e.g. dust, $a \propto t^{2/3}$, or radiation, $a \propto t^{1/2}$.) If there are points p, p' with disjoint causal past, then a particle at p could never have been in contact with a particle at p' – one says that "there are particle horizons". Thus, we get regions in spacetime which are causally disjoint, and for this reason, will not have had the opportunity to equilibrate with each other. Current observations seem to exclude the



presence of such horizons, meaning for instance that the scale factor $a(t)$ could not have behaved like that of radiation or dust all the way to the big bang ($t = 0$). On the other hand, particle horizons are not present e.g. for an exponential scale factor $a(t) \propto e^{tH}$ because the integral $\int_t^{t_1} dt'/a(t')$ then clearly diverges at the lower end $t \to -\infty$. Therefore such an "inflationary phase" is consistent with the absence of horizons. It is indeed currently believed that our universe underwent such a phase shortly after the big bang.

Black Holes

In the previous subsection, we have obtained solutions to Einstein's equation with a non-zero stress tensor representing varous types of fluid matter. However, in General Relativity, one can have interesting non-trivial solutions even for a vanishing stress tensor. Such solutions are called "vacuum solutions" and obey

$$R_{ab} = 0 \ . \tag{39}$$

In particular, unlike in the case of Newtonian Gravity [8], one can obtain *static* vacuum solutions. It turns out that these solutions describe objects that have properties unlike any other objects known before: black holes.

[8] In Newtonian Gravity, the only globally defined solution to Poisson's equation with vanishing ρ which is decaying at spatial infinity is the trivial one. The possibility of non-trivial static vacuum solutions in General Relativity can be ascribed to the fact that Einstein equations are non-linear, meaning that gravity can act as its own source in a sense.

Derivation of the Schwarzschild Solution

Given the complexity of Einstein's equations, it is somewhat surprising that this family of static solutions, known as the "Schwarzschild solution", is actually rather easy to derive. To get started one assumes, as seems evidently reasonable, that

1. g_{ab} has as its isometry group the group of rotations $SO(3)$ and, of course, that
2. g_{ab} is static. This means that there are 'time-shift' isometries whose orbits are orthogonal to a spacelike surface Σ.

[The first assumption can actually be shown to be a consequence of the second one – this is known as "Israels theorem".] One can show that time shifts and rotations commute, and that there exists a local coordinate system $(x^\mu) = (t, r, \theta, \phi)$ such that

$$g = -\underbrace{f(r)}_{\text{free functions}} dt^2 + \underbrace{h(r)} dr^2 + \underbrace{r^2\left(d\theta^2 + \sin^2\theta\ d\phi^2\right)}_{\text{metric on } \mathbb{S}^2_r}$$

with a time shift acting by $t \to t+$const and rotations acting on the spherical polar coordinates ϕ and θ in the usual way.

It is not hard, although tedious, to compute the Ricci tensor of this metric in the coordinates $(x^\mu) = (t, r, \theta, \phi)$. The off-diagonal components vanish identically, whereas the diagonal components give the equations

$$0 = \frac{1}{2}\frac{1}{\sqrt{fh}}\left[\frac{f'}{\sqrt{fh}}\right]' + (rfh)^{-1}f' \tag{40}$$

$$0 = -\frac{1}{2}\frac{1}{\sqrt{fh}}\left[\frac{f'}{\sqrt{fh}}\right]' + (rh^2)^{-1}h' \tag{41}$$

$$0 = -\frac{f'}{2fh} + (2h^2)^{-1}h' + \frac{1-\frac{1}{h}}{r} \tag{42}$$

Equations (40) and (41) give

$$\frac{f'}{f} + \frac{h'}{h} = 0 \quad \Leftrightarrow \quad f = \frac{k}{h} \qquad (k > 0 \text{ a constant,})$$

and by rescaling $t \to \sqrt{k}t$ we may set $k = 1$. Then (42) gives

$$\begin{aligned} -f' + \frac{1-f}{r} &= 0 \\ \Leftrightarrow \quad (rf)' &= 1 \\ \Leftrightarrow \quad f(r) &= 1 + \frac{C}{r}, \end{aligned}$$

where C is some constant. The desired vacuum solution is thus:

$$ds^2 = -\left(1 + \frac{C}{r}\right) dt^2 + \left(1 + \frac{C}{r}\right)^{-1} dr^2 + r^2 \left(d\theta^2 + \sin^2\theta \; d\phi^2\right).$$

