Mathematical Tripos Part III

Black Holes

Harvey Reall

Preface

These are lecture notes for the course on Black Holes in Part III of the Cambridge Mathematical Tripos.

Acknowledgment

I am grateful to Andrius Štikonas and Josh Kirklin for producing most of the figures.

Conventions

We will use units such that the speed of light is $c = 1$ and Newton's constant is $G = 1$. This implies that length, time and mass have the same units.

The metric signature is $(- + ++)$

The cosmological constant is so small that is is important only on the largest length scales, i.e., in cosmology. We will assume $\Lambda = 0$ in this course.

We will use *abstract index notation.* Greek indices $\mu, \nu, \ldots$ refer to tensor components with respect to some basis. Such indices take values from 0 to 3. An equation written with such indices is valid only in a particular basis. Spacetime coordinates are denoted x^μ. *Abstract indices* are Latin indices $a, b, c \ldots$. These are used to denote tensor equations, i.e., equations valid in any basis. Any object carrying abstract indices must be a tensor of the type indicates by its indices e.g. $X^a{}_b$ is a tensor of type $(1, 1)$. Any equation written with abstract indices can be written out in a basis by replacing Latin indices with Greek ones ($a \to \mu$, $b \to \nu$ etc). Conversely, if an equation written with Greek indices is valid in *any* basis then Greek indices can be replaced with Latin ones.

For example: $\Gamma^\mu_{\nu\rho} = \frac{1}{2} g^{\mu\sigma} \left(g_{\sigma\nu,\rho} + g_{\sigma\rho,\nu} - g_{\nu\rho,\sigma} \right)$ is valid only in a *coordinate* basis. Hence we cannot write it using abstract indices. But $R = g^{ab} R_{ab}$ is a tensor equation so we can use abstract indices.

Riemann tensor: $R(X, Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z$.

Bibliography

1. N.D. Birrell and P.C.W. Davies, *Quantum fields in curved space*, Cambridge University Press, 1982.

2. *Spacetime and Geometry*, S.M. Carroll, Addison Wesley, 2004.

3. V.P. Frolov and I.D. Novikov, *Black holes physics*, Kluwer, 1998.

4. S.W. Hawking and G.F.R. Ellis, *The large scale structure of space-time*, Cambridge University Press, 1973.

5. R.M. Wald, *General relativity*, University of Chicago Press, 1984.

6. R.M. Wald, *Quantum field theory in curved spacetime and black hole thermodynamics*, University of Chicago Press, 1994.

Most of this course concerns classical aspects of black hole physics. The books that I found most useful in preparing this part of the course are Wald's GR book, and Hawking and Ellis. The final chapter of this course concerns quantum field theory in curved spacetime. Here I mainly used Birrell and Davies, and Wald's second book. The latter also contains a nice discussion of the laws of black hole mechanics.

1 Spherical stars

1.1 Cold stars

To understand the astrophysical significance of black holes we must discuss stars. In particular, how do stars end their lives?

A normal star like our Sun is supported against contracting under its own gravity by pressure generated by nuclear reactions in its core. However, eventually the star will use up its nuclear "fuel". If the gravitational self-attraction is to be balanced then some new source of pressure is required. If this balance is to last forever then this new source of pressure must be *non-thermal* because the star will eventually cool.

A non-thermal source of pressure arises quantum mechanically from the Pauli principle, which makes a gas of cold fermions resist compression (this is called degeneracy pressure). A *white dwarf* is a star in which gravity is balanced by electron degeneracy pressure. The Sun will end its life as a white dwarf. White dwarfs are very dense compared to normal stars e.g. a white dwarf with the same mass as the Sun would have a radius around a hundredth of that of the Sun. Using Newtonian gravity one can show that a white dwarf cannot have a mass greater than the *Chandrasekhar limit* $1.4M_\odot$ where $M_\odot$ is the mass of the Sun. A star more massive than this cannot end its life as a white dwarf (unless it somehow sheds some of its mass).

Once the density of matter approaches nuclear density, the degeneracy pressure of neutrons becomes important (at such high density, inverse beta decay converts protons into neutrons). A *neutron star* is supported by the degeneracy pressure of neutrons. These stars are tiny: a solar mass neutron star would have a radius of around 10km (the radius of the Sun is 7×10^5km). Recall that validity of Newtonian gravity requires $|\Phi| \ll 1$ where Φ is the Newtonian gravitational potential. At the surface of a such a neutron star one has $|\Phi| \sim 0.1$ and so a Newtonian description is inadequate: one has to use GR.

In this chapter we will see that GR predicts that there is a maximum mass for neutron stars. Remarkably, this is independent of the (unknown) properties of matter at extremely high density and so it holds for *any* cold star. As we will explain, detailed calculations reveal the maximum mass to be around $3M_\odot$. Hence a hot star more massive than this cannot end its life as a cold star (unless it sheds some mass e.g. in a supernova). Instead the star will undergo complete gravitational collapse to form a black hole.

We will make the simplifying assumption that the star is spherically symmetric. As we will see, the Schwarzschild solution is the unique spherically symmetric vacuum solution and hence describes the gravitational field outside any spherically symmetric star. The interior of the star can be modelled using a perfect fluid and so spacetime inside the star is determined by solving the Einstein equation with a perfect fluid source and matching onto the Schwarzschild solution outside the star.

1.2 Spherical symmetry

We need to define what we mean by a spacetime being spherically symmetric. You are familiar with the idea that a round sphere is invariant under rotations, which form the group $SO(3)$. In more mathematical language, this can be phrased as follows. The set of all isometries of a manifold with metric forms a group. Consider the unit round metric on S^2:

$$d\Omega^2 = d\theta^2 + \sin^2\theta \, d\phi^2. \tag{1.1}$$

The isometry group of this metric is $SO(3)$ (actually $O(3)$ if we include reflections). Any 1-dimensional subgroup of $SO(3)$ gives a 1-parameter group of isometries, and hence a Killing vector field. A *spacetime* is spherically symmetric if it possesses the same symmetries as a round S^2:

Definition. A spacetime is *spherically symmetric* if its isometry group contains an $SO(3)$ subgroup whose orbits are 2-spheres. (The *orbit* of a point p under a group of diffeomorphisms is the set of points that one obtains by acting on p with all of the diffeomorphisms.)

The statement about the orbits is important: there are examples of spacetimes with $SO(3)$ isometry group in which the orbits of $SO(3)$ are 3-dimensional (e.g. Taub-NUT space: see Hawking and Ellis).

Definition. In a spherically symmetric spacetime, the *area-radius function* $r : M \to \mathbb{R}$ is defined by $r(p) = \sqrt{A(p)/4\pi}$ where $A(p)$ is the area of the S^2 orbit through p. (In other words, the S^2 passing through p has induced metric $r(p)^2 d\Omega^2$.)

1.3 Time-independence

Definition. A spacetime is *stationary* if it admits a Killing vector field k^a which is everywhere timelike: $g_{ab}k^a k^b < 0$.

We can choose coordinates as follows. Pick a hypersurface Σ nowhere tangent to k^a and introduce coordinates x^i on Σ. Assign coordinates (t, x^i) to the point parameter distance t along the integral curve through the point on Σ with coordinates x^i. This gives a coordinates chart such that $k^a = (\partial/\partial t)^a$. Since k^a is a Killing vector field, the metric is independent of t and hence takes the form

$$ds^2 = g_{00}(x^k)dt^2 + 2g_{0i}(x^k)dtdx^i + g_{ij}(x^k)dx^i dx^j \tag{1.2}$$

where $g_{00} < 0$. Conversely, given a metric of this form, $\partial/\partial t$ is obviously a timelike Killing vector field and so the metric is stationary.

Next we need to introduce the notion of hypersurface-orthogonality. Let Σ be a hypersurface in M specified by $f(x) = 0$ where $f : M \to \mathbb{R}$ is smooth with $df \neq 0$ on Σ. Then the 1-form df is normal to Σ. (Proof: let t^a be any vector tangent to Σ then $df(t) = t(f) = t^\mu \partial_\mu f = 0$ because f is constant on Σ.) Any other 1-form n normal to Σ can be written as $n = gdf + fn'$ where g is a smooth function with $g \neq 0$ on Σ and n' is a smooth 1-form. Hence we have $dn = dg \wedge df + df \wedge n' + f dn'$ so $(dn)|_\Sigma = (dg - n') \wedge df$. So if n is normal to Σ then

$$(n \wedge dn)|_\Sigma = 0 \tag{1.3}$$

Conversely:

Theorem (Frobenius). If n is a non-zero 1-form such that $n \wedge dn = 0$ *everywhere* then there exist functions f, g such that $n = gdf$ so n is normal to surfaces of constant f i.e. n is *hypersurface-orthogonal*.

Definition. A spacetime is *static* if it admits a hypersurface-orthogonal timelike Killing vector field. (So static implies stationary.)

For a static spacetime, we know that k^a is hypersurface-orthogonal so when defining coordinates (t, x^i) we can choose Σ to be orthogonal to k^a. But Σ is the surface $t = 0$, with normal dt. It follows that, at $t = 0$, $k_\mu \propto (1, 0, 0, 0)$ in our chart, i.e., $k_i = 0$. However $k_i = g_{0i}(x^k)$ so we must have $g_{0i}(x^k) = 0$. So in adapted coordinates a static metric takes the form

$$ds^2 = g_{00}(x^k)dt^2 + g_{ij}(x^k)dx^i dx^j \tag{1.4}$$

where $g_{00} < 0$. Note that this metric has a discrete time-reversal isometry: $(t, x^i) \to (-t, x^i)$. So static means "time-independent *and* invariant under time reversal". For example, the metric of a rotating star can be stationary but not static because time-reversal changes the sense of rotation.

1.4 Static, spherically symmetric, spacetimes

We're interested in determining the gravitational field of a time-independent spherical object so we assume our spacetime to be stationary and spherically symmetric. By this we mean that the isometry group is $\mathbb{R} \times SO(3)$ where the the $\mathbb{R}$ factor corresponds to "time translations" (i.e., the associated Killing vector field is timelike) and the orbits of $SO(3)$ are 2-spheres as above. It can be shown that any such spacetime must actually be static. (The gravitational field of a rotating star can be stationary but the rotation defines a preferred axis and so the spacetime would not be spherically symmetric.) So let's consider a spacetime that is both static and spherically symmetric.

Staticity means that we have a timelike Killing vector field k^a and we can foliate our spacetime with surfaces Σ_t orthogonal to k^a. One can argue that the orbit of $SO(3)$ through $p \in \Sigma_t$ lies within Σ_t. We can define spherical polar coordinates on Σ_0 as follows. Pick a S^2 symmetry orbit in Σ_0 and define spherical polars (θ, ϕ) on it. Extend the definition of (θ, ϕ) to the rest of Σ_0 by defining them to be constant along (spacelike) geodesics normal to this S^2 within Σ_0. Now we use (r, θ, ϕ) as coordinates on Σ_0 where r is the area-radius function defined above (which assume is not constant, i.e., $dr \neq 0$). The metric on Σ_0 must take the form

$$ds^2 = e^{2\Psi(r)} dr^2 + r^2 d\Omega^2 \tag{1.5}$$

$drd\theta$ and $drd\phi$ terms cannot appear because they would break spherical symmetry. Note that r is *not* "the distance from the origin". Finally, we define coordinates (t, r, θ, ϕ) with t the parameter distance from Σ_0 along the integral curves of k^a. The metric must take the form

$$ds^2 = -e^{2\Phi(r)} dt^2 + e^{2\Psi(r)} dr^2 + r^2 d\Omega^2 \tag{1.6}$$

We'll model the matter inside a star as a perfect fluid, with energy momentum tensor

$$T_{ab} = (\rho + p) u_a u_b + p g_{ab} \tag{1.7}$$

where u^a is the 4-velocity of the fluid (a unit timelike vector: $g_{ab} u^a u^b = -1$), and ρ,p are the energy density and pressure measured in the fluid's local rest frame (i.e. by an observer with 4-velocity u^a).

Since we're interested in a time-independent and spherically symmetric situation we assume that the fluid is at rest, so u^a is in the time direction:

$$u^a = e^{-\Phi} \left(\frac{\partial}{\partial t} \right)^a \tag{1.8}$$

Our assumptions of staticity and spherical symmetry implies that ρ and p depend only on r. Let R denote the (area-)radius of the star. Then ρ and p vanish for $r > R$.

1.5 Tolman-Oppenheimer-Volkoff equations

Recall that a perfect fluid's equations of motion are determined by energy-momentum tensor conservation. But the latter follows from the Einstein equation and the contracted Bianchi identity. Hence we can obtain the equations of motion from just the Einstein equation. Now the Einstein tensor inherits the symmetries of the metric and so there are only three non-trivial components of the Einstein equation. These are the tt, rr and $\theta\theta$ components (spherical symmetry implies that the $\phi\phi$ component is proportional to the $\theta\theta$ component).

Define $m(r)$ by

$$e^{2\Psi(r)} = \left(1 - \frac{2m(r)}{r}\right)^{-1} \tag{1.9}$$

and note that the LHS is positive so $m(r) < r/2$. The tt component of the Einstein equation gives

$$\frac{dm}{dr} = 4\pi r^2 \rho \tag{1.10}$$

The rr component of the Einstein equation gives

$$\frac{d\Phi}{dr} = \frac{m + 4\pi r^3 p}{r(r-2m)} \tag{1.11}$$

The final non-trivial component of the Einstein equation is the $\theta\theta$ component This gives a third equation of motion. But this is more easily derived from the r-component of energy-momentum conservation $\nabla_\mu T^{\mu\nu} = 0$, i.e., from the fluid equations of motion. This gives

$$\frac{dp}{dr} = -(p+\rho)\frac{(m + 4\pi r^3 p)}{r(r-2m)} \tag{1.12}$$

We have 3 equations but 4 unknowns (m, Φ, ρ, p) so we need one more equation. We are interested in a *cold* star, i.e., one with vanishing temperature T. Thermodynamics tells us that T, p and ρ are not independent: they are related by the fluid's equation of state e.g. $T = T(\rho, p)$. Hence the condition $T = 0$ implies a relation between p and ρ, i.e, a *barotropic equation of state* $p = p(\rho)$. So, for a cold star, p is not an independent variable so we have 3 equations for 3 unknowns. These are called the *Tolman-Oppenheimer-Volkoff equations.*

We assume that $\rho > 0$ and $p > 0$, i.e., the energy density and pressure of matter are positive. We also assume that p is an increasing function of ρ. If this were not the case then the fluid would be unstable: a fluctuation in some region that led to an increase in ρ would decrease p, causing the fluid to move into this region and hence further increase in ρ, i.e., the fluctuation would grow.

1.6 Outside the star: the Schwarzschild solution

Consider first the spacetime outside the star: $r > R$. We then have $\rho = p = 0$. For $r > R$ (1.10) gives $m(r) = M$, constant. Integrating (1.11) gives

$$\Phi = \frac{1}{2}\log\left(1 - 2M/r\right) + \Phi_0 \tag{1.13}$$

for some constant Φ_0. We then have $g_{tt} \to -e^{2\Phi_0}$ as $r \to \infty$. The constant Φ_0 can be eliminated by defining a new time coordinate $t' = e^{\Phi_0} t$. So without loss of generality we can set $\Phi_0 = 0$ and we have arrived at the *Schwarzschild solution*

$$ds^2 = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\Omega^2 \tag{1.14}$$

The constant M is the mass of the star. One way to see this is to note that for large r, the Schwarzschild solution reduces to the solution of linearized theory describing the gravitational field far from a body of mass M (a change of radial coordinate is required to see this). We will give a precise definition of mass later in this course.

The components of the above metric are singular at the *Schwarzschild radius* $r = 2M$, where g_{tt} vanishes and g_{rr} diverges. A solution describing a static spherically symmetric star can exist only if $r = 2M$ corresponds to a radius inside the star, where the Schwarzschild solution does not apply. Hence a static, spherically symmetric star must have a radius greater than its Schwarzschild radius:

$$R > 2M \tag{1.15}$$

Normal stars have $R \gg 2M$ e.g. for the Sun, $2M \approx 3$km whereas $R \approx 7 \times 10^5$km.

1.7 The interior solution

Integrating (1.10) gives

$$m(r) = 4\pi \int_0^r \rho(r') r'^2 dr' + m_\star \tag{1.16}$$

where $m_\star$ is a constant.

Now Σ_t should be smooth at $r = 0$ (the centre of the star). Recall that any smooth Riemannian manifold is locally flat, i.e., measurements in a sufficiently small region will be the same as in Euclidean space. In Euclidean space, a sphere of area-radius r also has proper radius r, i.e., all points on the sphere lie proper distance r from the centre. Hence the same must be true for a small sphere on Σ_t. The proper radius of a sphere of area-radius r is $\int_0^r e^{\Psi(r')} dr' \approx e^{\Psi(0)} r$ for small r. Hence we need $e^{\Psi(0)} = 1$ for the metric to be smooth at $r = 0$. This implies $m(0) = 0$ and so $m_\star = 0$.

Now at $r = R$, our interior solution must match onto the exterior Schwarzschild solution. For $r > R$ we have $m(r) = M$ so continuity of $m(r)$ determines M:

$$M = 4\pi \int_0^R \rho(r) r^2 dr \tag{1.17}$$

This is *formally* the same as the equation relating total mass to density in Newtonian theory. But there is an important difference: in the Euclidean space of Newtonian theory, the volume element on a surface of constant t is $r^2 \sin\theta dr \wedge d\theta \wedge d\phi$ and so the RHS above gives the total energy of matter. However, in GR, the volume element on Σ_t is $e^\Psi r^2 \sin\theta dr \wedge d\theta \wedge d\phi$ so the total energy of the matter is

$$E = 4\pi \int_0^R \rho e^\Psi r^2 dr \tag{1.18}$$

and since $e^\Psi > 1$ (as $m > 0$) we have $E > M$: the energy of the matter in the star is greater than the *total* energy M of the star. The difference $E - M$ can be interpreted as the gravitational binding energy of the star.

In GR there is a lower limit on the size of stars that has no Newtonian analogue. To see this, note that the definition (1.9) implies $m(r)/r < 1/2$ for all r. Evaluating at $r = R$ recovers the result $R > 2M$ discussed above. (To see that this has no Newtonian analogue, we can reinsert factors of G and c to write it as $GM/(c^2 R) < 1/2$. Taking the Newtonian limit $c \to \infty$ the equation becomes trivial.)

This lower bound can be improved. Note that (1.12) implies $dp/dr \leq 0$ and hence $d\rho/dr \leq 0$. Using this it can be shown (examples sheet 1) that

$$\frac{m(r)}{r} < \frac{2}{9}\left[1 - 6\pi r^2 p(r) + (1 + 6\pi r^2 p(r))^{1/2}\right] \tag{1.19}$$

Evaluating at $r = R$ we have $p = 0$ and hence obtain the *Buchdahl inequality*

$$R > \frac{9}{4} M \tag{1.20}$$

The derivation of this inequality assumes only $\rho \geq 0$ and $d\rho/dr \leq 0$ and nothing about the equation of state, so it also applies to hot stars satisfying these assumptions. This inequality is sharp: on examples sheet 1 it is shown that stars with constant density ρ can get arbitrarily close to saturating it (the pressure at the centre diverges in the limit in which the inequality becomes an equality).

The TOV equations can be solved by numerical integration as follows. Regard (1.10) and (1.12) as a pair of coupled first order ordinary differential equations for $m(r)$ and $\rho(r)$ (recall that $p = p(\rho)$ and $dp/d\rho > 0$). These can be solved, at least numerically on a computer, given initial conditions for $m(r)$ and $\rho(r)$ at $r = 0$. We have just seen that $m(0) = 0$. Hence just need to specify the value $\rho_c = \rho(0)$ for the density at the centre of the star.

Given a value for ρ_c we can solve (1.10) and (1.12). The latter equation shows that p (and hence ρ) decreases as r increases. Since the pressure vanishes at the surface of the star, the radius R is determined by the condition $p(R) = 0$. This determines R as a function of ρ_c. Equation (1.17) then determines M as a function of ρ_c. Finally we determine $\Phi(r)$ inside the star by integrating (1.11) inwards from $r = R$ with initial condition $\Phi(R) = (1/2)\log(1 - 2M/R)$ (from (1.13)). Hence, for a given equation of state, static, spherically symmetric, cold stars form a 1-parameter family of solutions, labelled by ρ_c.

1.8 Maximum mass of a cold star

When one follows the above procedure then one finds that, as ρ_c increases, M increases to a maximum value but then decreases for larger ρ_c as shown in Fig. 1.

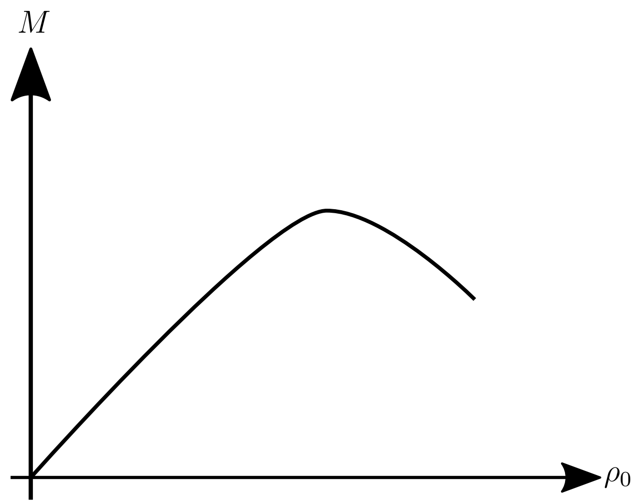


Figure 1. Plot of M against ρ_c for typical equation of state.

The maximum mass will depend on the details of the equation of state of cold matter. For example, taking an equation of state corresponding to white dwarf matter reproduces the Chandrasekhar bound (as mentioned above, one does not need GR for this, it can be obtained using Newtonian gravity). Experimentally we know this equation of state up to some density ρ_0 (around nuclear density) but we don't know its form for $\rho > \rho_0$. One might expect that by an appropriate choice of the equation of state for $\rho > \rho_0$ one could arrange for the maximum mass to be very large, say $100M_\odot$.

This is not the case. Remarkably, GR predicts that there is an upper bound on the mass of a cold, spherically symmetric star, which is independent of the form of the equation of state at high density. This upper bound is around $5M_\odot$.

Recall that ρ is a decreasing function of r. Define the *core* of the star as the region in which $\rho > \rho_0$ where we don't know the equation of state and the *envelope* as the region $\rho < \rho_0$ where we do know the equation of state. Let r_0 be the radius of the core, i.e., the core is the region $r < r_0$ and the envelope the region $r_0 < r < R$. The mass of the core is defined as $m_0 = m(r_0)$. Equation (1.16) gives

$$m_0 \geq \frac{4}{3}\pi r_0^3 \rho_0 \tag{1.21}$$

We would have the same result in Newtonian gravity. In GR we have the extra constraint (1.19). Evaluating this at $r = r_0$ gives

$$\frac{m_0}{r_0} < \frac{2}{9}\left[1 - 6\pi r_0^2 p_0 + (1 + 6\pi r_0^2 p_0)^{1/2}\right] \tag{1.22}$$

where $p_0 = p(r_0)$ is determined from ρ_0 using the equation of state. Note that the RHS is a decreasing function of p_0 so we obtain a simpler (but weaker) inequality by evaluating the RHS at $p_0 = 0$:

$$m_0 < \frac{4}{9} r_0 \tag{1.23}$$

i.e., the core satisfies the Buchdahl inequality. The two inequalities (1.21) and (1.23) define a finite region of the $m_0 - r_0$ plane shown in Fig. 2. From this, the upper bound on the mass of the core is

$$m_0 < \sqrt{\frac{16}{243\pi\rho_0}} \tag{1.24}$$

Hence although we don't know the equation of state inside the core, GR predicts that its mass cannot be indefinitely large. Experimentally, we don't know the equation of state of cold matter at densities much higher than the density of atomic nuclei so we take $\rho_0 = 5 \times 10^{14}$ g/cm^3, the density of nuclear matter. This gives an upper bound on the core mass $m_0 < 5M_\odot$.

Now, given a core with mass m_0 and radius r_0, the envelope region is determined uniquely by solving numerically (1.10) and (1.12) with initial conditions $m = m_0$ and $\rho = \rho_0$ at $r = r_0$, using the known equation of state at density $\rho < \rho_0$. This show that the total mass M of the star is a function of the core parameters m_0 and r_0. By investigating (numerically) the behaviour of this function as m_0 and r_0 range over the allowed region of Fig. 2, it is found that the M is maximised at the maximum of m_0 (actually one uses the stricter inequality (1.22) instead of (1.23) to define the allowed

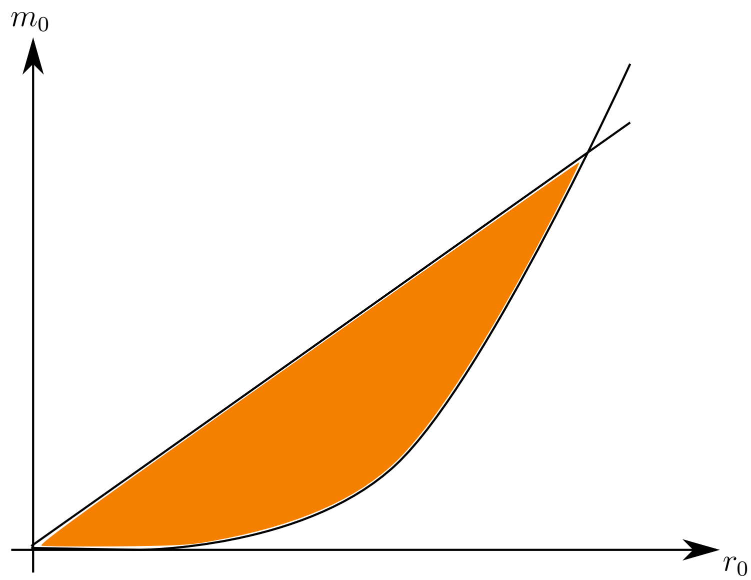


Figure 2. Allowed region of $m_0 - r_0$ plane

region). At this maximum, the envelope contributes less than 1% of the total mass so the maximum value of M is almost the same as the maximum value of m_0, i.e., $5M_\odot$.

It should be emphasized that this is an upper bound that applies for *any* physically reasonable equation of state for $\rho > \rho_0$. But any particular equation of state will have its own upper bound, which will be less than the above bound. Indeed, one can improve the above bound by adding further criteria to what one means by "physically reasonable". For example, the speed of sound in the fluid is $(dp/d\rho)^{1/2}$. It is natural to demand that this should not exceed the speed of light, i.e. one could add the extra condition $dp/d\rho \leq 1$. This has the effect of reducing the upper bound to about $3M_\odot$.

2 The Schwarzschild black hole

We have seen that GR predicts that a cold star cannot have a mass more than a few times $M_\odot$. A very massive hot star cannot end its life as a cold star unless it somehow sheds some of its mass. Instead it will undergo complete gravitational collapse to form a *black hole*. The simplest black hole solution is described by the Schwarzschild geometry. So far, we have used the Schwarzschild metric to describe the spacetime outside a spherical star. In this chapter we will investigate the geometry of spacetime under the assumption that the Schwarzschild solution is valid everywhere, i.e., no matter is present.

2.1 Birkhoff's theorem

In *Schwarzschild coordinates* (t, r, θ, ϕ), the Schwarzschild solution is

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\Omega^2 \tag{2.1}$$

This is actually a 1-parameter family of solutions. The parameter M take either sign but, as mentioned above, it has the interpretation of a mass so we will assume $M > 0$ here. The case $M < 0$ will be discussed later.

Previously we assumed that we were dealing with $r > 2M$. But the above metric is also a solution of the vacuum Einstein equation for $0 < r < 2M$. We will see below how these are related. $r = 2M$ is called the *Schwarzschild radius*.

We derived the Schwarzschild solution under the assumptions of staticity and spherical symmetry. It turns out that the former is not required:

Theorem (Birkhoff). Any spherically symmetric solution of the vacuum Einstein equation is isometric to the Schwarzschild solution.

Proof. See Hawking and Ellis.

This theorem assumes only spherical symmetry but the Schwarzschild solution has an additional isometry: $\partial/\partial t$ is a hypersurface-orthogonal Killing vector field. It is timelike for $r > 2M$ so the $r > 2M$ Schwarzschild solution is static.

Birkhoff's theorem implies that the spacetime outside any spherical body is described by the time-independent (exterior) Schwarzschild solution. This is true even if the body itself is time-dependent. For example, consider a spherical star that "uses up its nuclear fuel" and collapses to form a white dwarf or neutron star. The spacetime outside the star will be described by the static Schwarzschild solution even during the collapse.

2.2 Gravitational redshift

Consider two observers A and B who remain at fixed (r, θ, ϕ) in the Schwarzschild geometry. Let A have $r = r_A$ and B have $r = r_B$ where $r_B > r_A$. Now assume that A sends two photons to B separated by a coordinate time Δt as measured by A. Since $\partial/\partial t$ is an isometry, the path of the second photon is the same as the path of the first one, just translated in time through an interval Δt.

Exercise. Show that the *proper* time between the photons emitted by A, as measured by A is $\Delta\tau_A = \sqrt{1 - 2M/r_A}\Delta t$.

Similarly the proper time interval between the photons received by B, as measured by B is $\Delta\tau_B = \sqrt{1 - 2M/r_B}\Delta t$. Eliminating Δt gives

$$\frac{\Delta\tau_B}{\Delta\tau_A} = \sqrt{\frac{1 - 2M/r_B}{1 - 2M/r_A}} > 1 \tag{2.2}$$

Now imagine that we are considering light *waves* propagating from A to B. Applying the above argument to two successive wavecrests shows that the above formula relates the period $\Delta\tau_A$ of the waves emitted by A to the period $\Delta\tau_B$ of the waves received by B. For light, the period is the same as the wavelength (since $c = 1$): $\Delta\tau = \lambda$. Hence $\lambda_B > \lambda_A$: the light undergoes a *redshift* as it climbs out of the gravitational field.

If B is at large radius, i.e., $r_B \gg 2M$, then we have

$$1 + z \equiv \frac{\lambda_B}{\lambda_A} = \sqrt{\frac{1}{1 - 2M/r_A}} \tag{2.3}$$

Note that this diverges as $r_A \to 2M$. We showed above that a spherical star must have radius $R > 9M/4$ so (taking $r_A = R$) it follows that the maximum possible redshift of light emitted from the surface of a spherical star is $z = 2$.

2.3 Geodesics of the Schwarzschild solution

Let $x^\mu(\tau)$ be an affinely parameterized geodesic with tangent vector $u^\mu = dx^\mu/d\tau$. Since $k = \partial/\partial t$ and $m = \partial/\partial\phi$ are Killing vector fields we have the conserved quantities

$$E = -k \cdot u = \left(1 - \frac{2M}{r}\right)\frac{dt}{d\tau} \tag{2.4}$$

and

$$h = m \cdot u = r^2 \sin^2\theta \frac{d\phi}{d\tau} \tag{2.5}$$

For a timelike geodesic, we choose τ to be proper time and then E has the interpretation of energy per unit rest mass and h is the angular momentum per unit rest mass. (To see this, evaluate the expressions for E and h at large r where the metric is almost flat so one can use results from special relativity.) For a null geodesic, the freedom to rescale the affine parameter implies that E and h do not have direct physical significance. However, the ratio h/E is invariant under this rescaling. For a null geodesic which propagates to large r (where the metric is almost flat and the geodesic is a straight line), $b = |h/E|$ is the *impact parameter*, i.e., the distance of the null geodesic from "a line through the origin", more precisely the distance from a line of constant ϕ parallel (at large r) to the geodesic.

Exercise. Determine the Euler-Lagrange equation for $\theta(\tau)$ and eliminate $d\phi/d\tau$ to obtain

$$r^2 \frac{d}{d\tau}\left(r^2 \frac{d\theta}{d\tau}\right) - h^2 \frac{\cos\theta}{\sin^3\theta} = 0 \tag{2.6}$$

One can define spherical polar coordinates on S^2 in many different ways. It is convenient to rotate our (θ, ϕ) coordinates so that our geodesic has $\theta = \pi/2$ and $d\theta/d\tau = 0$ at $\tau = 0$, i.e., the geodesic initially lies in, and is moving tangentially to, the "equatorial plane" $\theta = \pi/2$. We emphasize: this is just a choice of the coordinates (θ, ϕ). Now, whatever $r(\tau)$ is (and we don't know yet), the above equation is a second order ODE for θ with initial conditions $\theta = \pi/2$, $d\theta/d\tau = 0$. One solution of this initial value problem is $\theta(\tau) = \pi/2$ for all τ. Standard uniqueness results for ODEs guarantee that this is the unique solution. Hence we have shown that we can always choose our θ, ϕ coordinates so that the geodesic is confined to the equatorial plane. We shall assume this henceforth.

Exercise. Choosing τ to be proper time in the case of a timelike geodesic, and arclength (proper distance) in the case of a spacelike geodesic implies $g_{\mu\nu}u^\mu u^\nu = -\sigma$ where $\sigma = 1, 0, -1$ for a timelike, null or spacelike geodesic respectively. Rearrange this equation to obtain

$$\frac{1}{2}\left(\frac{dr}{d\tau}\right)^2 + V(r) = \frac{1}{2}E^2 \tag{2.7}$$

where

$$V(r) = \frac{1}{2}\left(1 - \frac{2M}{r}\right)\left(\sigma + \frac{h^2}{r^2}\right) \tag{2.8}$$

Hence the radial motion of the geodesic is determined by the same equation as a Newtonian particle of unit mass and energy $E^2/2$ moving in a 1d potential $V(r)$.

2.4 Eddington-Finkelstein coordinates

Consider the Schwarzschild solution with $r > 2M$. Let's consider the simplest type of geodesic: radial null geodesics. "Radial" means that θ and ϕ are constant along the geodesic, so $h = 0$. By rescaling the affine parameter τ we can arrange that $E = 1$. The geodesic equation reduces to

$$\frac{dt}{d\tau} = \left(1 - \frac{2M}{r}\right)^{-1}, \qquad \frac{dr}{d\tau} = \pm 1 \tag{2.9}$$

where the upper sign is for an outgoing geodesic (i.e. increasing r) and the lower for ingoing. From the second equation it is clear that an ingoing geodesic starting at some

$r > 2M$ will reach $r = 2M$ in finite affine parameter. Dividing gives

$$\frac{dt}{dr} = \pm \left(1 - \frac{2M}{r}\right)^{-1} \tag{2.10}$$

The RHS has a simple pole at $r = 2M$ and hence t diverges logarithmically as $r \to 2M$. To investigate what is happening at $r = 2M$, define the "Regge-Wheeler radial coordinate" r_* by

$$dr_* = \frac{dr}{\left(1 - \frac{2M}{r}\right)} \qquad \Rightarrow \qquad r_* = r + 2M \log \left| \frac{r}{2M} - 1 \right| \tag{2.11}$$

where we made a choice of constant of integration. (We're interested only in $r > 2M$ for now, the modulus signs are for later use.) Note that $r_* \sim r$ for large r and $r_* \to -\infty$ as $r \to 2M$. (Fig. 3). Along a radial null geodesic we have

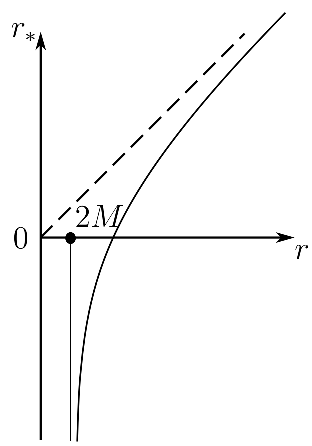


Figure 3. Regge=Wheeler radial coordinate

$$\frac{dt}{dr_*} = \pm 1 \tag{2.12}$$

so

$$t \mp r_* = \text{constant}. \tag{2.13}$$

Let's define a new coordinate v by

$$v = t + r_* \tag{2.14}$$

so that v is constant along ingoing radial null geodesics. Now let's use (v, r, θ, ϕ) as coordinates instead of (t, r, θ, ϕ). The new coordinates are called *ingoing Eddington-Finkelstein coordinates*. We eliminate t by $t = v - r_*(r)$ and hence

$$dt = dv - \frac{dr}{\left(1 - \frac{2M}{r}\right)} \tag{2.15}$$

Substituting this into the metric gives

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dv^2 + 2dvdr + r^2 d\Omega^2 \tag{2.16}$$

Written as a matrix we have, in these coordinates,

$$g_{\mu\nu} = \begin{pmatrix} -(1-2M/r) & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2\theta \end{pmatrix} \tag{2.17}$$

Unlike the metric components in Schwarzschild coordinates, the components of the above matrix are smooth for all $r > 0$, in particular they are smooth at $r = 2M$. Furthermore, this matrix has determinant $-r^4 \sin^2\theta$ and hence is non-degenerate for any $r > 0$ (except at $\theta = 0, \pi$ but this is just because the coordinates (θ, ϕ) are not defined at the poles of the spheres). This implies that its signature is Lorentzian for $r > 0$ since a change of signature would require an eigenvalue passing through zero.

The Schwarzschild spacetime can now be *extended* through the surface $r = 2M$ to a new region with $r < 2M$. Is the metric (2.16) a solution of the vacuum Einstein equation in this region? Yes. The metric components are *real analytic* functions of the above coordinates, i.e., they can be expanded as convergent power series about any point. If a real analytic metric satisfies the Einstein equation in some open set then it will satisfy the Einstein equation everywhere. Since we know that the (2.16) satisfies the vacuum Einstein equation for $r > 2M$ it must also satisfy this equation for $r > 0$.

Note that the new region with $0 < r < 2M$ is spherically symmetric. How is this consistent with Birkhoff's theorem?

Exercise. For $r < 2M$, define r_* by (2.11) and t by (2.14). Show that if the metric (2.16) is transformed to coordinates (t, r, θ, ϕ) then it becomes (2.1) but now with $r < 2M$.

Note that ingoing radial null geodesics in the EF coordinates have $dr/d\tau = -1$ (and constant v). Hence such geodesics will reach $r = 0$ in finite affine parameter. What happens there? Since the metric is Ricci flat, the simplest non-trivial scalar construced from the metric is $R_{abcd}R^{abcd}$ and a calculation gives

$$R_{abcd}R^{abcd} \propto \frac{M^2}{r^6} \tag{2.18}$$

This diverges as $r \to 0$. Since this is a scalar, it diverges in all charts. Therefore there exists no chart for which the metric can be smoothly extended through $r = 0$.

$r = 0$ is an example of a *curvature singularity*, where tidal forces become infinite and the known laws of physics break down. Strictly speaking, $r = 0$ is not part of the spacetime manifold because the metric is not defined there.

Recall that in $r > 2M$, Schwarzschild solution admits the Killing vector field $k = \partial/\partial t$. Let's work out what this is in ingoing EF coordinates. Denote the latter by x^μ so we have

$$k = \frac{\partial}{\partial t} = \frac{\partial x^\mu}{\partial t}\frac{\partial}{\partial x^\mu} = \frac{\partial}{\partial v} \tag{2.19}$$

since the EF coordinates are independent of t except for $v = t + r_*(r)$. We use this equation to extend the definition of k to $r \leq 2M$. Note that $k^2 = g_{vv}$ so k is null at $r = 2M$ and spacelike for $0 < r < 2M$. Hence the extended Schwarzschild solution is static only in the $r > 2M$ region.

2.5 Finkelstein diagram

So far we have considered ingoing radial null geodesics, which have $v = \text{constant}$ and $dr/d\tau = -1$. Now consider the outgoing geodesics. For $r > 2M$ in Schwarzschild coordinates these have $t - r_* = \text{constant}$. Converting to EF coordinates gives $v = 2r_* + \text{constant}$, i.e.,

$$v = 2r + 4M \log\left|\frac{r}{2M} - 1\right| + \text{constant} \tag{2.20}$$

To determine the behaviour of geodesics in $r \leq 2M$ we need to use EF coordinates from the start. This gives

Exercise. Consider radial null geodesics in ingoing EF coordinates. Show that these fall into two families: "ingoing" with $v = \text{constant}$ and "outgoing" satisfying either (2.20) or $r \equiv 2M$.

It is interesting to plot the radial null geodesics on a spacetime diagram. Let $t_* = v - r$ so that the ingoing radial null geodesics are straight lines at 45° in the (t_*, r) plane. This gives the *Finkelstein diagram* of Fig. 4.

Knowing the ingoing and outgoing radial null geodesics lets us draw light "cones" on this diagram. Radial *timelike* curves have tangent vectors that lie inside the light cone at any point.