Before embarking on a more detailed analysis of this metric, we make the following crude observations:

1. As $r \to \infty$

   $$ds^2 \to -dt^2 + \underbrace{dr^2 + r^2 \left(d\theta^2 + \sin^2\theta d\phi^2\right)}_{dx^2+dy^2+dz^2} \quad = \text{Minkowski space}$$

   So the metric is "asymptotically flat".

2. From our discussion surrounding the derivation of Einstein's equations, we have

   $$g_{00} \cong -1 - 2\Phi \tag{43}$$

   where Φ is the Newtonian potential. This leads us to identify the constant C as (restoring the speed of light c temporarily in our formulas)

   $$-\frac{C}{2} \cong \frac{G_N M}{c^2}, \tag{44}$$

   implying in particular that we should take $C \leq 0$. In Newtonian gravity, the radius $r_S = \frac{2G_N M}{c^2}$ is precisely the surface of a spherical object of mass M such that the escape velocity for a particle is equal to the speed of light (Laplace). This crudely suggests that the metric might have something to do with a "black hole", as we will confirm below. With the notation $r_S =$ "Schwarzschild radius,"

   $$ds^2 = -\left(1 - \frac{r_S}{r}\right) dt^2 + \left(1 - \frac{r_S}{r}\right)^{-1} dr^2 + r^2 \left(d\theta^2 + \sin^2\theta \; d\phi^2\right).$$

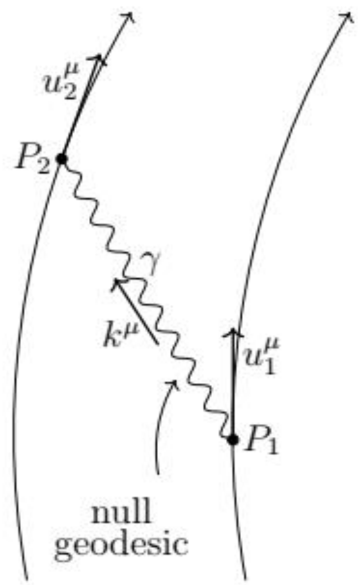
One gets a better intuition about the size of the constants if one writes for instance

$$\begin{aligned} r_S &= \frac{2G_N M}{c^2} \approx 3\left(\frac{M}{M_\odot}\right)\text{km} \qquad \text{with} \\ M_\odot &= \text{mass of the sun} = 2\times 10^{33}\text{g} , \end{aligned}$$

i.e. the sun is vastly bigger than its Schwarzschild radius.

The redshift effect

Consider two 'static' observers, each following a curve of constant r, θ, ϕ, exchanging a light signal. The tangents are denoted u_1^a, u_2^a, respectively.



The locally measured frequencies are (compare the corresponding discussion in cosmological spacetimes) at two points p_1, p_2 are:

$$\begin{aligned} \omega_1 &= -k_a u_1^a\big|_{p1} \\ \omega_2 &= -k_a u_2^a\big|_{p2} \end{aligned}$$

We have $u_1^a u_{1a} = -1 = u_2^a u_{2a}$, and the static observers are tangent to the Killing vector field $\xi^a = \left(\frac{\partial}{\partial t}\right)^a$. Note that $\xi_a k_a$ does not change along the null geodesic representing the signal, since ξ^a is Killing, and since k^a is geodesic.