The "outgoing" radial null geodesics have increasing r if $r > 2M$. But if $r < 2M$ then r decreases for *both* families of null geodesics. Both reach the curvature singularity at $r = 0$ in finite affine parameter. Since nothing can travel faster than light, the same is true for radial timelike curves. We will show below that r decreases along any timelike or null curve (irrespective of whether or not it is radial or geodesic) in $r < 2M$. Hence no signal can be sent from a point with $r < 2M$ to a point with $r > 2M$, in particular

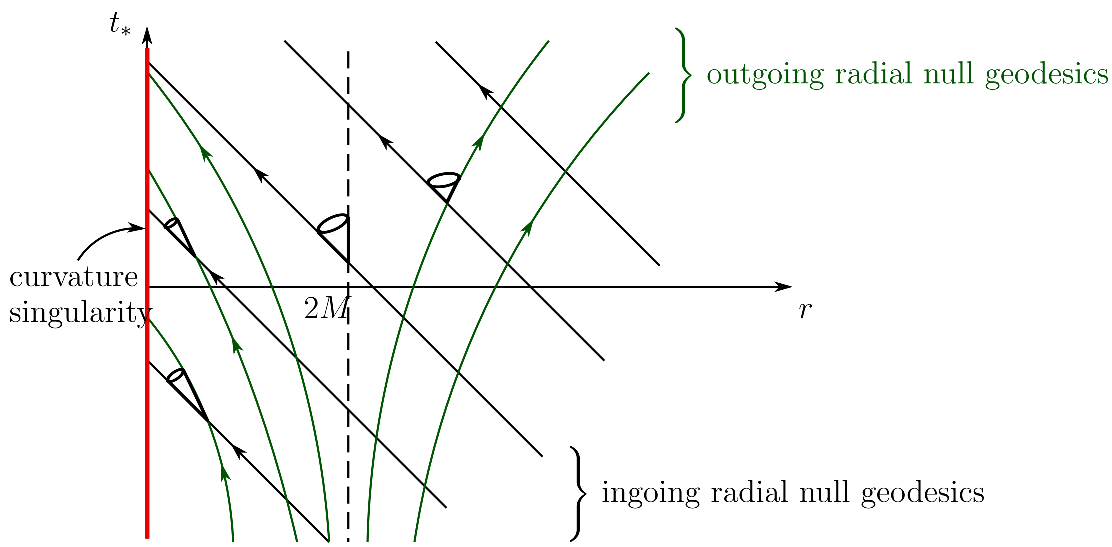


Figure 4. Finkelstein diagram

to a point with $r = \infty$. This is the defining property of a *black hole*: a region of an "asymptotically flat" spacetime from which it is impossible to send a signal to infinity.

2.6 Gravitational collapse

Consider the fate of a massive spherical star once it exhausts its nuclear fuel. The star will shrink under its own gravity. As mentioned above, Birkhoff's theorem implies that the geometry outside the star is given by the Schwarzschild solution even when the star is time-dependent. If the star is not too massive then eventually it might settle down to a white dwarf or neutron star. But if it is sufficiently massive then this is not possible: nothing can prevent the star from shrinking until it reaches its Schwarzschild radius $r = 2M$.

We can visualize this process of gravitational collapse on a Finkelstein diagram. We just need to remove the part of the diagram corresponding the interior of the star. By continuity, points on the surface of the collapsing star will follow radial timelike curves in the Schwarzschild geometry. This is shown in Fig. 5.

On examples sheet 1, it is shown that the total proper time along a timelike curve in $r \leq 2M$ cannot exceed πM. (For $M = M_\odot$ this is about 10^{-5}s.) Hence the star will collapse and form a curvature singularity in finite proper time as measured by an (unlucky) observer on the star's surface.

Note the behaviour of the outgoing radial null geodesics, i.e., light rays emitted from the surface of the star. As the star's surface approaches $r = 2M$, light from the

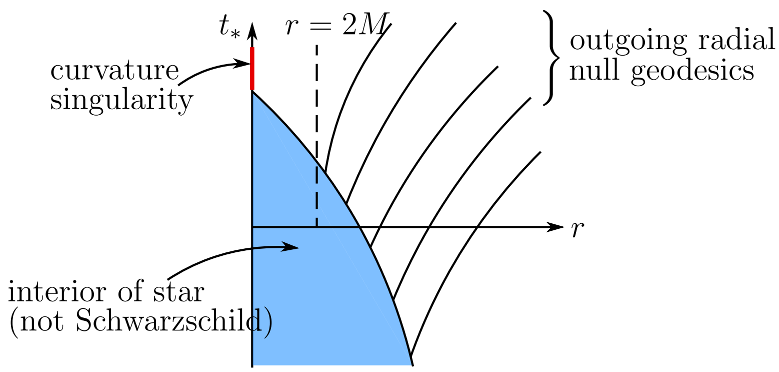


Figure 5. Finkelstein diagram for gravitational collapse

surface takes longer and longer to reach a distant observer. The observer will never see the star cross $r = 2M$. Equation (2.3) shows that the redshift of this light diverges as $r \to 2M$. So the distant observer will see the star fade from view as $r \to 2M$.

2.7 Black hole region

We will show that the region $r \leq 2M$ of the extended Schwarzschild solution describes a black hole. First recall some definitions.

Definition. A vector is *causal* if it is timelike or null (we adopt the convention that a null vector must be non-zero). A curve is causal if its tangent vector is everywhere causal.

At any point of a spacetime, the metric determines two light cones in the tangent space at that point. We would like to regard one of these as the "future" light-cone and the other as the "past" light-cone. We do this by picking a causal vector field and defining the future light cone to be the one in which it lies:

Definition. A spacetime is *time-orientable* if it admits a *time-orientation*: a causal vector field T^a. Another causal vector X^a is *future-directed* if it lies in the same light cone as T^a and *past-directed* otherwise.

Note that any other time orientation is either everywhere in the same light cone as T^a or everywhere in the opposite light cone. Hence a time-orientable spacetime admits exactly two inequivalent time-orientations.

In the $r > 2M$ region of the Schwarzschild spacetime, we choose $k = \partial/\partial t$ as our time-orientation. (We could just as well choose $-k$ but this is related by the isometry $t \to -t$ and therefore leads to equivalent results.) k is not a time-orientation in $r < 2M$ because in ingoing EF coordinates we have $k = \partial/\partial v$, which is spacelike for $r < 2M$. However, $\pm\partial/\partial r$ is globally null ($g_{rr} = 0$) and hence defines a time-orientation. We just need to choose the sign that gives a time orientation equivalent to k for $r > 2M$. Note that

$$k \cdot (-\partial/\partial r) = -g_{vr} = -1 \tag{2.21}$$

and if the inner product of two causal vectors is negative then they lie in the same light cone (remind yourself why!). Therefore we can use $-\partial/\partial r$ to define our time orientation for $r > 0$. Note that $-\partial/\partial r$ is tangent to ingoing radial null geodesics.

Proposition. Let $x^\mu(\lambda)$ be any future-directed causal curve (i.e. one whose tangent vector is everywhere future-directed and causal). Assume $r(\lambda_0) \leq 2M$. Then $r(\lambda) \leq 2M$ for $\lambda \geq \lambda_0$.

Proof. The tangent vector is $V^\mu = dx^\mu/d\lambda$. Since $-\partial/\partial r$ and V^a both are future-directed causal vectors we have

$$0 \geq \left(-\frac{\partial}{\partial r}\right) \cdot V = -g_{r\mu}V^\mu = -V^v = -\frac{dv}{d\lambda} \qquad \Rightarrow \qquad \frac{dv}{d\lambda} \geq 0 \tag{2.22}$$

hence v is non-decreasing along any future-directed causal curve. We also have

$$V^2 = -\left(1 - \frac{2M}{r}\right)\left(\frac{dv}{d\lambda}\right)^2 + 2\frac{dv}{d\lambda}\frac{dr}{d\lambda} + r^2\left(\frac{d\Omega}{d\lambda}\right)^2 \tag{2.23}$$

where $(d\Omega/d\lambda)^2 = (d\theta/d\lambda)^2 + \sin^2\theta(d\phi/d\lambda)^2$. Rearranging gives

$$-2\frac{dv}{d\lambda}\frac{dr}{d\lambda} = -V^2 + \left(\frac{2M}{r} - 1\right)\left(\frac{dv}{d\lambda}\right)^2 + r^2\left(\frac{d\Omega}{d\lambda}\right)^2 \tag{2.24}$$

Note that every term on the RHS is non-negative if $r \leq 2M$. Consider a point on the curve for which $r \leq 2M$ so

$$\frac{dv}{d\lambda}\frac{dr}{d\lambda} \leq 0 \tag{2.25}$$

Assume, to obtain a contradiction, that $dr/d\lambda > 0$ at this point. Then this inequality is consistent with (2.22) only if $dv/d\lambda = 0$. Plugging this into (2.24) and using the fact that the terms on the RHS are non-negative implies that $V^2 = 0$ and $d\Omega/d\lambda = 0$. But now the only non-zero component of V^μ is $V^r = dr/d\lambda > 0$ so V is a positive multiple of $\partial/\partial r$ and hence is past-directed, a contradiction.

We have shown that $dr/d\lambda \leq 0$ if $r \leq 2M$, which completes the proof. (If $r < 2M$ then $dr/d\lambda < 0$ for if $dr/d\lambda = 0$ then (2.24) implies $d\Omega/d\lambda = dv/d\lambda = 0$ but then we have $V^\mu = 0$, a contradiction. Hence if $r(\lambda_0) < 2M$ then $r(\lambda)$ is monotonically decreasing for $\lambda \geq \lambda_0$.)

This result implies that no future-directed causal curve connects a point with $r \leq 2M$ to a point with $r > 2M$. More physically: it is impossible to send a signal from a point with $r \leq 2M$ to a point with $r > 2M$, in particular to a point at $r = \infty$. A *black hole* is defined to be a region of spacetime from which it is impossible to send a signal to infinity. (We will define "infinity" more precisely later.) The boundary of this region is the *event horizon*.

Our result shows that points with $r \leq 2M$ of the extended Schwarzschild spacetime lie inside a black hole. However, it is easy to show that there do exist future-directed causal curves from a point with $r > 2M$ to $r = \infty$ (e.g. an outgoing radial null curve) so points with $r > 2M$ are not inside a black hole. Hence $r = 2M$ is the event horizon.

2.8 Detecting black holes

There are two important properties that underpin detection methods:

First: *there is no upper bound on the mass of a black hole.* This contrasts with cold stars, which have an upper bound around $3M_\odot$.

Second: *black holes are very small.* A black hole has radius $R = 2M$. A solar mass black holes has radius 3km. A black hole with the same mass as the Earth would have radius 0.9cm.

There are other systems which satisfy either one of these conditions. For example, there is no upper limit on the mass of a *cluster* of stars or a cloud of gas. But these would have size much greater than $2M$. On the other hand, neutron stars are also very small, with radius not much greater than $2M$. But a neutron star cannot be arbitrarily massive. It is the *combination* of a large mass concentrated into a small region which distinguishes black holes from other kinds of object.

Since black hole do not emit electromagnetic radiation directly, we infer their existence from their effect on nearby luminous matter. For example, stars near the centre of our galaxy are observed to be orbiting around the galactic centre (Fig. 6). From the shapes of the orbits, one can deduce that there is an object with mass $4 \times 10^6 M_\odot$ at the centre of the galaxy. Since some of the stars get close to the galactic centre, one can infer that this mass must be concentrated within a radius of about 6 light hours (6×10^9km about the same size as the Solar System) since otherwise these stars would be ripped apart by tidal effects. The only object that can contain so much mass in such a small region is a black hole.

Many other galaxies are also believed to contain enormous black holes at their centres (some with masses greater than $10^9 M_\odot$). Black holes with mass greater than about $10^6 M_\odot$ are referred to as *supermassive*. There appears to be a correlation between the mass of the black hole and the mass of its host galaxy, with the former typically about a thousandth of the latter. Supermassive black holes do not form directly from gravitational collapse of a normal star (since the latter cannot have a mass much greater than about $100M_\odot$). It is still uncertain how such large black holes form.

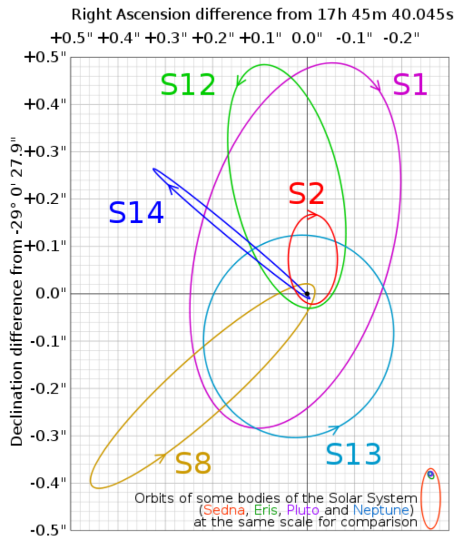


Figure 6. Stars orbiting the galactic centre.

To understand the motion of matter around a black hole, let's consider timelike geodesics in more detail. The effective potential has turning points where

$$r_{\pm} = \frac{h^2 \pm \sqrt{h^4 - 12h^2M^2}}{2M} \tag{2.26}$$

If $h^2 < 12M^2$ then there are no turning points, the effective potential is a monotonically increasing function of r. If $h^2 > 12M^2$ then there are two turning points. $r = r_+$ is a minimum and $r = r_-$ a maximum (Fig. 7). Hence there exist stable circular orbits

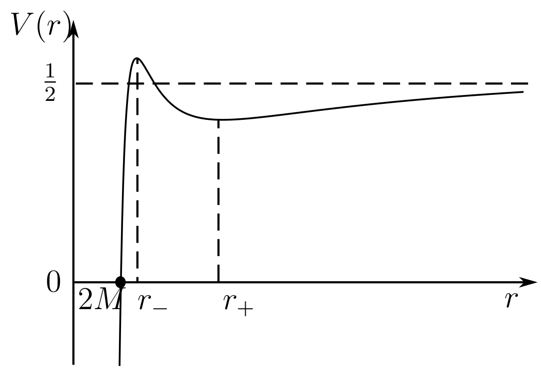


Figure 7. Timelike geodesics: effective potential for $h^2 > 12M^2$

with $r = r_+$ and unstable circular orbits with $r = r_-$.

Exercise. Show that $3M < r_- < 6M < r_+$.

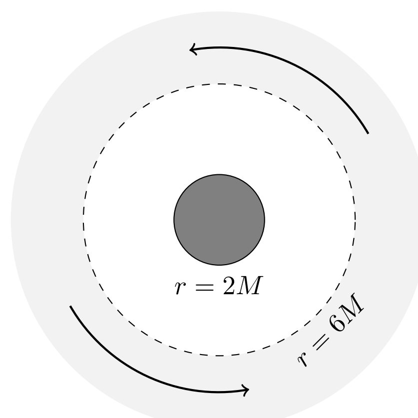
$r_+ = 6M$ is called the *innermost stable circular orbit* (ISCO). For a normal star, this lies well inside the star, where the Schwarzschild solution is not valid. But for a black hole it lies outside the event horizon. There is no analogue of the ISCO in Newtonian theory, for which all circular orbits are stable and exist down to arbitrarily small r.

The energy per unit rest mass of a circular orbit can be calculated using $E^2/2 = V(r)$ (since $dr/d\tau = 0$):

Exercise. Show that the energy per unit rest mass of a circular orbit $r = r_\pm$ can be written

$$E = \frac{r - 2M}{r^{1/2}(r - 3M)^{1/2}} \tag{2.27}$$

Hence a body following a circular orbit with large r has $E \approx 1 - M/(2r)$, i.e., its energy is $m - Mm/(2r)$ where m is the mass of the body. The first term is just the rest mass energy ($E = mc^2$) and the second term can be interpreted in Newtonian terms as the sum of its kinetic energy and gravitational potential energy.



Black holes formed in gravitational collapse of a star have M less than about $100M_\odot$ since (hot) stars with significantly higher mass than this do not exist. Such holes are referred to as *solar mass black holes*. The main way that such black holes are detected is to look for a binary system consisting of a black hole and a normal star. In such a system, the black hole can be surrounded by an *accretion disc*: a disc of gas orbiting the black hole, stripped off the star by tidal forces due to the black hole's gravitational field. Supermassive black holes can also have (much bigger) accretion discs: in this case, the disc is formed from matter present near the centre of the host galaxy.

As a first approximation, we can treat particles in an accretion disc as moving on geodesics. A particle in this material will gradually lose energy because of friction in the disc and so its value of E will decrease. This implies that r will decrease: the particle will gradually spiral in to smaller and smaller r. This process can be approximated by the particle moving slowly from one stable circular orbit to another. Eventually the particle will reach the ISCO, which has $E = \sqrt{8/9}$, after which it falls rapidly into the hole. The fraction of rest mass converted to radiation in this process is

$1-\sqrt{8/9} \approx 0.06$. This is an enormous amount of energy, much higher than the fraction of rest mass energy liberated in nuclear reactions. That is why accretion discs around supermassive black holes are believed to power some of the most energetic phenomena in the universe e.g. quasars.

The energy that the particle loses as it moves towards the ISCO leaves the disc as electromagnetic radiation. The first detections of black holes were made in the 1970s by observing X-rays emitted from accretion discs around solar mass black holes in our galaxy. The X-rays exhibits a characteristic cut-off in red-shift, corresponding to the ISCO. In 2019, radio observations were used to produce an image (Fig. 8) of the accretion disc around the supermassive black hole at the centre of the nearby galaxy M87, which has an estimated mass of $6 \times 10^9 M_\odot$.

Figure 8. Image of the accretion disc around the supermassive black hole at the centre of M87. The disc is nearly face-on to us and there is a dark area in the centre corresponding roughly to the ISCO. (Credit: Event Horizon Telescope.)

Of course we are no longer restricted to electromagnetic observations of black holes. The subject was revolutionized in 2015 by the LIGO/VIRGO collaboration's direct detection of gravitational waves from a (solar mass) black hole *merger*. (See my General Relativity lecture notes for more on this.) The evidence that the objects involved were black holes is that they had to be very compact (or else they could not get close enough to emit significant gravitational waves) and their masses (around $30M_\odot$) were too large for them to be neutron stars. Furthermore, the detected gravitational waves were in agreement with predictions from supercomputer simulations of black hole mergers. The post-merger gravitational waves exhibited damped oscillations, just as expected of a black hole settling down to equilibrium. Many other detections were made subsequently, including mergers where one, or both, compact objects were neutron stars. Future gravitational wave experiments will include LISA, a space-based detector, which will detect gravitational waves of much lower frequency than LIGO/VIRGO, including waves emited by mergers involving supermassive black holes.

2.9 White holes

We defined ingoing EF coordinates using ingoing radial null geodesics. What happens if we do the same thing with outgoing radial null geodesics? Starting with the Schwarzschild solution in Schwarzschild coordinates with $r > 2M$, let

$$u = t - r_* \tag{2.28}$$

so $u =$ constant along outgoing radial null geodesics. Now introduce *outgoing Eddington-Finkelstein* (u, r, θ, ϕ). The Schwarzschild metric becomes

$$ds^2 = -\left(1 - \frac{2M}{r}\right) du^2 - 2dudr + r^2 d\Omega^2 \tag{2.29}$$

Just as for the ingoing EF coordinates, this metric is smooth with non-vanishing determinant for $r > 0$ and hence can be extended to a new region $r \leq 2M$. Once again we can define Schwarzschild coordinates in $r < 2M$ to see that the metric in this region is simply the Schwarzschild metric. There is a curvature singularity at $r = 0$.

This $r < 2M$ region is *not the same* as the $r < 2M$ region in the ingoing EF coordinates. An easy way to see this is to look at the outgoing radial null geodesics, i.e., lines of constant u. We saw above (in the Schwarzschild coordinates) that these have $dr/d\tau = 1$ hence they propagate from the curvature singularity at $r = 0$, through the surface $r = 2M$ and then extend to large r. This is impossible for the $r < 2M$ region we discussed previously since that region is a black hole.

Exercise. Show that $k = \partial/\partial u$ in outgoing EF coordinates and that the time-orientation which is equivalent to k for $r > 2M$ is given by $+\partial/\partial r$.

The $r < 2M$ region of the outgoing EF coordinates is a *white hole*: a region which no signal from infinity can enter. A white hole is the time reverse of a black hole. To see this, make the substitution $u = -v$ to see that the above metric is isometric to (2.16). The only difference is the sign of the time orientation. It follows that no signal can be sent from a point with $r > 2M$ to a point with $r < 2M$. Any timelike curve starting with $r < 2M$ must pass through the surface $r = 2M$ within finite proper time.

White holes are believed to be unphysical. A black hole is formed from a normal star by gravitational collapse. But a white hole begins with a singularity, so to create a white hole one must first make a singularity. Black holes are stable objects: small perturbations of a black hole are believed to decay. Applying time-reversal implies that white holes must be unstable objects: small perturbations of a white hole become large under time evolution.

2.10 The Kruskal extension

We have seen that the Schwarzschild spacetime can be extended in two different ways, revealing the existence of a black hole region and a white hole region. How are these different regions related to each other? This is answered by introducing a new set of coordinates. Start in the region $r > 2M$. Define *Kruskal-Szekeres coordinates* (U, V, θ, ϕ) by

$$U = -e^{-u/(4M)}, \qquad V = e^{v/(4M)}, \tag{2.30}$$

so $U < 0$ and $V > 0$. Note that

$$UV = -e^{r_*/(2M)} = -e^{r/(2M)} \left(\frac{r}{2M} - 1 \right) \tag{2.31}$$

The RHS is a monotonic function of r (for $r > 0$) and hence this equation determines $r(U, V)$ uniquely. We also have

$$\frac{V}{U} = -e^{t/(2M)} \tag{2.32}$$

which determines $t(U, V)$ uniquely.