We may write

$$\begin{aligned} u_1^a &= \frac{\xi^a}{|\xi|}\Big|_{p_1} \\ u_2^a &= \frac{\xi^a}{|\xi|}\Big|_{p_2} \end{aligned}$$

where (setting $G_N = 1 = c$ in the follwing)

$$\begin{aligned} |\xi|^2 &= -g_{ab}\xi^a\xi^b \\ &= 1 - \frac{2M}{r}. \end{aligned}$$

So we find

$$\begin{aligned} \frac{\omega_1}{\omega_2} &= \frac{k_a u_1^a}{k_a u_2^a} = \frac{{}^{(k_a\xi^a)}/_{|\xi|}\Big|_{p_1}}{{}^{(k_a\xi^a)}/_{|\xi|}\Big|_{p_2}} \\ &= \sqrt{\frac{1-\frac{2M}{r_2}}{1-\frac{2M}{r_1}}}\ . \end{aligned}$$

If the emitter (1) is closer to the 'center' than the receiver (2), $r_1 < r_2$, the frequency will decrease ($\omega_2 < \omega_1$) and hence, by $E = \hbar\omega$, the energy of a photon climbing out the gravitational well is decreased. If $\frac{G_N M}{c^2} \ll r_1, r_2$ then

$$\begin{aligned} \frac{\Delta\omega}{\omega} &\approx -\frac{G_N M}{c^2 r_1} + \frac{G_N M}{c^2 r_2} \\ \text{or} \quad \Delta E &= \hbar\Delta\omega \approx \frac{\hbar\omega}{c^2}\left[-\frac{G_N M}{r_1} + \frac{G_N M}{r_2}\right]. \end{aligned}$$

Since a photon with reduced energy is redder according to Einstein's formula $E = h\nu$, the effect is called "gravitational redshift".

Geodesics

To determine the orbits of material particles and lightrays in the Schwarzschild geometry, we need to study geodesics. We denote the geodesic by $\gamma : \mathbb{R} \to M, s \mapsto \gamma(s)$ and by $\dot{\gamma}^a$ the tangent vector to the geodesic. In coordinates $(x^\mu) = (t, r, \theta, \phi)$

$$\frac{d\gamma^\mu}{ds} = \dot{\gamma}^\mu = (\dot{t}, \dot{r}, \dot{\theta}, \dot{\phi})\ .$$

We assume first that the geodesic is timelike, and choose s to be proper time. It follows that there holds

$$g_{ab}\dot{\gamma}^a\dot{\gamma}^b = -1$$

along the curve. (Proof: act with $\dot{\gamma}^c\nabla_c \cong \frac{\partial}{\partial s}$ on $g_{ab}\dot{\gamma}^a\dot{\gamma}^b$ and use $\dot{\gamma}^a\nabla_a\dot{\gamma}^b = 0$.) Additionally, the quantities

$$\left.\begin{array}{l} E = -\dot{\gamma}_a \left(\frac{\partial}{\partial t}\right)^a \\ L = \quad \dot{\gamma}_a \left(\frac{\partial}{\partial \phi}\right)^a \end{array}\right\} \tag{45}$$

are also constant along the curve, i.e. independent of s, because $(\partial/\partial t)^a$ and $(\partial/\partial\phi)^a$ are Killing fields. (This follows e.g. from eq.??, because the line element is independent of t, ϕ.) Furthermore, since $\theta \mapsto \pi - \theta$ is an isometry of Schwarzschild, it is consistent to assume that $\theta(s) = \frac{\pi}{2}$, so $\dot{\theta} = 0$ along the geodesic (exercise). Without loss of generality, we may choose γ to be in such an equatorial plane. E has the interpretation of the energy of the particle, and L that of the angular momentum in the equatorial plane. For null geodesics, we have the same constants of motion, but the normalization condition is now $g_{ab}\dot{\gamma}^a\dot{\gamma}^b = 0$.

Substitution of the constants of motion into the normalisation condition yields, for timelike geodesics,

$$\begin{aligned} \frac{1}{2}\dot{r}^2 + V(r) &= \frac{1}{2}E^2 \\ \text{where} \qquad V(r) &= \frac{1}{2} - \underbrace{\frac{M}{r}}_{\substack{\text{Newtonian} \\ \text{term}}} + \underbrace{\frac{L^2}{2r^2}}_{\substack{\text{Angular} \\ \text{momentum} \\ \text{barrier} \\ \text{(centrifugal} \\ \text{force)}}} - \underbrace{\frac{ML^2}{r^3}}_{\text{new}}. \end{aligned}$$

$V(r)$ can be viewed as the 'effective potential' seen by the geodesic. In the null case, it is given instead by

$$V(r) = \frac{L^2}{2r^2} - \frac{ML^2}{r^3}$$

The radial motion is thus the same as that of a particle in a potential $(V(r)$ in either case, although the form of the potential is different in the null case. Once the radial motion has been determined, the angular motion is found in either case by solving

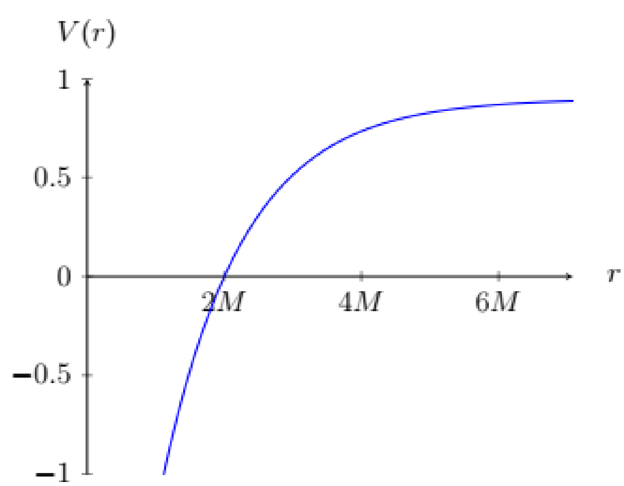
$$L = r^2\dot{\phi} \quad \text{and} \quad E = \left(1 - \frac{2M}{r}\right)\dot{t}.$$

Looking at the potential $V(r)$ in the timelike case, we have the familiar terms from the Kepler problem corresponding to graviational attraction and centrifugal force. In addition to the familiar terms, we have the new term $-\frac{ML^2}{r^3}$, which is of the same sign as the Newtonian term, but wins at small distances r (and is insignificant at large distances). It is thus plausible that the behavior of timelike geodesics will differ from the familiar motion of particles in a central $1/r$-potential for small r.

We are particularly interested in (quasi-) periodic orbits, corresponding to radial oscillations around the minima of the effective potential $V(r)$. To find the extrema of $V(r)$ in the timelike case, we compute

$$\begin{aligned} 0 &= V'(r) = \frac{Mr^2 - L^2 r + 3ML^2}{r^4} \\ \Rightarrow \qquad R_\pm &= \frac{L^2}{2M} \pm \left(\left(\frac{L^2}{2M}\right)^2 - 3L^2\right)^{\frac{1}{2}}. \end{aligned}$$

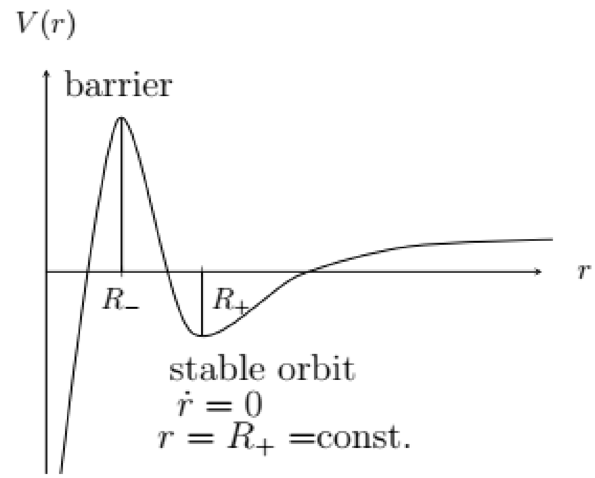
So if $L^2 < 12M^2$, then there are no extrema and the potential $V(r)$ looks like



Hence, there are no stable bound orbits, and a particle having $\dot{r} < 0$ initially will fall right into the "singularity[9]" $r = r_S$", and further into $r = 0$.

[9]We will later clarify the true nature of this, only apparent, "singularity".

On the other hand, if $L^2 \geq 12M^2$, it is easy to check that there is a minimum R_+ and a maximum R_-, and the potential $V(r)$ is as in the following figure.