Exercise. Show that in Kruskal-Szekeres coordinates, the metric is

$$ds^2 = -\frac{32M^3 e^{-r(U,V)/(2M)}}{r(U,V)} dUdV + r(U,V)^2 d\Omega^2 \tag{2.33}$$

Hint. First transform the metric to coordinates (u, v, θ, ϕ) and then to KS coordinates.

Let us now *define* the function $r(U, V)$ for $U \geq 0$ or $V \leq 0$ by (2.31). This new metric can be analytically extended, with non-vanishing determinant, through the surfaces $U = 0$ and $V = 0$ to new regions with $U > 0$ or $V < 0$.

Let's consider the surface $r = 2M$. Equation (2.31) implies that either $U = 0$ or $V = 0$. Hence KS coordinates reveal that $r = 2M$ is actually *two* surfaces, that intersect at $U = V = 0$. Similarly, the curvature singularity at $r = 0$ corresponds to $UV = 1$, a hyperbola with two branches. This information can be summarized on the *Kruskal diagram* of Fig. 9.

One should think of "time" increasing in the vertical direction on this diagram. Radial null geodesics are lines of constant U or V, i.e., lines at 45° to the horizontal. This diagram has four regions. Region I is the region we started in, i.e., the region $r > 2M$ of the Schwarzschild solution. Region II is the black hole that we discovered using ingoing EF coordinates (note that these coordinates cover regions I and II of the Kruskal diagram), Region III is the white hole that we discovered using outgoing EF coordinates. And region IV is an entirely new region. In this region, $r > 2M$

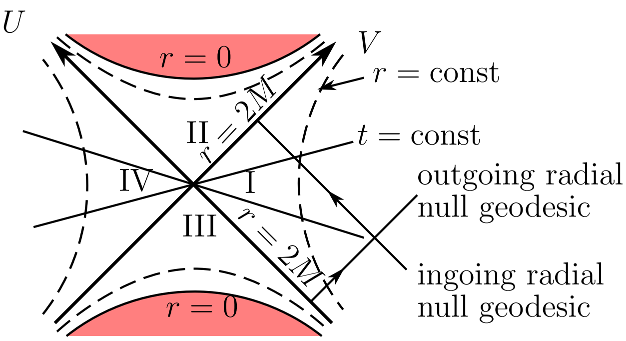


Figure 9. Kruskal diagram

and so this region is again described by the Schwarzschild solution with $r > 2M$. This is a new asymptotically flat region. It is isometric to region I: the isometry is $(U, V) \to (-U, -V)$. Note that it is impossible for an observer in region I to send a signal to an observer in region IV. If they want to communicate then one or both of them will have to travel into region II (and then hit the singularity).

Note that the singularity in region II appears to the future of any point. Therefore it is not appropriate to think of the singularity as a "place" inside the black hole. It is more appropriate to think of it as a "time" at which tidal forces become infinite. The black hole region is time-dependent because, in Schwarzschild coordinates, it is r, not t, that plays the role of time. This region can be thought of as describing a homogeneous but anisotropic universe approaching a "big crunch". Conversely, the white hole singularity resembles a "big bang" singularity.

Most of this diagram is unphysical. If we include a timelike worldline corresponding to the surface of a collapsing star and then replace the region to the left of this line by the (non-vacuum) spacetime corresponding to the star's interior then we get a diagram in which only regions I and II appear (Fig. 10). Inside the matter, $r = 0$ is just the origin of polar coordinates, where the spacetime is smooth.

Finally, let's discuss time translations in Kruskal coordinates:

Exercise. Show that, in Kruskal coordinates

$$k = \frac{1}{4M}\left(V\frac{\partial}{\partial V} - U\frac{\partial}{\partial U}\right) \qquad k^2 = -\left(1 - \frac{2M}{r}\right) \tag{2.34}$$

The result for k^2 can be deduced either by direct calculation or by noting that it is true for $r > 2M$ (e.g. use Schwarzschild coordinates) and the the LHS and RHS are both analytic functions of U, V (since the metric is analytic). Hence the result must be true everywhere.

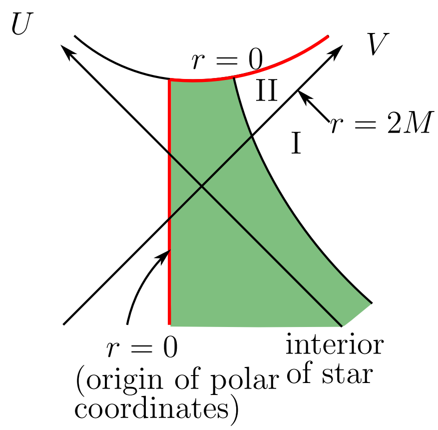


Figure 10. Kruskal diagram for gravitational collapse. The region to the left of the shaded region is not part of the spacetime.

k is timeline in regions I and IV, spacelike in regions II and III, and null (or zero) where $r = 2M$ i.e. where $U = 0$ or $V = 0$. The orbits (integral curves) of k on a Kruskal diagram are shown in Fig. 11. Note that the sets $\{U = 0\}$ and $\{V = 0\}$ are fixed (mapped into themselves) by k and that $k = 0$ on the "bifurcation 2-sphere" $U = V = 0$. Hence points on the latter are also fixed by k.

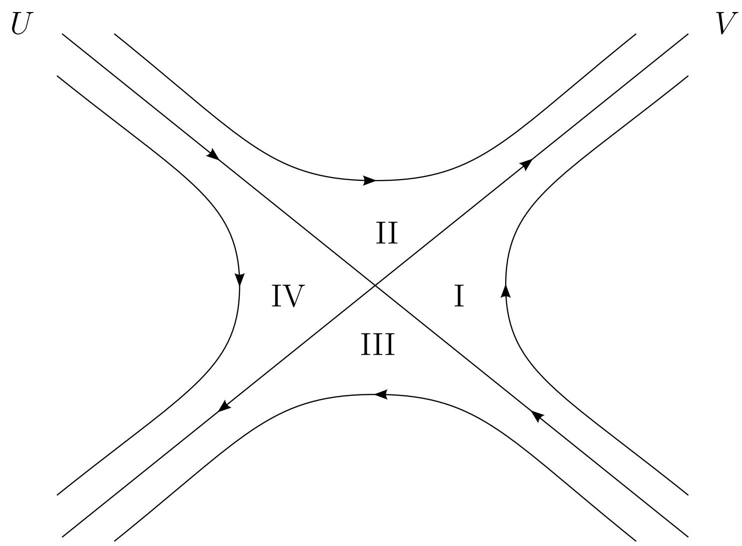


Figure 11. Orbits of k in Kruskal spacetime.

2.11 Einstein-Rosen bridge

Recall equation (2.32): in region I we have $V/U = -e^{t/(2M)}$. Hence a surface of constant t in region I is a straight line through the origin in the Kruskal diagram. These extend naturally into region IV (see Fig. 9). Let's investigate the geometry of these hypersurfaces. Define a new coordinate ρ by

$$r = \rho + M + \frac{M^2}{4\rho} \tag{2.35}$$

Given r, there are two possible solutions for ρ (see Fig. 12). We choose $\rho > M/2$

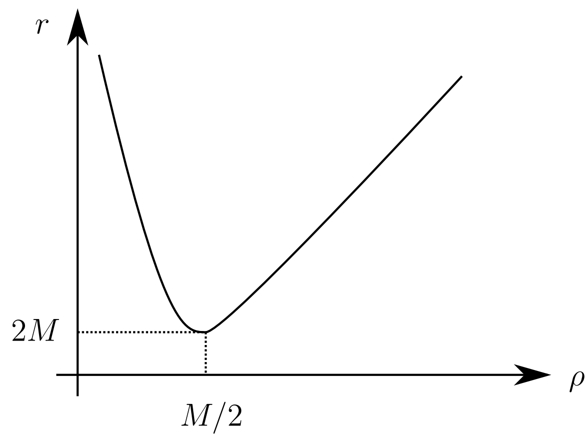


Figure 12. Area-radius function r as a function of isotropic radial coordinate ρ.

in region I and $0 < \rho < M/2$ in region IV. The Schwarzschild metric in *isotropic coordinates* (t, ρ, θ, ϕ) is then (exercise)

$$ds^2 = -\frac{(1 - M/(2\rho))^2}{(1 + M/(2\rho))^2} dt^2 + \left(1 + \frac{M}{2\rho}\right)^4 \left(d\rho^2 + \rho^2 d\Omega^2\right) \tag{2.36}$$

The transformation $\rho \to M^2/(4\rho)$ is an isometry that interchanges regions I and IV. Of course the above metric is singular at $\rho = M/2$ but we know this is just a coordinate singularity. Now consider the metric of a surface of constant t:

$$ds^2 = \left(1 + \frac{M}{2\rho}\right)^4 \left(d\rho^2 + \rho^2 d\Omega^2\right) \tag{2.37}$$

This metric is non-singular for $\rho > 0$. It defines a Riemannian 3-manifold with topology $\mathbb{R} \times S^2$ (where $\mathbb{R}$ is parameterized by ρ). Its geometry can be visualized by embedding the surface into 4d Euclidean space (examples sheet 1). If we suppress the θ

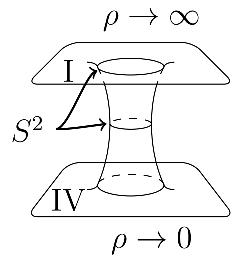


Figure 13. Einstein-Rosen bridge

direction, this gives the diagram shown in Fig 13. The geometry has two asymptotically flat regions ($\rho \to \infty$ and $\rho \to 0$) connected by a "throat" with minimum radius $2M$ at $\rho = M/2$. A surfaces of constant t in the Kruskal spacetime is called an "Einstein-Rosen bridge".

2.12 Extendibility

Definition. A spacetime $(\mathcal{M}, g)$ is *extendible* if it is isometric to a proper subset of another spacetime $(\mathcal{M}', g')$. The latter is called an *extension* of $(\mathcal{M}, g)$. (In GR we require that the spacetime manifold M is connected so both M and M' should be connected.)

For example, let $(\mathcal{M}, g)$ denote the Schwarzschild solution with $r > 2M$ and let $(\mathcal{M}', g')$ denote the Kruskal spacetime. Then $\mathcal{M}$ is a subset of $\mathcal{M}'$ (i.e. region I). If we define a map to take a point of $\mathcal{M}$ to the corresponding point of $\mathcal{M}'$ then this is just the identity map in region I, which is obviously an isometry. The Kruskal spacetime $(\mathcal{M}', g')$ is *inextendible* (not extendible). In particular, it cannot be extended beyond $r = 0$ (this is obvious for a C^2 metric, because of the curvature singularity, but it is true even for a C^0 metric (Sbierski 2015)).

2.13 Singularities

We say that the metric is singular in some basis if its components are not smooth or its determinant vanishes. A *coordinate singularity* can be eliminated by a change of coordinates (e.g. $r = 2M$ in the Schwarzschild spacetime). These are unphysical. However, if it is not possible to eliminate the bad behaviour by a change of coordinates then we have a physical singularity. We have already seen an example of this: a *scalar curvature singularity*, where some scalar constructed from the Riemann tensor blows up, cannot be eliminated by a change of coordinates and hence is physical. It is also possible to have more general curvature singularities for which no scalar constructed from the Riemann tensor diverges but, nevertheless, there exists no chart in which the Riemann tensor remains finite.

Not all physical singularities are curvature singularities. For example consider the manifold $M = \mathbb{R}^2$, introduce plane polar coordinates (r, ϕ) (so $\phi \sim \phi + 2\pi$) and define the 2d Riemannian metric

$$g = dr^2 + \lambda^2 r^2 d\phi^2 \tag{2.38}$$

where $\lambda > 0$. The metric determinant vanishes at $r = 0$. If $\lambda = 1$ then this is just Euclidean space in plane polar coordinates, so we can convert to Cartesian coordinates to see that $r = 0$ is just a coordinate singularity, i.e., g can be smoothly extended to $r = 0$. But consider the case $\lambda \neq 1$. In this case, let $\phi' = \lambda\phi$ to obtain

$$g = dr^2 + r^2 d\phi'^2 \tag{2.39}$$

which is *locally* isometric to Euclidean space and hence has vanishing Riemann tensor (so there is no curvature singularity at $r = 0$). However, it is not globally isometric to Euclidean space because the period of ϕ' is $2\pi\lambda$. Consider a circle $r = \epsilon$. This has

$$\frac{\text{circumference}}{\text{radius}} = \frac{2\pi\lambda\epsilon}{\epsilon} = 2\pi\lambda \tag{2.40}$$

which does *not* tend to 2π as $\epsilon \to 0$. Recall that any smooth Riemannian manifold is locally flat, i.e., one recovers results of Euclidean geometry on sufficiently small scales (one can introduce normal coordinates to show this). The above result shows that this is not true for small circles centred on $r = 0$. Hence the above metric cannot be smoothly extended to $r = 0$. This is an example of a *conical singularity*.

A problem in defining singularities is that they are not "places": they do not belong to the spacetime manifold because we define spacetime as a pair (M, g) where g is a smooth Lorentzian metric. For example, $r = 0$ is not part of the Kruskal manifold. Similarly, in the example just discussed if we want a smooth Riemannian manifold then we must take $M = \mathbb{R}^2 \backslash (0, 0)$ so that $r = 0$ is not a point of M. But in both of these examples, the existence of the singularity implies that some geodesics cannot be extended to arbitrarily large affine parameter because they "end" at the singularity. It is this property that we will use to define what we mean by "singular".

First we must eliminate a trivial case, corresponding to the possibility of a geodesic ending simply because we haven't taken the range of its parameter to be large enough. Recall that a curve is a smooth map $\gamma : (a, b) \to M$. Sometimes a curve can be *extended*, i.e., it is part of a bigger curve. If this happens then the first curve will have an endpoint, which is defined as follows.

Definition. $p \in M$ is a *future endpoint* of a future-directed causal curve $\gamma : (a, b) \to M$ if, for any neighbourhood O of p, there exists t_0 such that $\gamma(t) \in O$ for all $t > t_0$. We say that γ is *future-inextendible* if it has no future endpoint. Similary for past endpoints and past inextendibility. γ is inextendible if it is both future and past inextendible.

For example, let (M, g) be Minkowski spacetime. Let $\gamma : (-\infty, 0) \to M$ be $\gamma(t) = (t, 0, 0, 0)$. Then the origin is a future endpoint of γ. However, if we instead let (M, g) be Minkowski spacetime with the origin deleted then γ is future-inextendible.

Definition. A geodesic is *complete* if an affine parameter for the geodesic extends to $\pm\infty$. A *spacetime* is *geodesically complete* if all inextendible causal geodesics are complete.

For example, Minkowski spacetime is geodesically complete, as is the spacetime describing a static spherical star. However, the Kruskal spacetime is geodesically *incomplete* because some geodesics have $r \to 0$ in finite affine parameter and hence cannot be extended to infinite affine parameter. A similar definition applies to Riemannian manifolds.

A spacetime that is extendible will usually also be geodesically incomplete. But in this case, it is clear that the incompleteness arises because we are not considering "the whole spacetime". So we will regard a spacetime as singular if it is geodesically incomplete and inextendible. This is the case for the Kruskal spacetime.

3 The initial value problem

In the next chapter we will explain why GR predicts that black holes necessarily form under certain circumstances. To do this, we need to understand the initial value problem for GR.

3.1 Predictability

Definition. Let (M, g) be a time-orientable spacetime. A *partial Cauchy surface* Σ is a hypersurface for which no two points are connected by a causal curve in M. The *future domain of dependence* of Σ, denoted $D^+(\Sigma)$, is the set of $p \in M$ such that every past-inextendible causal curve through p intersects Σ. The past domain of dependence, $D^-(\Sigma)$, is defined similarly. The *domain of dependence of* Σ is $D(\Sigma) = D^+(\Sigma) \cup D^-(\Sigma)$.

$D(\Sigma)$ is the region of spacetime in which one can determine what happens from data specified on Σ. For example, any causal geodesic (i.e. free particle worldline) in $D(\Sigma)$ must intersect Σ at some point p. The geodesic is determined uniquely by specifying its tangent vector (velocity) at p. More generally, solutions of hyperbolic partial differential equations are uniquely determined in $D(\Sigma)$ by initial data prescribed on Σ.

Here, by "hyperbolic partial differential equations" we mean second order partial differential equations for a set of tensor fields $T^{(i)ab\dots}{}_{cd\dots}$ $(i = 1, \dots N)$ for which the equations of motion take the form

$$g^{ef}\nabla_e\nabla_f T^{(i)ab\dots}{}_{cd\dots} = \dots \tag{3.1}$$

where the RHS is a tensor that depends smoothly on the metric and its derivatives, and linearly on the fields $T^{(j)}$ and their first derivatives, but not their second or higher derivatives. The Klein-Gordon equation is of this form, as are the Maxwell equations when written using a vector potential A_a in Lorenz gauge.

For example, let Σ be the positive x-axis in 2d Minkowski spacetime (M, g) (figure 14). $D^+(\Sigma)$ is the set of points with $0 \leq t < x$, $D^-(\Sigma)$ is the set of points with $-x < t \leq 0$. The boundary of $D(\Sigma)$ is the pair of null rays $t = \pm x$ for $x > 0$. Let Σ' be the entire x-axis. This gives $D(\Sigma') = M$.

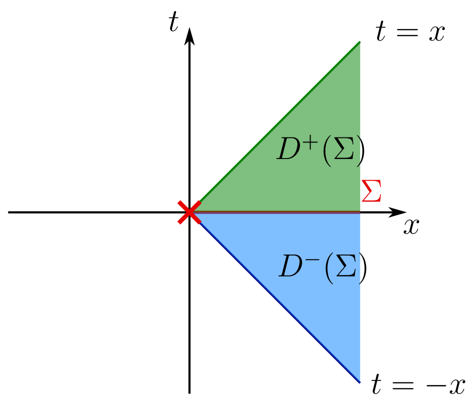


Figure 14. The regions $D^{\pm}(\Sigma)$

Consider the wave equation $\nabla^a\nabla_a\psi = -\partial_t^2\psi + \partial_x^2\psi = 0$ in this spacetime. Specifying the initial data $(\psi, \partial_t\psi)$ on Σ determines the solution uniquely in $D(\Sigma)$. Specifying initial data on Σ' determines the solution uniquely throughout M. Two such solutions whose initial data agrees on the subset Σ of Σ' will agree within $D(\Sigma)$ but differ on $M \backslash D(\Sigma)$.

This is true in general: if $D(\Sigma) \neq M$ then solutions of hyperbolic equations will not be uniquely determined in $M \backslash D(\Sigma)$ by data on Σ. Given only this data, there will be infinitely many different solutions on M which agree within $D(\Sigma)$.

Definition. A spacetime (M, g) is *globally hyperbolic* if it admits a *Cauchy surface*: a partial Cauchy surface Σ such that $M = D(\Sigma)$.

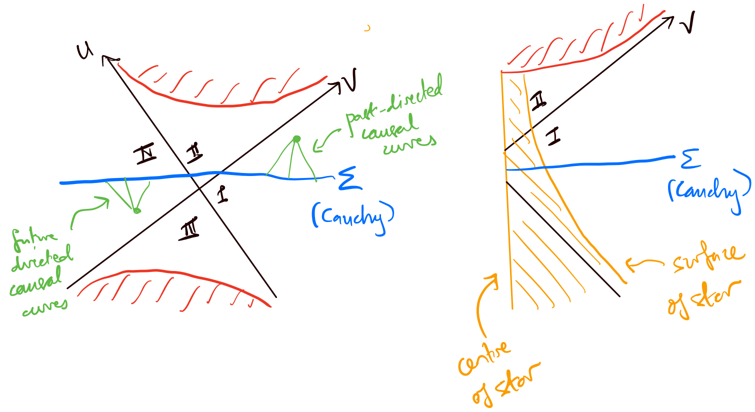


Figure 15. Examples of Cauchy surfaces for the Kruskal spacetime and the spacetime describing spherically symmetric gravitational collapse.

(If Σ is not a Cauchy surface then the past/future boundary of $D(\Sigma)$ is called the past/future *Cauchy horizon*. We will define it more precisely later.)

Hence a globally hyperbolic spacetime is one in which one can predict what happens everywhere from data on Σ. Minkowski spacetime is globally hyperbolic e.g. a surface of constant t is a Cauchy surface. Other examples are the the Kruskal spacetime and the spacetime describing spherically symmetric gravitational collapse, see Fig. 15.

To obtain an example of a spacetime which is not globally hyperbolic, delete the origin from 2d Minkowski spacetime (the cross in Fig. 14). For any partial Cauchy surface Σ, there will be some inextendible causal curves which don't intersect Σ because they "end" at the origin.

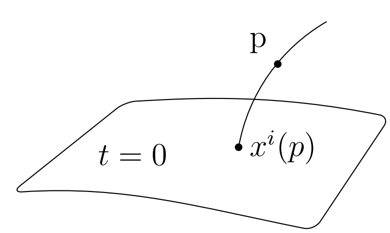
The following theorem is proved in Wald:

Theorem. Let (M, g) be globally hyperbolic. Then (i) there exists a *global time function*: a map $t : M \to \mathbb{R}$ such that $-(dt)^a$ (normal to surfaces of constant t) is future-directed and timelike (ii) surfaces of constant t are Cauchy surfaces, and these all have the same topology Σ (iii) the topology of M is $\mathbb{R} \times \Sigma$.

Exercise. Show that $U + V$ is a global time function in the Kruskal spacetime.

Since the surface $U + V = 0$ is an Einstein-Rosen bridge, it follows that Σ has topology $\mathbb{R} \times S^2$ in this case. The topology of M is $\mathbb{R}^2 \times S^2$.

If (M, g) is globally hyperbolic then we can perform a 3 + 1 split ("Arnowitt-Deser-Misner (ADM) decomposition") of spacetime as follows. Let t be a time function. Introduce coordinates x^i



$(i = 1, 2, 3)$ on the Cauchy surface $t = 0$. Pick an everywhere timelike vector field T^a. Given $p \in M$, consider the integral curve of T^a through p. This intersects the surface $t = 0$ at a unique point. Let $x^i(p)$ be the coordinates of this point. This defines functions $x^i : M \to \mathbb{R}$. We use (t, x^i) as our coordinate chart. It is conventional to use the following notation for the metric components:

$$ds^2 = -N^2 dt^2 + h_{ij}(dx^i + N^i dt)(dx^j + N^j dt) \tag{3.2}$$

where $N(t, x)$ is called the *lapse function* (sometimes denoted α) and $N^i(t, x)$ the *shift vector* (sometimes denoted β^i). The metric on a surface of constant t is $h_{ij}(t, x)$.

3.2 Extrinsic curvature

In GR, not only do we need to determine the metric tensor but we also need to determine the spacetime on which this tensor is defined. So it is not obvious what constitutes a suitable set of initial data for solving Einstein's equation. However, it seems likely that we will want to prescribe data on an "initial" hypersurface Σ which should correspond to a "moment of time". This we interpret as the requirement that Σ should be a *spacelike hypersurface*:

Definition. A hypersurface Σ is spacelike if its normal 1-form n_a is everywhere timelike. (A vector X^a is tangent to Σ iff $n_a X^a = 0$, which implies that X^a is spacelike.)

What data should be prescribed on Σ? Since the Einstein equation, like the Klein-Gordon equation, is second order in derivatives, one would expect that prescribing the spacetime metric and the "time derivative of the metric" on Σ should be enough. In fact, it turns out that we do not need to prescribe a full spacetime metric tensor on Σ, but only a Riemannian metric h_{ab} describing the intrinsic geometry of Σ, obtained from the spacetime metric by pull-back. A notion of "time derivative of the metric" on Σ is provided by the *extrinsic curvature* tensor of Σ, which we will now introduce. First we let n_a be the normal 1-form to Σ, which we assume to have unit norm:

$$n_a n^a = -1 \tag{3.3}$$

We now define the *projection* tensor onto Σ as

$$h^a_b = \delta^a_b + n^a n_b \tag{3.4}$$

Note that $h_{ab} = g_{ab} + n_a n_b$ is symmetric which means that it doesn't matter whether we write $h^a{}_b$ or h_b^a. If X^a and Y^a are tangent to Σ then $h_{ab} X^a Y^b = g_{ab} X^a Y^b$, so h_{ab} can be interpreted as the metric induced on Σ (the pull-back of g_{ab} to Σ). This is sometimes called the *first fundamental form* of Σ.

It is easy to check that

$$h^a_b n^b = 0 \qquad h^a_c h^c_b = h^a_b \tag{3.5}$$

which shows that h^a_b is a projection onto Σ. We can decompose any vector on Σ as

$$X^a = \delta^a_b X^b = h^a_b X^b - n^a n_b X^b \equiv X^a_\parallel + X^a_\perp \tag{3.6}$$

where $X^a_\parallel = h^a_b X^b$ is tangent to Σ and $X^a_\perp = -n_b X^b n^a$ is normal to Σ.

Let N_a be normal to Σ at p and consider parallel transport of N_a along a curve in Σ with tangent vector X^a, i.e., $X^b \nabla_b N_a = 0$. Does N_a remain normal to Σ? To answer this, let Y^a be another vector tangent to Σ so $Y^a N_a = 0$ at p. Consider how $Y^a N_a$ varies along the curve: $X(Y^a N_a) = X^b \nabla_b (Y^a N_a) = N_a X^b \nabla_b Y^a$. So $Y^a N_a$ vanishes along the curve iff the RHS vanishes. So if parallel transport within Σ preserves the property of being normal to Σ then $(\nabla_X Y)_\perp = 0$ for any X, Y tangent to Σ. The converse is also true. This motivates the following:

Definition. Up to now, n_a has been defined defined only on Σ so first extend it to a neighbourhood of Σ in an arbitrary way (with unit norm). The *extrinsic curvature tensor* (also called the *second fundamental form*) K_{ab} is defined at $p \in \Sigma$ by $K(X, Y) = -n_a \left(\nabla_{X_\parallel} Y_\parallel\right)^a$ where X, Y are vector fields on M.

Lemma.

$$K_{ab} = h^c_a h^d_b \nabla_c n_d \tag{3.7}$$

and K_{ab} is independent of how n_a is extended.

Proof. The RHS of the definition of $K(X, Y)$ is

$$-n_d X^c_\parallel \nabla_c Y^d_\parallel = X^c_\parallel Y^d_\parallel \nabla_c n_d = h^c_a X^a h^d_b Y^b \nabla_c n_d \tag{3.8}$$

where we used $n_d Y^d_\parallel = 0$ in the first equality. The final expression is linear in X^a and Y^b so the result follows. To demonstrate that the result is independent of how n_a is extended, consider a different extension n'_a, and let $m_a = n'_a - n_a$ so $m_a = 0$ on Σ. Then, on Σ,

$$X^a Y^b \left(K'_{ab} - K_{ab}\right) = X^c_\parallel Y^d_\parallel \nabla_c m_d = \nabla_{X_\parallel}(Y^d_\parallel m_d) = X_\parallel (Y^d_\parallel m_d) = 0 \tag{3.9}$$

where the second equality uses $m_a = 0$ on Σ and the final equality follows because it is the derivative along a curve tangent to Σ, along which $m_a = 0$.

Remark. $n^b\nabla_c n_b = (1/2)\nabla_c(n_b n^b) = 0$ because $n_b n^b = -1$. Hence we can also write

$$K_{ab} = h_a^c \nabla_c n_b \tag{3.10}$$

Lemma. $K_{ab} = K_{ba}$.

Proof. Let Σ be a surface of constant f with $df \neq 0$ on Σ. Recall a result from section 1.3: on Σ we have $n_a = g(df)_a$ for some function g chosen to make n_a a unit covector. We can extend n_a to a neighbourhood of Σ using this formula. We now have $\nabla_c n_d = g\nabla_c\nabla_d f + (\nabla_c g)g^{-1}n_d$ and so $K_{ab} = g h_a^c h_b^d \nabla_c\nabla_d f$, which is symmetric.

Lemma. K_{ab} is related to the Lie derivative of h_{ab} along n^a by

$$K_{ab} = \frac{1}{2}\mathcal{L}_n h_{ab} \tag{3.11}$$

Proof. Examples sheet 2.

This equation explains why K_{ab} can be interpreted as "the time derivative of the metric on Σ."

3.3 The Gauss-Codacci equations

A tensor at $p \in \Sigma$ is invariant under the projection h_b^a if

$$T^{a_1\ldots a_r}{}_{b_1\ldots b_s} = h^{a_1}_{c_1}\ldots h^{a_r}_{c_r} h^{d_1}_{b_1}\ldots h^{d_s}_{b_s} T^{c_1\ldots c_r}{}_{d_1\ldots d_s} \tag{3.12}$$

Tensors at p which are invariant under projection can be identified with tensors defined on the submanifold Σ, at p and vice-versa. (See Hawking and Ellis for more details on this correspondence.)

Proposition. A covariant derivative D on Σ can be defined by projection of the covariant derivative on M: for any tensor obeying (3.12) we define

$$D_a T^{b_1\ldots b_r}{}_{c_1\ldots c_s} = h_a^d h^{b_1}_{e_1}\ldots h^{b_r}_{e_r} h^{f_1}_{c_1}\ldots h^{f_s}_{c_s}\nabla_d T^{e_1\ldots e_r}{}_{f_1\ldots f_s} \tag{3.13}$$

Lemma. D is the Levi-Civita connection associated to the metric h_{ab} on Σ: $D_a h_{bc} = 0$ and D is torsion-free.

Proof. $\nabla_a h_{bc} = n_c\nabla_a n_b + n_b\nabla_a n_c$. Acting with the projections kills both terms. To prove the torsion-free property, let $f : \Sigma \to \mathbb{R}$ and extend to a function $f : M \to \mathbb{R}$.

$$D_a D_b f = h_a^c h_b^d \nabla_c\left(h_d^e \nabla_e f\right) = h_a^c h_b^e \nabla_c\nabla_e f + \left(h_a^c h_b^d \nabla_c h_d^e\right)\nabla_e f \tag{3.14}$$

The first term is symmetric (because ∇ is torsion-free). The second term involves

$$h_a^c h_b^d \nabla_c h_d^e = g^{ef} h_a^c h_b^d \nabla_c h_{df} = g^{ef} h_a^c h_b^d n_f \nabla_c n_d = n^e K_{ab} \tag{3.15}$$

which is also symmetric on a, b. Hence $D_a D_b f$ is symmetric so D is torsion-free.

We can now calculate the Riemann tensor associated to D, which measures the *intrinsic* curvature of Σ. The following result shows that this can be written in terms of the Riemann tensor of ∇ and the extrinsic curvature of Σ.

Proposition. Denote the Riemann tensor associated to D_a on Σ as $R'^a{}_{bcd}$. This is given by *Gauss' equation*:

$$R'^a{}_{bcd} = h_e^a h_b^f h_c^g h_d^h R^e{}_{fgh} - 2K_{[c}{}^a K_{d]b} \tag{3.16}$$

Proof. Let X^a be tangent to Σ. The Ricci identity for D is

$$R'^a{}_{bcd} X^b = 2D_{[c} D_{d]} X^a \tag{3.17}$$

Let's calculate the RHS

$$\begin{aligned} D_c D_d X^a &= h_c^e h_d^f h_g^a \nabla_e \left(D_f X^g\right) \\ &= h_c^e h_d^f h_g^a \nabla_e \left(h_f^h h_i^g \nabla_h X^i\right) \\ &= h_c^e h_d^h h_i^a \nabla_e \nabla_h X^i + h_c^e h_d^f h_i^a \left(\nabla_e h_f^h\right) \nabla_h X^i + h_c^e h_d^h h_g^a \left(\nabla_e h_i^g\right) \nabla_h X^i \\ &= h_c^e h_d^f h_g^a \nabla_e \nabla_f X^g + K_{cd} h_i^a n^h \nabla_h X^i + K_c{}^a n_i h_d^h \nabla_h X^i \end{aligned} \tag{3.18}$$

where we used (3.15) in the final two terms. The final term can be written

$$K_c{}^a h_d^h \nabla_h (n_i X^i) - K_c{}^a X^i h_d^h \nabla_h n_i = -K_c{}^a X^b h_b^i h_d^h \nabla_h n_i = -K_c{}^a K_{bd} X^b \tag{3.19}$$

where we used $X^i = h_b^i X^b$ because X^a is tangent to Σ. We can now plug (3.18) into (3.17): the second term on the RHS drops out when we antisymmetrize, leaving

$$R'^a{}_{bcd} X^b = 2h_{[c}^e h_{d]}^f h_g^a \nabla_e \nabla_f X^g - 2K_{[c}{}^a K_{d]b} X^b \tag{3.20}$$

The first term can be written

$$2h_c^e h_d^f h_g^a \nabla_{[e} \nabla_{f]} X^g = h_c^e h_d^f h_g^a R^g{}_{hef} X^h = h_c^e h_d^f h_g^a h_b^h R^g{}_{hef} X^b \tag{3.21}$$

where we used the Ricci identity for ∇ in the first equality and the fact that X^a is parallel to Σ in the second. Moving everything to the LHS now gives

$$\left(R'^a{}_{bcd} - h_c^e h_d^f h_g^a h_b^h R^g{}_{hef} + 2K_{[c}{}^a K_{d]b}\right) X^b = 0 \tag{3.22}$$

The expression in brackets is invariant under projection onto Σ and hence can be identified with a tensor on Σ. X^b is an arbitrary vector parallel to Σ. It follows that the expression in brackets must vanish. The result follows upon relabelling dummy indices.

Lemma. The Ricci scalar of Σ is

$$R' = R + 2R_{ab}n^a n^b - K^2 + K^{ab}K_{ab} \tag{3.23}$$

where $K \equiv K^a{}_a$.

Proof. $R' = h^{bd}R'^c{}_{bcd}$ (since h^{bd} can be identified with the inverse metric on Σ). Now use Gauss' equation.

Proposition. (Codacci's equation).

$$D_a K_{bc} - D_b K_{ac} = h_a^d h_b^e h_c^f n^g R_{defg} \tag{3.24}$$

Proof.

$$\begin{aligned} D_a K_{bc} &= h_a^d h_b^g h_c^f \nabla_d K_{gf} \\ &= h_a^d h_b^g h_c^f \nabla_d \left(h_g^e \nabla_e n_f \right) \\ &= h_a^d h_b^g h_c^f h_g^e \nabla_d \nabla_e n_f + h_a^d h_b^g h_c^f (\nabla_d h_g^e) \nabla_e n_f \\ &= h_a^d h_b^e h_c^f \nabla_d \nabla_e n_f + K_{ab} n^e h_c^f \nabla_e n_f \end{aligned} \tag{3.25}$$

where we used (3.15) in the final line. Antisymmetrizing on a, b now gives

$$2D_{[a}K_{b]c} = 2h^d_{[a}h^e_{b]}h^f_c \nabla_d \nabla_e n_f = 2h_a^d h_b^e h_c^f \nabla_{[d}\nabla_{e]} n_f = h_a^d h_b^e h_c^f R_{fgde} n^g \tag{3.26}$$

The result follows using $R_{defg} = R_{fgde}$.

Lemma.

$$D_a K^a{}_b - D_b K = h_b^c R_{cd} n^d \tag{3.27}$$

Proof. Contract Codacci's equation with h^{ac}.

Some people refer to equation (3.27) as Codacci's equation.

3.4 The constraint equations

Consider the "normal-normal" component of the Einstein equation, i.e., contract $G_{ab} = 8\pi T_{ab}$ with $n^a n^b$. On the LHS we get

$$G_{ab}n^a n^b = R_{ab}n^a n^b + \frac{1}{2}R = \frac{1}{2}\left(R' - K^{ab}K_{ab} + K^2\right) \tag{3.28}$$

where we have used (3.23) in the final step. Therefore we have

$$R' - K^{ab}K_{ab} + K^2 = 16\pi\rho \tag{3.29}$$

where $\rho \equiv T_{ab}n^a n^b$ is the matter energy density measured by an observer with 4-velocity n^a. R' is determined by the metric on Σ, i.e., by h_{ab}. This equation reveals that we are not free to specify h_{ab} and K_{ab} arbitrarily on Σ: they must be related by this equation, which is called the *Hamiltonian constraint*.

Now consider the "normal-tangential" components of the Einstein equation by contracting it with n^a and then projecting onto Σ:

$$8\pi h_a^b T_{bc} n^c = h_a^b G_{bc} n^c = h_a^b R_{bc} n^c \tag{3.30}$$

Using (3.27) we have

$$D_b K^b{}_a - D_a K = 8\pi h_a^b T_{bc} n^c \tag{3.31}$$

Note that the RHS is (8π times) minus the momentum density measured by an observer with 4-velocity n^a. The LHS of this equation involves only the metric on Σ and K_{ab} so this is another constraint equation, called the *momentum constraint*.

The constraint equations involve the metric h_{ab} on Σ and its "time-derivative" $K_{ab} = (1/2)\mathcal{L}_n h_{ab}$. The remaining components of the Einstein equation, i.e., those tangential to Σ, involve second time derivatives of h_{ab}. More precisely, they involve $\mathcal{L}_n \mathcal{L}_n h_{ab}$. These components are *evolution equations*, which determine how to evolve the initial data "forward in time".

3.5 The initial value problem in GR

Initial data for Einstein's equation consists of a triple (Σ, h_{ab}, K_{ab}) where (Σ, h_{ab}) is a Riemannian 3-manifold and K_{ab} is a symmetric tensor. The idea is that Σ corresponds to a spacelike hypersurface in spacetime, h_{ab} is the pull-back of the spacetime metric to Σ, and K_{ab} is the extrinsic curvature tensor of Σ, i.e., the "rate of change" of the metric on Σ. The initial data is not completely free: the Einstein equation implies that it must satisfy the constraint equations.

The following result is of fundamental significance in GR:

Theorem (Choquet-Bruhat & Geroch 1969). Let (Σ, h_{ab}, K_{ab}) be initial data satisfying the vacuum Hamiltonian and momentum constraints (i.e. equations (3.29,3.31) with vanishing RHS). Then there exists a unique (up to diffeomorphism) spacetime (M, g_{ab}), called the *maximal Cauchy development* of (Σ, h_{ab}, K_{ab}) such that (i) (M, g_{ab}) satisfies the vacuum Einstein equation; (ii) (M, g_{ab}) is globally hyperbolic with Cauchy surface Σ; (iii) The induced metric and extrinsic curvature of Σ are h_{ab} and K_{ab} respectively; (iv) Any other spacetime satisfying (i),(ii),(iii) is isometric to a subset of (M, g_{ab}).

Analogous theorems exist in the non-vacuum case for suitable matter e.g. a perfect fluid or tensor fields whose equations of motion are hyperbolic partial differential equations (e.g. Maxwell field, scalar field, perfect fluid).

It is possible that the maximal Cauchy development (M, g_{ab}) is *extendible*, i.e., isometric to a proper subset of another spacetime (M', g'_{ab}). By the above theorem, Σ cannot be a Cauchy surface for (M', g'_{ab}). Instead we will have $M = D(\Sigma) \subset M'$, and the boundaries of $D(\Sigma)$ in M' will be future/past Cauchy horizons. If this happens then we cannot predict physics in $M' \backslash D(\Sigma)$ from the initial data on Σ. In particular, we cannot say determine the metric g'_{ab} in $M' \backslash D(\Sigma)$: there will be infinitely possible solutions of the vacuum Einstein equation that are consistent with the initial data on Σ. Let's look at some examples for which this happens.

First consider initial data given by a surface $\Sigma = \{(x, y, z) : x > 0\}$ with flat 3-metric $\delta_{\mu\nu}$ and vanishing extrinsic curvature. The maximal development of this initial data is the region $|t| < x$ of Minkowski spacetime, which is extendible. Outside this region, we cannot predict the spacetime, in particular it need not be flat. In this example we could have anticipated that the maximal development would be extendible because the initial data is extendible (to $x \leq 0$). If we are given initial conditions only in part of space then we do not expect to be able to predict the entire spacetime.

Now consider the Schwarzschild solution with $M < 0$:

$$ds^2 = -\left(1 + \frac{2|M|}{r}\right) dt^2 + \left(1 + \frac{2|M|}{r}\right)^{-1} dr^2 + r^2 d\Omega^2 \tag{3.32}$$

This solution has a curvature singularity at $r = 0$ but no event horizon. Let (Σ, h_{ab}, K_{ab}) be the data on a surface $t = 0$ in this spacetime (in fact $K_{ab} = 0$). In this case, (Σ, h_{ab}) is inextendible. However, viewed as a Riemannian manifold, (Σ, h_{ab}) is not geodesically complete because some of its geodesics have $r \to 0$ in finite affine parameter. So in this case, the initial data is "singular".

The resulting maximal development is not the whole $M < 0$ Schwarzschild spacetime. This is because some inextendible causal curves do not intersect Σ. For example, consider an outgoing radial null geodesic, which satisfies

$$\frac{dt}{dr} = \left(1 + \frac{2|M|}{r}\right)^{-1} = \frac{r}{r + 2|M|} \approx \frac{r}{2|M|} \qquad \text{at small } r \tag{3.33}$$

hence $t \approx t_0 + r^2/(4|M|)$ at small r so t has a finite limit t_0 as $r \to 0$. So this null geodesic emerges from the singularity at time t_0 and then has $t > t_0$. Hence if

$t_0 > 0$ then this geodesic does not intersect Σ so Σ is not a Cauchy surface for the full spacetime. One can show that the boundary of $D(\Sigma)$ is given precisely by those radial null geodesics which have $t_0 = 0$, i.e., they "emerge from the singularity on Σ": see Fig. 16.