We conclude that

$$\begin{aligned} R_+ &= r \quad \text{corresponds to a stable circular orbit,} \\ R_- &= r \quad \text{corresponds to an unstable circular orbit.} \end{aligned}$$

For $L \gg M$, the stable orbit is approximately at $R_+ \approx \frac{L^2}{M}$, which gives the Newtonian formula for the orbit of a mass with angular momentum L orbiting a central point mass M. (This is another way of seeing that the identification $C \to -2M$ is physically correct.) Note that the minimum value of R_+ such that stable circular orbit exist is attained when $R_+ = R_- \Rightarrow L^2 = 12M^2 \Rightarrow R_+ \geq 6M$. That shows that no stable circular orbits exist for sufficiently small r-values in General Relativity, because the new term in $V(r)$ wins for $r < 6M$. The energy E of a particle in the stable circular orbit $r = R_+$ is

$$\begin{aligned} \frac{1}{2}E^2 &= V(R_+) = \frac{1}{2} - \frac{M}{R_+} + \frac{L^2}{2R_+^2} - \frac{ML^2}{R_+^3} \\ &= \frac{1}{2}\frac{(R_+ - 2M)^2}{R_+(R_+ - 3M)} \\ V'(R_+) &= 0, \\ \text{so} \qquad E &= \frac{R_+ - 2M}{R_+^{\frac{1}{2}}(R_+ - 3M)^{\frac{1}{2}}} \longrightarrow 1 \text{ as } R_+ \to \infty \end{aligned}$$

Therefore, a particle in an unstable circular orbit in the range $3M < R < 4M$ having an energy which is bigger than that of $E(\infty)$ escapes to infinity. The binding energy $E_B = E(\infty) - E = 1 - E$ for the smallest stable circular orbit ($R_+ = 6M$) is given by

$$E_B = 1 - \sqrt{\frac{8}{9}} \approx 0.06 = 6\%.$$

Due to gravitational radiation (covered later), a body starting in a circular stable orbit will loose some of its energy and therefore gradually decrease r down to $r = R_{\min} = 6M$. The total energy lost (and hence emitted by gravitational radiation) is thus at most about 6% of its total energy. For a body rotating around an ultra-spinning Kerr black hole (a rotating analogoue of the Schwarzschild solution which is beyond the scope of these notes), the ratio is even as big as $\approx 40\%$. Thereby, a substantial portion of energy can be converted into gravitational radiation.

In order to find the oscillations of r around the minimum of $V(r)$, we carry out a Taylor exansion around $r = R_+$,

$$V(r) \approx V(R_+) + \frac{1}{2}\omega_r^2(r - R_+)^2 + \mathcal{O}\left[(r - R_+)^3\right]$$

The "oscillation frequency" in the radial direction is given by

$$\omega_r^2 = V''(R_+) = \frac{M(R_+ - 6M)}{R_+^3(R_+ - 3M)}.$$

On the other hand, the angular frequency of the geodesic is $\omega_\phi = \dot{\phi}$ and therefore

$$\omega_\phi^2 = \frac{L^2}{R_+^4} = \frac{M}{R_+^2(R_+ - 3M)}.$$

Hence, for $R_+ \gg M$ (Newtonian limit), $\omega_\phi \approx \omega_r$ and the particle returns to the original r value after each orbit. In full General Relativity (without taking the Newtonian limit), there is instead a precession of the perhelion with frequency

$$\omega_p = \omega_\phi - \omega_r = -\left(\sqrt{1 - \frac{6M}{R_+}} - 1\right)\omega_\phi.$$

For $R_+ \gg M$ we get to the lowest non-vanishing order

$$\omega_p \approx \frac{3M\omega_\phi}{R_+} \approx \frac{3M}{R_+}\left(\frac{M}{R_+^3}\right)^{\frac{1}{2}} = \frac{3M^{\frac{3}{2}}}{R_+^{\frac{5}{2}}},$$

and restoring c, G_N:

$$\omega_p \approx \frac{3(G_N M)^{\frac{3}{2}}}{c^2 R_+^{\frac{5}{2}}}.$$

For Mercury, this gives (taking also the eccentricity of the orbit into account) $\omega_p(☿) = \frac{43''}{100\text{a}}$. It is surprising – and a lucky coincidence – that this minute precession had been observed around the time General Relativity was conceived. Newtonian calculations, including even the perturbations by the other planets, which could be carried out thanks to the enormous advances in Celestial Mechanics in the end of the 19th century, could not account for this effect, whereas General Relativity could. Hence, the prediction of the perihelion precession of Mercury was historically the first test of General Relativity.