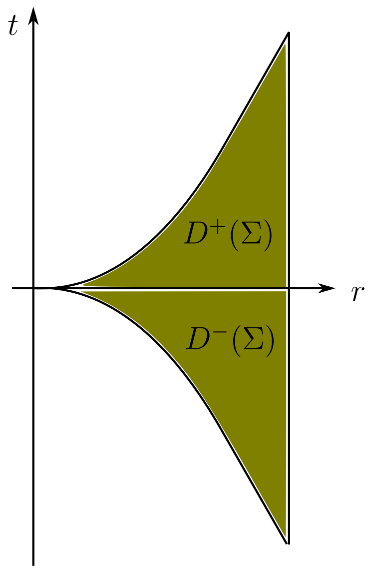


Figure 16. Domain of dependence of $t = 0$ surface in negative M Schwarzschild geometry.

We emphasize that the solution outside $D(\Sigma)$ is not determined by the initial data on Σ. The data on Σ does *not* predict that the solution outside $D(\Sigma)$ must coincide with the $M < 0$ Schwarzschild solution. This is just one possibility amongst infinitely many alternatives. These alternatives cannot be spherically symmetric because of Birkhoff's theorem.

In this case, the extendibility of the maximal development arises because the initial data is singular (not geodesically complete) and one "can't predict what comes out of a singularity". Henceforth we will restrict to initial data which is geodesically complete (and therefore also inextendible).

Even when (Σ, h_{ab}) is geodesically complete, the maximal development may be extendible. For example, let Σ be the hyperboloid $-t^2 + x^2 + y^2 + z^2 = -1$ with $t < 0$ in Minkowski spacetime (Fig. 17). Take h_{ab} to be the induced metric and K_{ab} the extrinsic curvature of this surface. Clearly there are inextendible null curves in Minkowski spacetime which do not intersect Σ. The maximal development of the initial data on Σ is the interior of the past light cone through the origin in Minkowsi spacetime. In this case, the maximal development is extendible because the initial data surface is "asymptotically null" so "information can arrive from infinity". Extensions of

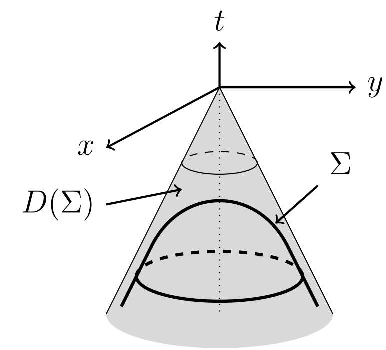


Figure 17. Hyperboloidal initial data surface in Minkowski spacetime.

the maximal development need not be flat in $M \backslash D(\Sigma)$, even if they satisfy the vacuum Einstein equation there. Such extensions include solutions describing gravitational waves being sent in from infinity to the future of the Cauchy horizon.

3.6 Asymptotically flat initial data

To avoid all of these ways in which $D(\Sigma)$ is extendible, we will restrict to geodesically complete initial data which is "asymptotically flat" in the sense that, at large distance, it looks like a surface of constant t in Minkowski spacetime. (Recall that such surfaces are Cauchy surfaces for Minkowski spacetime.) We also want to allow for the possibility of having several asymptotically flat regions, as in the Kruskal spacetime.

Definition. (a) An initial data set (Σ, h_{ab}, K_{ab}) is an *asymptotically flat end* if (i) Σ is diffeomorphic to $\mathbb{R}^3 \backslash B$ where B is a closed ball centred on the origin in $\mathbb{R}^3$; (ii) if we pull-back the $\mathbb{R}^3$ coordinates to define coordinates x^i on Σ then $h_{ij} = \delta_{ij} + \mathcal{O}(1/r)$ and $K_{ij} = \mathcal{O}(1/r^2)$ as $r \to \infty$ where $r = \sqrt{x^i x^i}$ (iii) derivatives of the latter expressions also hold e.g. $h_{ij,k} = \mathcal{O}(1/r^2)$ etc. (If matter fields are present then these should also decay at a suitable rate at large r.)

(b) An initial data set is *asymptotically flat with N ends* if it is the union of a compact set with N asymptotically flat ends.

For example, in the $(M > 0)$ Schwarzschild solution consider the surface $\Sigma = \{t = \text{constant}, r > 2M\}$. On examples sheet 2 it is shown that this data is an asymptotically flat end. Of course this initial data is not geodesically complete (since it stops at $r = 2M$). But now consider the Kruskal spacetime. Then Σ corresponds to part of an Einstein-Rosen bridge. The full Einstein-Rosen bridge is asymptotically flat with 2 ends. This is because it is the union of the bifurcation sphere $U = V = 0$ (a compact set) with two copies of the asymptotically flat end just discussed (one in region I and one in region IV).

3.7 Strong cosmic censorship

For geodesically complete, asymptotically flat, initial data it would be very disturbing if the maximal Cauchy development were extendible. It would imply that GR suffers from a lack of determinism (predictability). The *strong cosmic censorship conjecture* asserts that this does *not* happen:

Strong cosmic censorship conjecture (Penrose). Let (Σ, h_{ab}, K_{ab}) be a geodesically complete, asymptotically flat (with N ends), initial data set for the vacuum Einstein equation. Then generically the maximal Cauchy development of this initial data is inextendible.

This conjecture is known to be correct for initial data which is sufficiently close to initial data for Minkowski spacetime. For such data, a theorem of Christdoulou and Klainerman (1994) asserts that the resulting spacetime "settles down to Minkowski spacetime at late time". In more physical terms, it says that Minkowski spacetime is stable against small gravitational perturbations. The spacetime has no Cauchy horizon so strong cosmic censorship is true for such initial data.

The word "generically" is included because of known counter-examples. Later we will discuss charged and rotating black hole solutions and find that they exhibit a Cauchy horizon (for a geodesically complete, asymptotically flat initial data set) inside the black hole. However, this is believed to be unstable in the sense that an arbitrarily small perturbation of this initial data has an inextendible maximal development. More formally, if one introduces some measure on the space of geodesically complete, asymptotically flat, initial data, strong cosmic censorship asserts that the maximal development is inextendible except for a set of initial data of measure zero.

The above conjecture can be extended to include matter. We need to assume that the matter is such that the Choquet-Bruhat-Geroch theorem applies, as will be the case if the matter fields satisfy hyperbolic equations of motion. We also restrict to matter that is "physical" in the sense that it has positive energy density and does not travel faster than light. We do this by imposing the *dominant energy condition* (to be discussed later). This condition is satisfied by all "normal" matter. One also has to deal with the fact that some types of matter (e.g. a perfect fluid) can form singularities (shocks) even in the absence of gravity and so the conjecture is most straightforward to formulate for matter that doesn't do this e.g. a Maxwell field or Klein-Gordon field.

Proving the strong cosmic censorship conjecture, and the related weak cosmic censorship conjecture, is one of the main goals of mathematical relativity.

4 The singularity theorem

We have seen how spherically symmetric gravitational collapse results in the formation of a singularity. But maybe this is just a consequence of spherical symmetry. For example, in Newtonian theory, spherically symmetric collapse of a ball of matter produces a "singularity", i.e., infinite density at the origin. But this does not happen without spherical symmetry. In this case, the singularity is non-generic: a tiny perturbation (breaking spherical symmetry) of the initial state results in a "bouncing" solution without a singularity. Could the same be true in GR? No. In this chapter we will discuss the Penrose singularity theorem, which shows that singularities are a generic prediction of GR.

4.1 Null hypersurfaces

Definition. A *null hypersurface* is a hypersurface whose normal is everywhere null.

Example. Consider surfaces of constant r in the Schwarzschild spacetime. The 1-form $n = dr$ is normal to such surfaces. Using ingoing Eddington-Finkelstein coordinates, the inverse metric is

$$g^{\mu\nu} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 1-\frac{2M}{r} & 0 & 0 \\ 0 & 0 & \frac{1}{r^2} & 0 \\ 0 & 0 & 0 & \frac{1}{r^2\sin^2\theta} \end{pmatrix} \tag{4.1}$$

hence

$$n^2 \equiv g^{\mu\nu} n_\mu n_\nu = g^{rr} = 1 - \frac{2M}{r} \tag{4.2}$$

so the surface $r = 2M$ is a null hypersurface. Since $n^\mu = g^{\mu\nu} n_\nu = g^{\mu r}$ we have

$$n^a|_{r=2M} = \left(\frac{\partial}{\partial v}\right)^a \tag{4.3}$$

Let n_a be normal to a null hypersurface $\mathcal{N}$. Then any (non-zero) vector X^a tangent to the hypersurface obeys $n_a X^a = 0$ which implies that either X^a is spacelike or X^a is parallel to n^a. In particular, note that n^a is tangent to the hypersurface. Hence, on $\mathcal{N}$, the integral curves of n^a lie within $\mathcal{N}$.

Proposition. The integral curves of n^a are null geodesics. These are called the *generators* of $\mathcal{N}$.

Proof. Let $\mathcal{N}$ be given by an equation $f = $ constant for some function f with $df \neq 0$ on $\mathcal{N}$. Then we have $n = h df$ for some function h. Let $N = df$. The integral curves of n^a and N^a are the same up to a choice of parameterization.

Then since $\mathcal{N}$ is null we have that $N^a N_a = 0$ on $\mathcal{N}$. Hence the function $N^a N_a$ is constant on $\mathcal{N}$ which implies that the gradient of this function is normal to $\mathcal{N}$:

$$\nabla_a \left(N^b N_b\right)|_{\mathcal{N}} = 2\alpha N_a \tag{4.4}$$

for some function α on $\mathcal{N}$. Now we also have $\nabla_a N_b = \nabla_a \nabla_b f = \nabla_b \nabla_a f = \nabla_b N_a$. So the LHS above is $2N^b \nabla_a N_b = 2N^b \nabla_b N_a$. Hence we have

$$N^b \nabla_b N_a|_{\mathcal{N}} = \alpha N_a \tag{4.5}$$

which is the geodesic equation for a non-affinely parameterized geodesic. Hence, on $\mathcal{N}$, the integral curves of N^a (and therefore also n^a) are null geodesics. □

Example. In the Kruskal spacetime, let $N = dU$ which is null *everywhere* ($g^{UU} = 0$) and normal to a *family* of null hypersurfaces ($U =$ constant), which gives

$$N^b \nabla_b N_a = N^b \nabla_b \nabla_a U = N^b \nabla_a \nabla_b U = N^b \nabla_a N_b = (1/2)\nabla_a (N^2) = 0 \tag{4.6}$$

so in this case N^a is tangent to *affinely parameterized* null geodesics. Raising an index gives (exercise)

$$N^a = -\frac{r}{16M^3} e^{r/(2M)} \left(\frac{\partial}{\partial V}\right)^a \tag{4.7}$$

Now let $\mathcal{N}$ be the surface $U = 0$. Since $r = 2M$ on $\mathcal{N}$ we see that N^a is a constant multiple of $\partial/\partial V$. Hence V is an affine parameter for the generators of $\mathcal{N}$. Similarly U is an affine parameter for the generators of the null hypersurface $V = 0$.

4.2 Geodesic deviation

You encountered the geodesic deviation equation in the GR course. Recall the following definitions:

Definition. A *1-parameter family of geodesics* is a map $\gamma : I \times I' \to M$ where I and I' both are open intervals in $\mathbb{R}$, such that (i) for fixed s, $\gamma(s, \lambda)$ is a geodesic with affine parameter λ (so s is the parameter that labels the geodesic); (ii) the map $(s, \lambda) \mapsto \gamma(s, \lambda)$ is smooth and one-to-one with a smooth inverse. This implies that the family of geodesics forms a 2d surface $\Sigma \subset M$.

Let U^a be the tangent vector to the geodesics and S^a to be the vector tangent to the curves of constant t, which are parameterized by s (see Fig. 18). In a chart x^μ, the geodesics are specified by $x^\mu(s, \lambda)$ with $S^\mu = \partial x^\mu / \partial s$. Hence $x^\mu(s + \delta s, \lambda) = x^\mu(s, \lambda) + \delta s S^\mu(s, \lambda) + \mathcal{O}(\delta s^2)$. Therefore $(\delta s) S^a$ points from one geodesic to an infinitesimally nearby one in the family. We call S^a a *deviation vector*.

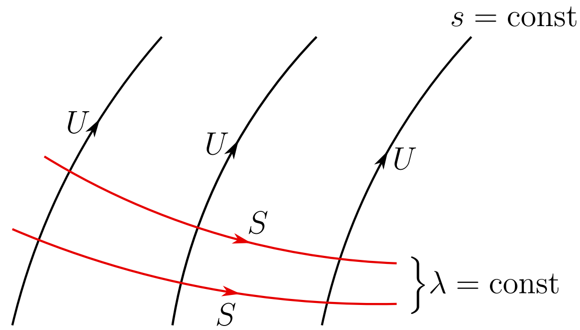


Figure 18. 1-parameter family of geodesics

In a neighbourhood of Σ we can use coordinates (s, λ, y^1, y^2) for suitable y^1, y^2. This gives a coordinate chart in which $S = \partial/\partial s$ and $U = \partial/\partial\lambda$ on Σ. Hence S^a and U^a commute:

$$[S, U] = 0 \qquad \Leftrightarrow \qquad U^b \nabla_b S^a = S^b \nabla_b U^a \tag{4.8}$$

Recall that this implies that S^a satisfies the *geodesic deviation equation*

$$U^c \nabla_c (U^b \nabla_b S^a) = R^a{}_{bcd} U^b U^c S^d \tag{4.9}$$

Given an affinely parameterized geodesic γ with tangent U^a, a solution S^a of this equation along γ is called a *Jacobi field*.

4.3 Geodesic congruences

Definition. Let $\mathcal{U} \subset M$ be open. A *geodesic congruence* in $\mathcal{U}$ is a family of geodesics such that exactly one geodesic passes through each $p \in \mathcal{U}$.

We will consider a congruence for which all the geodesics are of the same type (timelike or spacelike or null). Then by normalizing the affine parameter we can arrange that the tangent vector U^a satisfies $U^2 = \pm 1$ (in the spacelike or timelike case) or $U^2 = 0$ (in the null case) everywhere.

Now consider a 1-parameter family of geodesics belonging to a congruence. Write (4.8) as

$$U^b \nabla_b S^a = B^a{}_b S^b \tag{4.10}$$

where

$$B^a{}_b = \nabla_b U^a \tag{4.11}$$

measures the failure of S^a to be parallelly transported along the geodesic with tangent U^a. Note that

$$B^a{}_b U^b = 0 \tag{4.12}$$

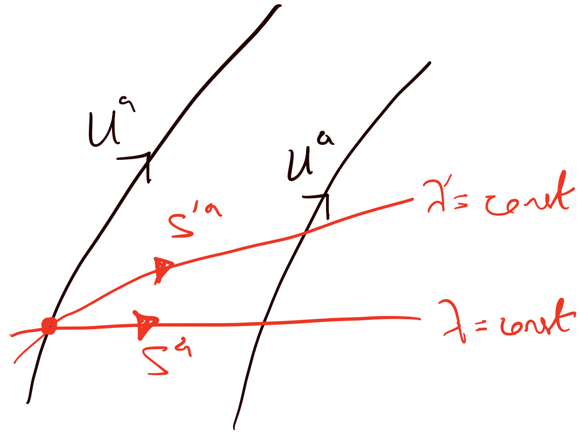


Figure 19. Shifting the affine parameter changes the deviation vector.

because U^a is tangent to affinely parameterized geodesics. Note also that

$$U_a B^a{}_b = \frac{1}{2}\nabla_b(U^2) = 0 \tag{4.13}$$

because we've arranged that U^2 is constant throughout $\mathcal{U}$. This implies that

$$U \cdot \nabla(U \cdot S) = (U \cdot \nabla U^a)S_a + U^a U \cdot \nabla S_a = U^a B_{ab} S^b = 0 \tag{4.14}$$

using the geodesic equation and (4.10). Hence $U \cdot S$ is constant along any geodesic in the congruence.

Now recall that, even after normalising so that $U^2 \in \{\pm 1, 0\}$, the affine parameter is not uniquely defined because we are free to shift it by a constant. We can choose this constant to be different on different geodesics, i.e., it can depend on s: $\lambda' = \lambda - a(s)$ is just as good an affine parameter as λ. But this changes the deviation vector to (exercise)

$$S'^a \equiv S^a + \frac{da}{ds}U^a \tag{4.15}$$

Hence S'^a is a deviation vector pointing to the *same* geodesic as S^a: Fig. 19. Now $U \cdot S' = U \cdot S + (da/ds)U^2$ so in the spacelike or timelike case, we can fix this "gauge freedom" by choosing $a(s)$ so that $U \cdot S = 0$ at some point on each geodesic (e.g. $\lambda = 0$). Since $U \cdot S$ is constant along each geodesic, this implies that $U \cdot S = 0$ everywhere.

4.4 Null geodesic congruences

In the null case, the above procedure does not work because $U \cdot S' = U \cdot S$. Instead we fix the gauge freedom as follows. Pick a spacelike hypersurface Σ which intersects each geodesic once. Let N^a be a vector field defined on Σ obeying $N^2 = 0$ and $N \cdot U = -1$ on Σ. Now extend N^a off Σ by parallel transport along the geodesics: $U \cdot \nabla N^a = 0$. This implies $N^2 = 0$ and $N \cdot U = -1$ everywhere (proof: exercise). In summary, we've constructed a vector field such that

$$N^2 = 0 \qquad U \cdot N = -1 \qquad U \cdot \nabla N^a = 0 \tag{4.16}$$

We can now decompose any deviation vector uniquely as

$$S^a = \alpha U^a + \beta N^a + \hat{S}^a \tag{4.17}$$

where

$$U \cdot \hat{S} = N \cdot \hat{S} = 0 \tag{4.18}$$

which implies that $\hat{S}^a$ is spacelike (or zero). Note that $U \cdot S = -\beta$ hence β is constant along each geodesic. So we can write a deviation vector S^a the sum of a part $\alpha U^a + \hat{S}^a$ orthogonal to U^a and a part βN^a that is parallelly transported along each geodesic.

An important case is when the congruence contains the generators of a null hypersurface $\mathcal{N}$ and we are interested only in the behaviour of these generators. In this case, if we pick a 1-parameter family of geodesics contained within $\mathcal{N}$ then the deviation vector S^a will be tangent to $\mathcal{N}$ and hence obey $U \cdot S = 0$ (since U^a is normal to $\mathcal{N}$) i.e. $\beta = 0$.

Note that we can write

$$\hat{S}^a = P^a_b S^b \tag{4.19}$$

where

$$P^a_b = \delta^a_b + N^a U_b + U^a N_b \tag{4.20}$$

is a *projection* (i.e. $P^a_b P^b_c = P^a_c$) of the tangent space at p onto $T_\perp$, the 2d space of vectors at p orthogonal to U^a and N^a. Since U^a and N^a are both parallelly transported, so is P^a_b:

$$U \cdot \nabla P^a_b = 0 \tag{4.21}$$

Proposition. A deviation vector for which $U \cdot S = 0$ satisfies $U \cdot \nabla \hat{S}^a = \hat{B}^a{}_b \hat{S}^b$ where $\hat{B}^a{}_b = P^a_c B^c{}_d P^d_b$.

Proof. $U \cdot \nabla \hat{S}^a = U \cdot \nabla (P^a_c S^c) = P^a_c U \cdot \nabla S^c = P^a_c B^c{}_d S^d = P^a_c B^c{}_d P^d_e S^e$ using $U \cdot S = 0$ and $B^c{}_d U^d = 0$ in the final step. Finally we can use $P^2 = P$ to write the RHS as $P^a_c B^c{}_d P^d_b P^b_e S^e = \hat{B}^a{}_b \hat{S}^b$.

4.5 Expansion, rotation and shear

Note that $\hat{B}^a{}_b$ can be regarded as a matrix that acts on the 2d space $T_\perp$. To understand its geometrical interpretation, it is useful to divide it into its trace, traceless symmetric, and antisymmetric parts as follows:

Definition. The *expansion, shear* and *rotation* of the null geodesic congruence are

$$\theta = \hat{B}^a{}_a \qquad \hat{\sigma}_{ab} = \hat{B}_{(ab)} - \frac{1}{2}P_{ab}\theta, \qquad \hat{\omega}_{ab} = \hat{B}_{[ab]} \tag{4.22}$$

This implies

$$\hat{B}^a{}_b = \frac{1}{2}\theta P^a_b + \hat{\sigma}^a{}_b + \hat{\omega}^a{}_b \tag{4.23}$$

Exercise. Show that $\theta = g^{ab}B_{ab} = \nabla_a U^a$.