A similar analysis can be carried out for null-geodesics "skimming the surface at the Schwarzschild radius", and leads to the prediction of light bending – another early test of General Relativity (exercises).

Kruskal extension

In order to gain insight into the nature of the apparent singularity of the Schwarzschild metric at $r = r_S$, we next consider radially outgoing null-geodesics, which exist due to the symmetries of the line element. As before, our geodesics are denoted by $(\gamma^\mu(s)) = (t(s), r(s), \theta(s), \phi(s))$, and for radial geodesics $(\dot{\gamma}^\mu) = (\dot{t}, \dot{r}, 0, 0)$. The null-condition leads to

$$\begin{aligned} 0 &= g_{ab}\dot{\gamma}^a\dot{\gamma}^b \\ &= -\left(1 - \frac{2M}{r}\right)\dot{t}^2 + \frac{\dot{r}^2}{1 - \frac{2M}{r}} \end{aligned}$$

or

$$\frac{dt}{dr} = \pm\left(\frac{r}{r - 2M}\right)$$

or

$$\begin{aligned} t &= \pm r_* + \text{const.} \\ \text{with} \qquad r_* &= r + 2M \log\left(\frac{r}{2M} - 1\right) \Rightarrow dr_* = \frac{dr}{1 - \frac{2M}{r}}. \end{aligned}$$

It seems a good idea to try to rewrite the line element in terms of coordinates obtained from the affine parameters along radial null geodesics. Since we have just seen that the latter are defined by the conditions $\pm r_* + t =$ constant and $\phi, \theta =$ constant, the appropriate coordinates are $u = t - r_*, v = t + r_*$ (and θ, ϕ), in terms of which the line element becomes

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dudv + r^2(d\theta^2 + \sin^2\theta \; d\phi^2) \; .$$

Here r is now viewed as a function of u and v; explicitly

$$r + 2M \log\left(\frac{r}{2M} - 1\right) = r_* = \frac{v - u}{2}.$$

We now make further transformations

$$\begin{aligned} U &= -\exp\left(\frac{-u}{4M}\right) \equiv T - X \; , \\ V &= \exp\left(\frac{v}{4M}\right) \equiv T + X \; . \end{aligned}$$

The relationship between the original coordinates (t, r) and the new ones (T, X) is summarized in the following equations

$$\left(\frac{r}{2M} - 1\right) \exp\left(\frac{r}{2M}\right) = X^2 - T^2 \; , \tag{46}$$

$$\frac{t}{2M} = \log\left(\frac{T + X}{X - T}\right) = 2 \, \text{arctanh} \, \frac{T}{X} \; . \tag{47}$$

Changing to the coordinates (T, X) leads to the "Kruskal form" of the line element

$$ds^2 = \frac{32M^3}{r} \exp\left(-\frac{r}{2M}\right)(-dT^2 + dX^2) + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{48}$$

The Kruskal form of the line element shows that $r = r_S$ is not a singularity, because we can clearly extend the metric analytically across this value, at least until $r = 0$. Thus, the true geometry is the analytically extended manifold labelled by the coordinates X, T, and the coordinates on $\mathbb{S}^2$, consistent with $r > 0$. We see:

1. The allowed range of T and X consistent with $r > 0$ is $X^2 - T^2 > -1$. The value $X^2 - T^2 = -1$ corresponds to $r = 0$. This is seen to be a true singularity, e.g. by evaluating the "Kretschmann invariant": $R_{abcd}R^{abcd} \to \infty$.

2. By contrast, $r = 2M$ ($r = r_S$) corresponds to $T = \pm X$, which is not a singularity.

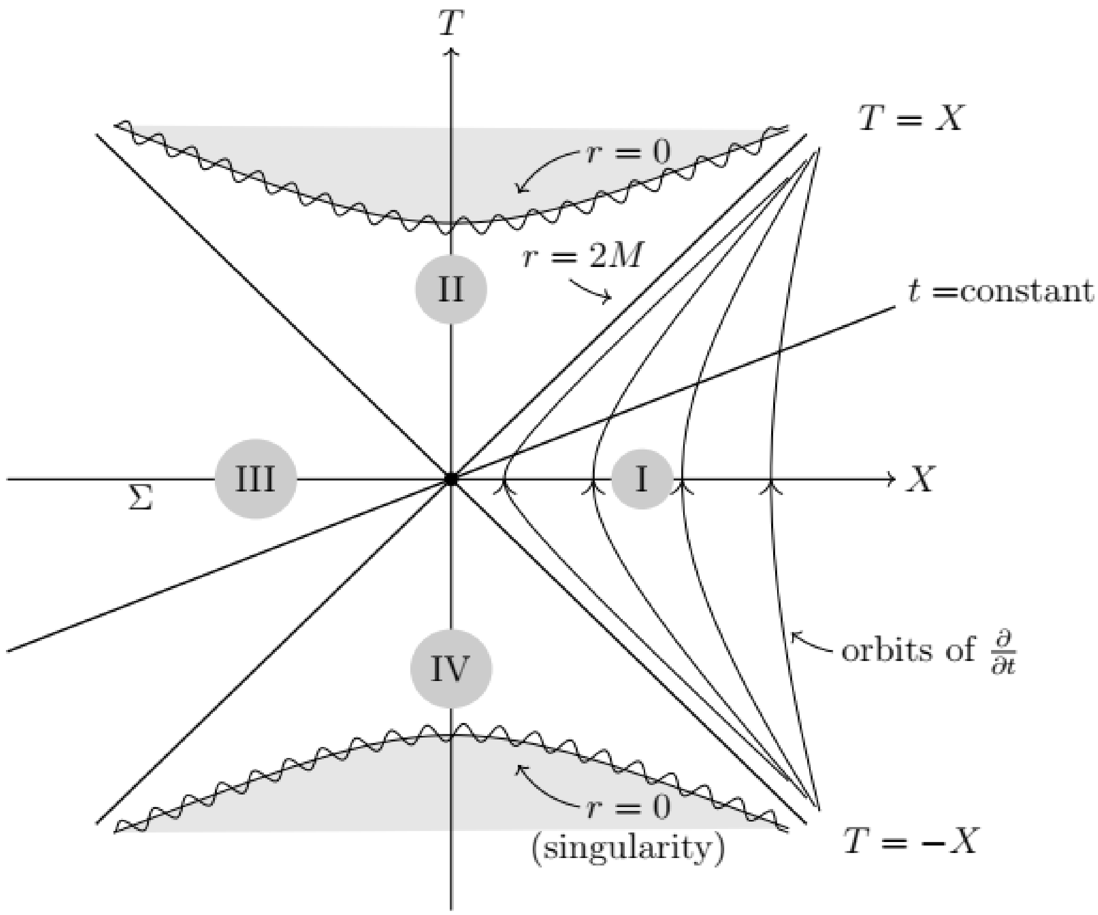
3. The surfaces of constant t corresponds to $\frac{T}{X} = \text{constant}$.

4. At $T = 0 = X$ we have, from equation (46), that

$$dr\Big|_{T=X=0} = \frac{4M}{e}(XdX - TdT)\Big|_{X=T=0} = 0$$

   and that $g^{ab}(dr)_b$ and $\left(\frac{\partial}{\partial t}\right)^a$ (when expressed in X, T-coordinates) become co-linear at $T = \pm X$. It follows that $\frac{\partial(r,t)}{\partial(X,T)}$ becomes singular there, too, showing that (r, t) are 'bad coordinates'. The apparent singularity at $r = r_S$ in the original coordinates is hence due to a bad choice of coordinates.

The causal structure of Schwarzschild following from equation (48) is best illustrated in a diagram in which the (ϕ, θ)-coordinates are suppressed.