This shows that the expansion is independent of the choice of N^a, i.e., it is an intrinsic property of the congruence. Scalar invariants of the rotation and shear (e.g. $\hat{\omega}_{ab}\hat{\omega}^{ab}$ or the eigenvalues of $\hat{\sigma}^a{}_b$) are also independent of the choice of N^a.

Proposition. If the congruence contains the generators of a (null) hypersurface $\mathcal{N}$ then $\hat{\omega}_{ab} = 0$ on $\mathcal{N}$. Conversely, if $\hat{\omega}_{ab} = 0$ everywhere then U^a is everywhere hypersurface orthogonal (i.e. orthogonal to a family of null hypersurfaces).

Proof. The definition of $\hat{B}$ and $U \cdot B = B \cdot U = 0$ implies

$$\hat{B}^b{}_c = B^b{}_c + U^b N_d B^d{}_c + U_c B^b{}_d N^d + U^b U_c N_d B^d{}_e N^e \tag{4.24}$$

Using this, we have

$$U_{[a}\hat{\omega}_{bc]} = U_{[a}\hat{B}_{bc]} = U_{[a}B_{bc]} \tag{4.25}$$

since the extra terms drop out of the antisymmetrization. Now using the definition of B_{ab} we have

$$U_{[a}\hat{\omega}_{bc]} = U_{[a}\nabla_c U_{b]} = -\frac{1}{6}\left(U \wedge dU\right)_{abc} \tag{4.26}$$

If U^a is normal to $\mathcal{N}$ then $U \wedge dU = 0$ on $\mathcal{N}$ and hence, on $\mathcal{N}$,

$$0 = U_{[a}\hat{\omega}_{bc]} = \frac{1}{3}\left(U_a\hat{\omega}_{bc} + U_b\hat{\omega}_{ca} + U_c\hat{\omega}_{ab}\right) \tag{4.27}$$

Contracting this with N^a gives $\hat{\omega}_{bc} = 0$ on $\mathcal{N}$ (using $U \cdot N = -1$ and $\hat{\omega} \cdot N = 0$). Conversely, if $\hat{\omega} = 0$ everywhere then (4.26) implies that U is hypersurface-orthogonal using Frobenius' theorem.

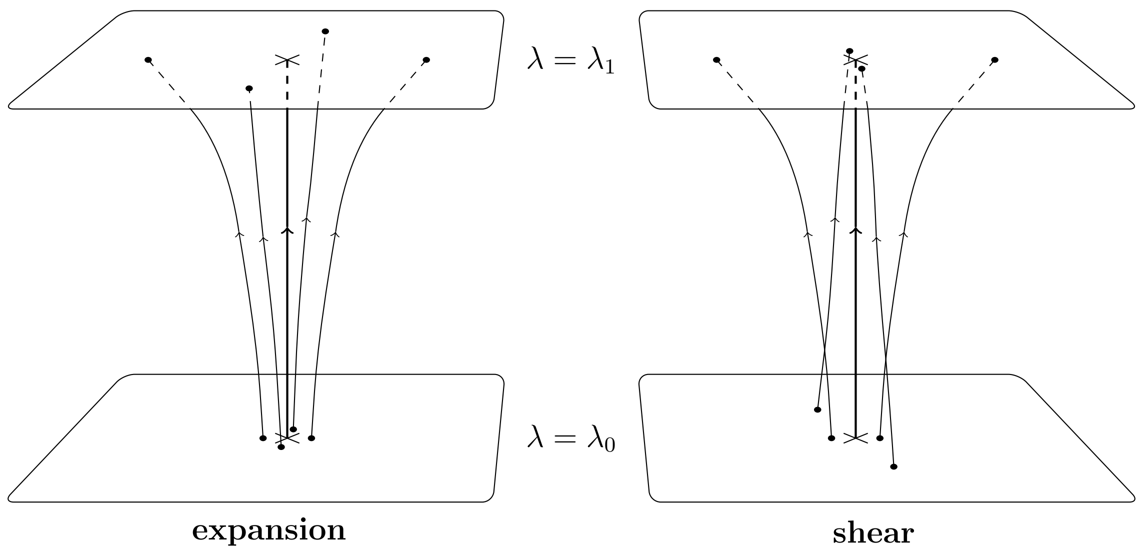


Figure 20. Effects of expansion and shear on the generators of a null hypersurface.

4.6 Expansion and shear of a null hypersurface

Assume that we have a congruence which includes the generators of a null hypersurface $\mathcal{N}$. The generators of $\mathcal{N}$ have $\hat{\omega} = 0$. To understand how these generators behave, restrict attention to deviation vectors tangent to $\mathcal{N}$ (i.e. consider a 1-parameter family of generators of $\mathcal{N}$). Consider the evolution of the generators of $\mathcal{N}$ as a function of affine parameter λ, as shown in Fig. 20.

Qualitatively: expansion θ corresponds to neighbouring generators moving apart (if $\theta > 0$) or together (if $\theta < 0$). Shear corresponds to geodesics moving apart in one direction, and together in the orthogonal direction whilst preserving the cross-sectional area.

We can make this more precise by introducing *Gaussian null coordinates* near $\mathcal{N}$ as follows (see Fig. 21). Pick a spacelike 2-surface S within $\mathcal{N}$ and let y^i ($i = 1, 2$) be coordinates on this surface. Assign coordinates (λ, y^i) to the point affine parameter distance λ from S along the generator of $\mathcal{N}$ (with tangent U^a) which intersects the surface S at the point with coordinates y^i. Now we have coordinates (λ, y^i) on $\mathcal{N}$ such that the generators are lines of constant y^i and $U^a = (\partial/\partial\lambda)^a$.

Let V^a be a null vector field on $\mathcal{N}$ satisfying $V \cdot \partial/\partial y^i = 0$ and $V \cdot U = 1$. Assign coordinates (r, λ, y^i) to the point affine parameter distance r along the null geodesic which starts at the point on $\mathcal{N}$ with coordinates (λ, y^i) and has tangent vector V^a there.

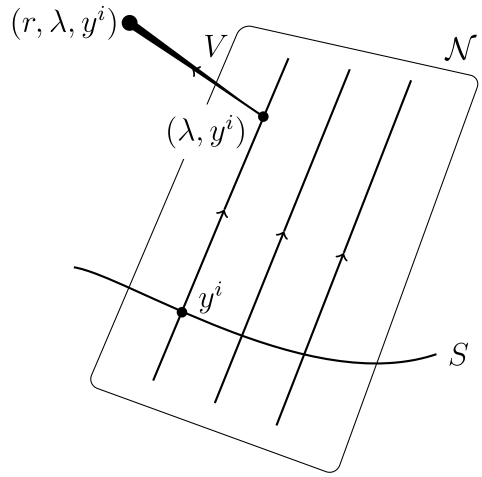


Figure 21. Construction of Gaussian null coordinates near a null hypersurface $\mathcal{N}$.

This defines a coordinate chart in a neighbourhood of $\mathcal{N}$ such that $\mathcal{N}$ is at $r = 0$, with $U = \partial/\partial\lambda$ on $\mathcal{N}$, and $\partial/\partial r$ is tangent to affinely parameterized null geodesics. The latter implies that $g_{rr} = 0$ everywhere.

Exercise. Use the geodesic equation for $\partial/\partial r$ to show $g_{r\mu,r} = 0$.

At $r = 0$ we have $g_{r\lambda} = V \cdot U = 1$ (as $V = \partial/\partial r$ on $\mathcal{N}$) and $g_{ri} = V \cdot (\partial/\partial y^i) = 0$. Since $g_{r\mu}$ is independent of r, these results are valid for all r. We also know that $g_{\lambda\lambda} = 0$ at $r = 0$ (as U^a is null) and $g_{\lambda i} = 0$ at $r = 0$ (as $\partial/\partial y^i$ is tangent to $\mathcal{N}$ and hence orthogonal to U^a). So we can write $g_{\lambda\lambda} = rF$ and $g_{\lambda i} = rh_i$ for some smooth functions F, h_i. Therefore the metric takes the form

$$ds^2 = 2drd\lambda + rFd\lambda^2 + 2rh_i d\lambda dy^i + h_{ij}dy^i dy^j \tag{4.28}$$

(We note that F must vanish at $r = 0$. To see this, we use the fact that the curves $\lambda \mapsto (0, \lambda, y^i)$, for constant y^i are affinely parameterized null geodesics: the generators of $\mathcal{N}$. For these the only-non vanishing component of the geodesic equation is the r component. This reduces to $\partial_r(rF) = 0$ Hence $F = 0$ at $r = 0$ so we can write $F = r\hat{F}$ for some smooth function $\hat{F}$.)

On $\mathcal{N}$ the metric is

$$g|_{\mathcal{N}} = 2drd\lambda + h_{ij}dy^i dy^j \tag{4.29}$$

so $U^\mu = (0, 1, 0, 0)$ on $\mathcal{N}$ implies that $U_\mu = (1, 0, 0, 0)$ on $\mathcal{N}$. Now $U \cdot B = B \cdot U = 0$ implies that $B^r{}_\mu = B^\mu{}_\lambda = 0$. We saw above that $\theta = B^\mu{}_\mu$. Hence on $\mathcal{N}$ we have

$$\theta = B^i{}_i = \nabla_i U^i = \partial_i U^i + \Gamma^i_{i\mu} U^\mu = \Gamma^i_{i\lambda} = \frac{1}{2} g^{i\mu} \left(g_{\mu i,\lambda} + g_{\mu\lambda,i} - g_{i\lambda,\mu} \right) \tag{4.30}$$

In the final expression, note that the form of the metric on $\mathcal{N}$ implies that $g^{i\mu}$ is non-vanishing only when $\mu = j$, and that $g^{ij} = h^{ij}$ (the inverse of h_{ij}) hence on $\mathcal{N}$

$$\theta = \frac{1}{2} h^{ij} \left(g_{ji,\lambda} + g_{j\lambda,i} - g_{i\lambda,j} \right) = \frac{1}{2} h^{ij} h_{ij,\lambda} = \frac{\partial_\lambda \sqrt{h}}{\sqrt{h}} \tag{4.31}$$

where we used $g_{i\lambda} = 0$ on $\mathcal{N}$ and defined $h = \det h_{ij}$. Hence we have

$$\frac{\partial}{\partial \lambda} \sqrt{h} = \theta \sqrt{h} \tag{4.32}$$

From (4.29), $\sqrt{h}$ is the area element on a surface of constant λ within $\mathcal{N}$, so θ measures the rate of increase of this area element with respect to affine parameter along the geodesics.

4.7 Trapped surfaces

Consider a 2d spacelike surface S, i.e., a 2d submanifold for which all tangent vectors are spacelike. For any $p \in S$ there will be precisely two future-directed null vectors U_1^a and U_2^a orthogonal to S (up to the freedom to rescale U_1^a and U_2^a). If we assume that S is orientable then U_1^a and U_2^a can be defined continuously over S. This defines two families of null geodesics which start on S and are orthogonal to S (with the freedom to rescale U^a corresponding to the freedom to rescale the affine parameter). These null geodesics form two null hypersurfaces $\mathcal{N}_1$ and $\mathcal{N}_2$. In simple situations, these correspond to the set of "outgoing" and "ingoing" light rays that start on S. Consider a null congruence that contains the generators of $\mathcal{N}_i$. By the proposition above, we will have $\hat{\omega}_{ab} = 0$ on $\mathcal{N}_1$ and $\mathcal{N}_2$.

Example. Let S be a 2-sphere $U = U_0$, $V = V_0$ in the Kruskal spacetime. By symmetry, the generators of $\mathcal{N}_i$ will be radial null geodesics, as shown in Fig. 22.

Hence $\mathcal{N}_i$ must be surfaces of constant U or constant V with generators tangent to dU and dV respectively. We saw above that dU and dV correspond to affine parameterization. Raising an index, equation (4.7) gives

$$U_1^a = r e^{r/2M} \left(\frac{\partial}{\partial V} \right)^a \qquad U_2^a = r e^{r/2M} \left(\frac{\partial}{\partial U} \right)^a \tag{4.33}$$

where we have discarded an overall constant and fixed the sign so that U_1^a and U_2^a are future-directed. ($\partial/\partial U$ and $\partial/\partial V$ are future-directed because they are globally null and hence define time-orientations. In region I they both give the same time orientation as the one defined by k^a.) We can now calculate the expansion of these congruences:

$$\theta_1 = \nabla_a U_1^a = \frac{1}{\sqrt{-g}} \partial_\mu \left(\sqrt{-g} U_1^\mu \right) = r^{-1} e^{r/2M} \partial_V \left(r e^{-r/2M} r e^{r/2M} \right) = 2 e^{r/2M} \partial_V r \tag{4.34}$$

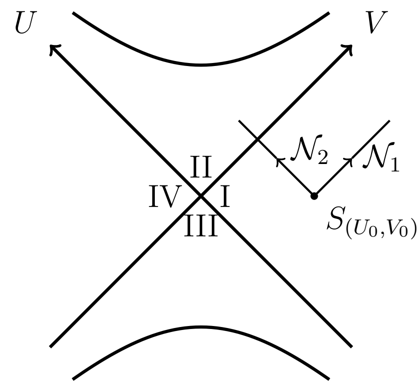


Figure 22. Null hypersurfaces orthogonal to a sphere S ($U = U_0$, $V = V_0$) in the Kruskal spacetime.

The RHS can be calculated from (2.31), giving

$$\theta_1 = -\frac{8M^2}{r}U \tag{4.35}$$

A similar calculation gives

$$\theta_2 = -\frac{8M^2}{r}V \tag{4.36}$$

We can now set $U = U_0$ and $V = V_0$ to study the expansion (on S) of the null geodesics normal to S. For S in region I, we have $\theta_1 > 0$ and $\theta_2 < 0$ i.e., the outgoing null geodesics normal to S are expanding and the ingoing geodesics are converging, as one expects under normal circumstances. In region IV we have $\theta_2 > 0$ and $\theta_1 < 0$ so again we have an expanding family and a converging family. However, in region II we have $\theta_1 < 0$ and $\theta_2 < 0$: both families of geodesics normal to S are converging. And in region III, $\theta_1 > 0$ and $\theta_2 > 0$ so both families are expanding.

Definition. A compact, orientable, spacelike, 2-surface is *trapped* if both families of null geodesics orthogonal to S have negative expansion everywhere on S. It is *marginally trapped* if both families have non-positive expansion everywhere on S.

So in the Kruskal spacetime, all 2-spheres $U = U_0$, $V = V_0$ in region II are trapped and 2-spheres on the event horizon ($U_0 = 0$, $V_0 > 0$) are marginally trapped.

4.8 Raychaudhuri's equation

Let's determine how the expansion evolves along the geodesics of a null geodesic congruence.

Proposition (Raychaudhuri's equation).

$$\frac{d\theta}{d\lambda} = -\frac{1}{2}\theta^2 - \hat{\sigma}^{ab}\hat{\sigma}_{ab} + \hat{\omega}^{ab}\hat{\omega}_{ab} - R_{ab}U^aU^b \tag{4.37}$$

Proof. From the definition of θ we have

$$\frac{d\theta}{d\lambda} = U \cdot \nabla \left(B^a{}_b P^b_a\right) = P^b_a U \cdot \nabla B^a{}_b = P^b_a U^c \nabla_c \nabla_b U^a \tag{4.38}$$

Now commute derivatives using the definition of the Riemann tensor:

$$\begin{aligned}\frac{d\theta}{d\lambda} &= P^b_a U^c \left(\nabla_b \nabla_c U^a + R^a{}_{dcb} U^d\right) \\ &= P^b_a \left[\nabla_b (U^c \nabla_c U^a) - (\nabla_b U^c)(\nabla_c U^a)\right] + P^b_a R^a{}_{dcb} U^c U^d \\ &= -B^c{}_b P^b_a B^a{}_c - R_{cd} U^c U^d \end{aligned} \tag{4.39}$$

where we used the geodesic equation and, in the final term, the antisymmetry of the Riemann tensor allows us to replace P^b_a with δ^b_a. Finally (exercise) we can rewrite the first term so that

$$\frac{d\theta}{d\lambda} = -\hat{B}^c{}_a \hat{B}^a{}_c - R_{ab} U^a U^b \tag{4.40}$$

The result then follows by using (4.23).

Similar calculations give equations governing the evolution of shear and rotation.

4.9 Energy conditions

Raychaudhuri's equation involves the Ricci tensor, which is related to the energy-momentum tensor of matter via the Einstein equation. We will want to consider only "physical" matter, which implies that the energy-momentum tensor should satisfy certain conditions. For example, an observer with 4-velocity u^a would measure an "energy-momentum current" $j^a = -T^a{}_b u^b$. We would expect "physically reasonable" matter not to move faster than light, so this current should be non-spacelike. This motivates:

Dominant energy condition. $-T^a{}_b V^b$ is a future-directed causal vector (or zero) for all future-directed timelike vectors V^a.

For matter satisfying the dominant energy condition, if T_{ab} is zero in some closed region S of a spacelike hypersurface Σ then T_{ab} will be zero within $D^+(S)$. (See Hawking and Ellis for a proof.)

Example. Consider a massless scalar field

$$T_{ab} = \partial_a \Phi \partial_b \Phi - \frac{1}{2} g_{ab} (\partial \Phi)^2 \tag{4.41}$$

Let

$$j^a = -T^a{}_b V^b = -(V \cdot \partial\Phi)\partial^a\Phi + \frac{1}{2}V^a(\partial\Phi)^2 \tag{4.42}$$

then, for timelike V^a,

$$j^2 = \frac{1}{4}V^2\left((\partial\Phi)^2\right)^2 \leq 0 \tag{4.43}$$

so j^a is indeed causal or zero. Now consider

$$V \cdot j = -(V \cdot \partial\Phi)^2 + \frac{1}{2}V^2(\partial\Phi)^2 = -\frac{1}{2}(V \cdot \partial\Phi)^2 + \frac{1}{2}V^2\left(\partial\Phi - \frac{V \cdot \partial\Phi}{V^2}V\right)^2 \tag{4.44}$$

the final expression in brackets is orthogonal to V^a and hence must be spacelike or zero, so its norm is non-negative. We then have $V \cdot j \leq 0$ using $V^2 < 0$. Hence j^a is future-directed (or zero).

A less restrictive condition requires only that the energy density measured by all observers is positive:

Weak energy condition. $T_{ab}V^aV^b \geq 0$ for any causal vector V^a.

A special case of this is

Null energy condition. $T_{ab}V^aV^b \geq 0$ for any null vector V^a.

The dominant energy condition implies the weak energy condition, which implies the null energy condition. Another energy condition is

Strong energy condition. $(T_{ab} - (1/2)g_{ab}T^c_c)\, V^aV^b \geq 0$ for all causal vectors V^a.

Using the Einstein equation, this is equivalent to $R_{ab}V^aV^b \geq 0$, or "gravity is attractive". Despite its name, the strong energy condition does not imply the weak energy condition. The strong energy condition is needed to prove some of the singularity theorems, but the dominant energy condition is the most important physically. For example, our universe appears to contain a positive cosmological constant. This violates the strong energy condition but respects the dominant energy condition.

4.10 Conjugate points

Lemma. In a spacetime satisfying Einstein's equation with matter obeying the null energy condition, the generators of a null hypersurface satisfy

$$\frac{d\theta}{d\lambda} \leq -\frac{1}{2}\theta^2 \tag{4.45}$$

Proof. Consider the RHS of Raychaudhuri's equation. The generators of a null hypersurface have $\hat{\omega} = 0$. Vectors in $T_\perp$ are all spacelike, so the metric restricted to $T_\perp$ is positive definite. Hence $\hat{\sigma}^{ab}\hat{\sigma}_{ab} \geq 0$. Einstein's equation gives $R_{ab}U^aU^b = 8\pi T_{ab}U^aU^b$ because U^a is null. Hence the null energy condition implies $R_{ab}U^aU^b \geq 0$. The result follows from Raychaudhuri's equation.

Corollary. If $\theta = \theta_0 < 0$ at a point p on a generator γ of a null hypersurface then $\theta \to -\infty$ along γ within an affine parameter distance $2/|\theta_0|$ provided γ extends this far.

Proof. Let $\lambda = 0$ at p. Equation (4.45) implies that θ is decreasing, so negative, and

$$\frac{d}{d\lambda}\theta^{-1} \geq \frac{1}{2} \tag{4.46}$$

Integrating gives $\theta^{-1} - \theta_0^{-1} \geq \lambda/2$, which can be rearranged to give

$$\theta \leq \frac{\theta_0}{1 + \lambda\theta_0/2} \tag{4.47}$$

if $\theta_0 < 0$ then the RHS $\to -\infty$ as $\lambda \to 2/|\theta_0|$.