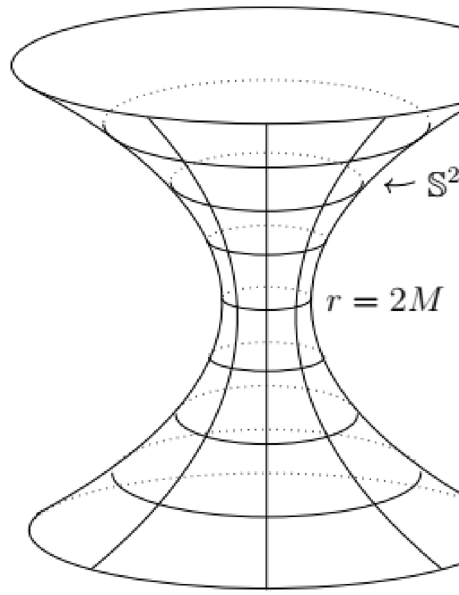
1. Region I (excluding $r = 2M$-lines) corresponds to the original coordinate range $r > 2M$ ('exterior ').

2. Region II has the property that no lightlike-future-directed curves can enter the exterior region I ('black hole').

3. Region III has the property that no lightlike-future-directed curves can stay within that region forever ('white hole').

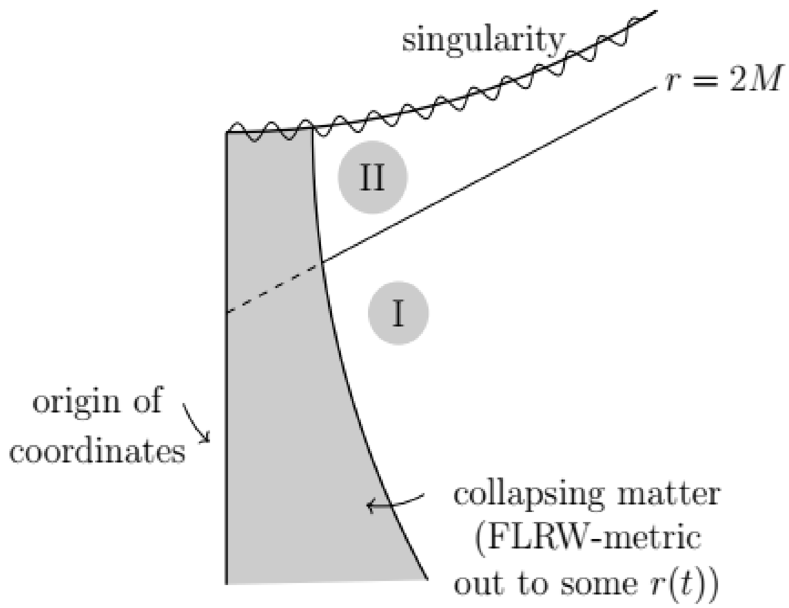
4. Region IV has properties identical to those of region I. If we consider the metric $h = h_{ij}dx^i dx^j$ with $(x^i) = (X, \theta, \phi)$ induced on $\Sigma = \{T = 0\}$, we obtain a Riemannian 3-manifold streching between regions I and IV. Its geometry is illustrated in the following figure (for a $2 + 1$ version of Schwarzschild for the sake of visualization). The embedding into flat space is intended to be such that the induced metric corresponds to

h_{ij}. We hence see the appearance of a "throat" connecting the exterior (region I) with a "parallel universe" (region IV).



5. The black hole/white hole regions are separated from the exterior region by a pair of null surfaces called 'event horizons '. In the old coordinates, these surfaces are both located at $r = r_S$. Thus, rather than being a singularity, the Schwarzschild radius describes the location of the event horizons.

The 'parallel universe' has attracted considerable attention in Science Fiction, but it is unrealistic that it could be formed in the real world. A more realistic spacetime diagram illustrating the formation of a black hole from collapsing matter is as follows. In this diagram (describing a solution suitably patched together from a part of Schwarzschild and a suitable spherically symmetric solution of the Einstein equation with a stress tensor in perfect fluid form), the parallel universe is covered up by the collapsing star.


singularity
$r = 2M$
II
I
origin of coordinates
collapsing matter (FLRW-metric out to some $r(t)$)