Definition. Points p, q on a geodesic γ are *conjugate* if there exists a Jacobi field (i.e. a solution of the geodesic deviation equation) along γ that vanishes at p and q but is not identically zero.

Roughly speaking, if p and q are conjugate then there exist multiple infinitesimally nearby geodesics which pass through p and q. The following results are proved in Hawking and Ellis:

Theorem 1. Consider a null geodesic congruence which includes all of the null geodesics through p (this congruence is singular at p). If $\theta \to -\infty$ at a point q on a null geodesic γ through p then q is conjugate to p along γ.

Theorem 2. Let γ be a causal curve with $p, q \in \gamma$. Then there does *not* exist a smooth 1-parameter family of causal curves γ_s connecting p, q with $\gamma_0 = \gamma$ and γ_s timelike for $s > 0$ (i.e. γ cannot be smoothly deformed to a timelike curve) if, and only if, γ is a null geodesic with no point conjugate to p along γ between p and q.

For example, consider the 3d spacetime $\mathbb{R} \times S^2$ with metric

$$ds^2 = -dt^2 + d\Omega^2 \tag{4.48}$$

Null geodesics emitted from the south pole at time $t = 0$ (the spacetime point p) all reconverge at the north pole at time $t = \pi$ (spacetime point r). Such geodesics correspond to great circles of S^2. r is conjugate to p along any of these geodesics. If

q lies beyond r along one of these geodesics then by deforming the great circle into a shorter path one can travel from p to q with velocity less than that of light hence there exists a timelike curve from p to q.

Now consider the case in which we have a 2d spacelike surface S. As discussed above, we can introduce two future-directed null vector fields U_1^a, U_2^a on S that are normal to S and consider the null geodesics which have one of these vectors as their tangent on S. These generate a null hypersurface $\mathcal{N}$. Let p be a point on a geodesic γ in this family. We say that p is conjugate to S if there exists a Jacobi field along γ that vanishes at p and, on S, is tangent to S. If p is conjugate to S then, roughly speaking, infinitesimally nearby geodesics normal to S intersect at p.

The analogue of theorem 1 in this case is: p is conjugate to S if $\theta \to -\infty$ at p along one of the geodesics just discussed, in a congruence containing the generators of $\mathcal{N}$. (We saw earlier that θ depends only in the geodesics in $\mathcal{N}$ and not on how the other geodesics in the congruence are chosen.)

4.11 Causal structure

Definition. Let (M, g) be a time-orientable spacetime and $U \subset M$. The *chronological future* of U, denoted $I^+(U)$, is the set of points of M which can be reached by a future-directed timelike curve starting on U. The *causal future* of U, denoted $J^+(U)$, is the union of U with the set of points of M which can be reached by a future-directed causal curve starting on U. The chronological past $I^-(U)$ and causal past $J^-(U)$ are defined similarly.

For example, let q be a point in Minkowski spacetime. Then $I^+(p)$ is the set of points strictly inside the future light cone of p and $J^+(p)$ is the set of points on or inside the future light cone of p, including p itself (Fig. 23).

Next we need to review some basic topological ideas. A subset S of M is *open* if, for any point $p \in S$, there exists a neighbourhood V of p (i.e. a set of points whose coordinates in some chart are a neighbourhood of the coordinates of p) such that $V \subset S$. Small deformations of timelike curves remain timelike. Hence $I^\pm(U)$ are open subsets of M.

We use an overbar to denote the *closure* of a set, i.e., the union of a set and its limit points. In Minkowski spacetime, we have $\overline{I^\pm(p)} = J^\pm(p)$ so $J^\pm(p)$ are closed sets, i.e., they contain their limit points. This is not true in general e.g. let (M, g) be the spacetime obtained by deleting a point from 2d Minkowski spacetime, see Fig. 24. In this example we see that $J^+(p) \neq \overline{I^+(p)}$ and $J^+(p)$ is not closed.

A point $p \in S$ is an *interior point* if there exists a neighbourhood of p contained in S. The *interior* of S, denoted $\mathrm{int}(S)$ is the set of interior points of S. If S is open

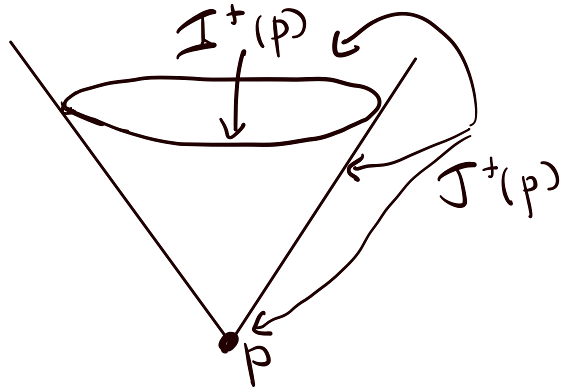


Figure 23. Chronological future and causal future of a point in Minkowski spacetime.

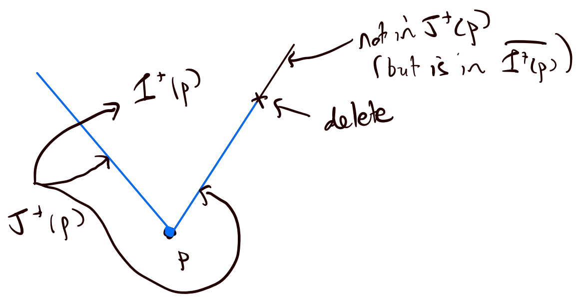


Figure 24. Minkowski spacetime with a point deleted. Here $J^+(p) \neq \overline{I^+(p)}$.

then $\text{int}(S) = S$. The *boundary* of S is $\dot{S} = \bar{S} \backslash \text{int}(S)$. This is a topological boundary rather than a boundary in the sense of manifold-with-boundary (to be defined later).

The boundary of $I^+(p)$ is $\dot{I}^+(p) = \overline{I^+(p)} \backslash I^+(p)$. In Minkowski spacetime, $I^+(p)$ is the set of points along future-directed timelike geodesics starting at p and $\dot{I}^+(p)$ is the set of points along future-directed null geodesics starting at p. These statements are not true in general, they are true only *locally* in the following sense:

Theorem 1. Given $p \in M$ there exists a *convex normal neighbourhood* of p. This is an open set U with $p \in U$ such that for any $q, r \in U$ there exists a unique geodesic connecting q, r that stays in U. The chronological future of p in the spacetime (U, g) consists of all points in U along future-directed timelike geodesics in U that start at p. The boundary of this region is the set of all points in U along future-directed null geodesics in U that start at p.

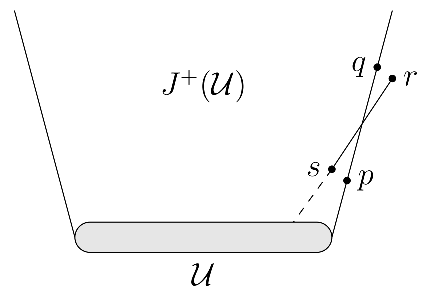


Figure 25. Proof of achronality of $\dot{J}^+(U)$.

Proof. See Hawking and Ellis or Wald.

Corollary. If $q \in J^+(p) \backslash I^+(p)$ then there exists a null geodesic from p to q.

Proof (sketch). Let γ be a future-directed causal curve with $\gamma(0) = p$ and $\gamma(1) = q$. Since $[0, 1]$ is compact, the set of points on γ is compact, hence we can cover a neighbourhood of this set by finitely many convex normal neighbourhoods. Use the above Theorem in each neighbourhood.

Lemma. Let $S \subset M$. Then $J^+(S) \subset \overline{I^+(S)}$ and $I^+(S) = \text{int}(J^+(S))$.

Proof. See Hawking and Ellis (this is an exercise in Wald).

Since $I^+(S) \subset J^+(S)$, $\overline{I^+(S)} \subset \overline{J^+(S)}$ so the first result implies that $\overline{J^+(S)} = \overline{I^+(S)}$. The second result then implies $\dot{J}^+(S) = \dot{I}^+(S)$.

Definition. $S \subset M$ is *achronal* if no two points of S are connected by a timelike curve.

Theorem 2. Let $U \subset M$. Then $\dot{J}^+(U)$ is an achronal 3d submanifold of M.

Proof. (See Fig. 25.) Assume $p, q \in \dot{J}^+(U)$ with $q \in I^+(p)$. Since $I^+(p)$ is open, there exists r (near q) with $r \in I^+(p)$ but $r \notin J^+(U)$. Similarly, since $I^-(r)$ is open, there exists s (near p) with $s \in I^-(r)$ and $s \in J^+(U)$. Hence there exists a causal curve from U to s to r so $r \in J^+(U)$, which is a contradiction. Hence we can't have $q \in I^+(p)$, which establishes achronality. For proof of the "submanifold" part see Wald.

For example, let $M = \mathbb{R} \times S^1$ with the flat metric

$$ds^2 = -dt^2 + d\phi^2 \tag{4.49}$$

where $\phi \sim \phi + 2\pi$ parameterizes S^1 (this is a 2d version of the "Einstein static universe"). The diagram shows $J^+(p)$ (shaded). Its boundary $\dot{J}^+(p)$ is a pair of null geodesic segments which start at p and end at q.

Note that q is a *future endpoint* of these geodesics. They could be extended to the future beyond q but then they would leave $\dot{J}^+(p)$. They also have a past endpoint at p.

The next theorem characterises the behaviour of $\dot{J}^+(U)$.

Theorem 3. Let $U \subset M$ be closed. Then every $p \in \dot{J}^+(U)$ with $p \notin U$ lies on a null geodesic λ lying entirely in $\dot{J}^+(U)$ and such that λ is either past-inextendible or has a past endpoint on U.

Proof (sketch). Since U is closed, $M\backslash U$ is a manifold. We will work in this manifold. Consider a compact neighbourhood V of p and a sequence of points $p_n \in I^+(U) = \text{int}(J^+(U))$ with limit point p. Let λ_n be a timelike curve from U to p_n and let q_n be the past endpoint of λ_n in V:

Then one can show that q_n has a limit point $q \in \overline{J^+(U)}$ and there is a causal "limit curve" λ from q to p lying in $\overline{J^+(U)}$ (see Wald). We need to show $\lambda \subset \dot{J}^+(U)$. Suppose there is a point $r \in \lambda$ such that $r \in I^+(U) = \text{int}(J^+(U))$. Then there is a timelike curve γ from $r' \in U$ to r. But then we can get from r' to r to p by following γ then λ. Hence $p \in J^+(r')$ but $p \notin I^+(r')$ (as $p \notin I^+(U)$) so theorem 1 implies that this curve must be a null geodesic, which is a contradiction because it's not null everywhere. Hence we must have $\lambda \subset \overline{J^+(U)} - I^+(U) = \dot{J}^+(U)$.

Theorem 2 tells us that $\dot{J}^+(U)$ is achronal so $p \notin I^+(q)$. Theorem 1 then tells us that λ must be a null geodesic. Now we repeat the argument starting at q, to get a point $r \in \dot{J}^+(U)$ with a null geodesic λ' from r to q lying in $\dot{J}^+(U)$, If λ' were not the past extension of λ we could "round off the corner" to construct a timelike curve from r to p, violating achronality. This argument can be repeated indefinitely, hence λ cannot have a past endpoint in $M\backslash U$.

In the case of a globally hyperbolic spacetime, this theorem can be strengthened as follows:

Theorem 4. Let S be a 2-dimensional orientable compact spacelike submanifold of a globally hyperbolic spacetime. Then every $p \in \dot{J}^+(S)$ lies on a future-directed null geodesic starting from S which is orthogonal to S and has no point conjugate to S between S and p.

Finally, we can use the notation of this section to define what we mean by a Cauchy horizon:

Definition. The *future Cauchy horizon* of a partial Cauchy surface Σ is $H^+(\Sigma) = \overline{D^+(\Sigma)} \backslash I^-(D^+(\Sigma))$. Similarly for the past Cauchy horizon $H^-(\Sigma)$.

We don't define $H^+(\Sigma)$ simply as $\dot{D}^+(\Sigma)$ since this includes Σ itself. However, one can show that $\dot{D}(\Sigma) = H^+(\Sigma) \cup H^-(\Sigma)$. One can also show that $H^\pm$ are null hypersurfaces in the same sense as $\dot{J}^+(U)$ in Theorems 2 and 3 above. (See Wald for details.)

4.12 Penrose singularity theorem

Theorem (Penrose 1965). Let (M, g) be globally hyperbolic with a non-compact Cauchy surface Σ. Assume that the Einstein equation and the null energy condition are satisfied and that M contains a trapped surface T. Let $\theta_0 < 0$ be the maximum value of θ on T for both sets of null geodesics orthogonal to T. Then at least one of these geodesics is future-inextendible and has affine length no greater than $2/|\theta_0|$.

Proof. Assume that all future inextendible null geodesics orthogonal to T have affine length greater than $2/|\theta_0|$. Along any of these geodesics, we will have $\theta \to -\infty$ (from the Corollary in section 4.10), and hence a point conjugate to T, within affine parameter no greater than $2/|\theta_0|$.

Let $p \in \dot{J}^+(T)$, $p \notin T$. From theorem 4 above, we know that p lies on a future-directed null geodesic γ starting from T which is orthogonal to T and has no point conjugate to T between T and p. It follows that p cannot lie beyond the point on γ conjugate to T on γ.

Therefore $\dot{J}^+(T)$ is a subset of the compact set consisting of the set of points along the null geodesics orthogonal to T, with affine parameter less than or equal to $2/|\theta_0|$. Since $\dot{J}^+(T)$ is closed this implies that $\dot{J}^+(T)$ is compact. Now recall (theorem 2 of section 4.11) that $\dot{J}^+(T)$ is a manifold, which implies that it can't have a boundary. If Σ were compact this might be possible because the "ingoing" and "outgoing" congruences orthogonal to T might join up:

But since Σ is non-compact, this can't happen and we'll now reach a contradiction as follows. Pick a timelike vector field T^a (possible because our manifold is time-orientable). By global hyperbolicity, integral curves of this vector field will intersect Σ exactly once. They will intersect $\dot{J}^+(T)$ at most once (because this set is achronal by theorem 2 of section 4.11). This defines a continuous one-to-one map $\alpha : \dot{J}^+(T) \to \Sigma$. This is a homeomorphism between $\dot{J}^+(T)$ and $\alpha(\dot{J}^+(T)) \subset \Sigma$. Since the former is a closed set, so must be the latter. Now $\dot{J}^+(T)$ is a 3d submanifold hence for any $p \in \dot{J}^+(T)$ we can find a neighbourhood V of p in $\dot{J}^+(T)$. Then $\alpha(V)$ gives a neighbourhood of $\alpha(p)$ in $\alpha(\dot{J}^+(T))$ hence the latter set is open (in Σ). Since it is both open and closed, and since Σ is connected (this follows from connectedness of M) we have $\alpha(\dot{J}^+(T)) = \Sigma$. But this is a contradiction because the set on the LHS is compact (because $\dot{J}^+(T)$ is). □

The formation of trapped surfaces is routinely observed in numerical simulations of gravitational collapse. There are also various mathematical results concerning the formation of trapped surfaces. The Einstein equation possesses the property of *Cauchy stability*, which implies that the solution in a *compact* region of spacetime depend continuously on the initial data. In a spacetime describing spherically symmetric gravitational collapse, choose a compact region that includes a trapped surface (e.g. a 2-sphere in region II of the Kruskal diagram). Cauchy stability implies that if one perturbs the initial data (breaking spherical symmetry) then the resulting spacetime will also have a trapped surface, for a small enough initial perturbation. This shows that trapped surfaces occur generically in gravitational collapse.

A theorem due to Schoen and Yau (1983) establishes that asymptotically flat initial data will contain a trapped surface if the energy density of matter is sufficiently large in a small enough region. Christodoulou (2009) has proved that trapped surfaces can be formed dynamically, even in the absence of matter and without any symmetry assumptions, by sending sufficiently strong gravitational waves into a small enough region.

The above theorem implies that if the maximal development of asymptotically flat initial data contains a trapped surface then this maximal development is not geodesically complete. Such incompletness might arise because the maximal development is extendible. However, this is (generically) excluded if the strong cosmic censorship conjecture is correct. Hence it is expected that, generically, the incompleteness arises because the maximal development is singular. In fact, a different singularity theorem (due to Hawking and Penrose) eliminates the assumption that spacetime is globally hyperbolic (at the cost of requiring the strong energy condition and a mild "genericity" assumption on the spacetime curvature) and still proves existence of incomplete geodesics. So even if the maximal development is extendible then the Hawking-Penrose theorem implies that this extended spacetime must be geodesically incomplete, i.e., singular.

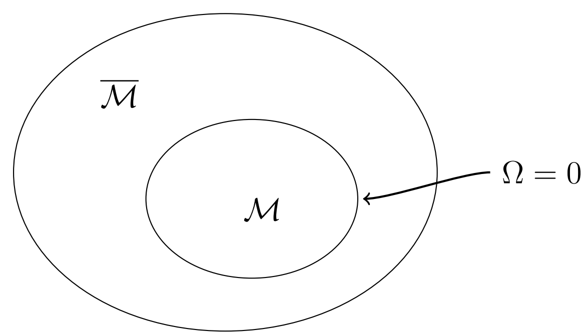
Hence there are very good reasons to believe that gravitational collapse leads to formation of a singularity. Notice that these theorems tell us nothing about the nature of this singularity e.g. we do not know that it must be a curvature singularity as occurs in spherically symmetric collapse.

5 Asymptotic flatness

We've already defined the notion of asymptotic flatness of an initial data set. In this chapter, we will define what it means for a spacetime to be asymptotically flat. We'll then be able to define the term "black hole".

5.1 Conformal compactification

Given a spacetime (M, g) we can define a new metric $\bar{g} = \Omega^2 g$ where Ω is a smooth positive function on M. We say that $\bar{g}$ is obtained from g by a *conformal transformation*. The metrics $g, \bar{g}$ agree on the definitions of "timelike", "spacelike" and "null" so they have the same light cones, i.e., the same causal structure. The idea of conformal compactification is to choose Ω so that "points at infinity" with respect to g are at "finite distance" w.r.t. the "unphysical" metric $\bar{g}$. To do this we need $\Omega \to 0$ "at infinity". More precisely, we try to choose Ω so that the spacetime $(M, \bar{g})$ is extendible in the sense we discussed previously, i.e., $(M, \bar{g})$ is part of a larger "unphysical" spacetime $(\bar{M}, \bar{g})$. M is then a proper subset of $\bar{M}$ with $\Omega = 0$ on the boundary ∂M of M in $\bar{M}$. This boundary ∂M corresponds to "infinity" in (M, g). It is easiest to see how this works by looking at some examples.



Minkowski spacetime

Let (M, g) be Minkowski spacetime. In spherical polars the metric is

$$g = -dt^2 + dr^2 + r^2 d\omega^2 \tag{5.1}$$

(We denote the metric on S^2 by $d\omega^2$ to avoid confusion with the conformal factor Ω.) Define retarded and advanced time coordinates

$$u = t - r \qquad v = t + r \tag{5.2}$$

In what follows it will be important to keep track of the ranges of the different coordinates: since $r \geq 0$ we have $-\infty < u \leq v < \infty$. The metric is

$$g = -dudv + \frac{1}{4}(u-v)^2 d\omega^2 \tag{5.3}$$

Now define new coordinates (p, q) by

$$u = \tan p \qquad v = \tan q \tag{5.4}$$

so the range of (p, q) is finite: $-\pi/2 < p \leq q < \pi/2$. This gives

$$g = (2\cos p \cos q)^{-2} \left[-4dpdq + \sin^2(q-p)d\omega^2\right] \tag{5.5}$$

"Infinity" in the original coordinates corresponds to $|t| \to \infty$ or $r \to \infty$. In the new coordinates this corresponds to $|p| \to \pi/2$ or $|q| \to \pi/2$.

To conformally compactify this spacetime, define the positive function

$$\Omega = 2\cos p \cos q \tag{5.6}$$

and let

$$\bar{g} = \Omega^2 g = -4dpdq + \sin^2(q-p)d\omega^2 \tag{5.7}$$

Finally define

$$T = q + p \in (-\pi, \pi) \qquad \chi = q - p \in [0, \pi) \tag{5.8}$$

so

$$\bar{g} = -dT^2 + d\chi^2 + \sin^2 \chi d\omega^2 \tag{5.9}$$

Now $d\chi^2 + \sin^2 \chi d\omega^2$ is the unit radius round metric on S^3. If we had $T \in (-\infty, \infty)$ and $\chi \in [0, \pi]$ then $\bar{g}$ would be the metric of the *Einstein static universe* $\mathbb{R} \times S^3$, given by the product of a flat time direction with the unit round metric on S^3. The ESU can be visualised as an infinite cylinder, whose axis corresponds to the time direction. In our case the restrictions on the ranges of p, q imply that M is just a finite portion of the ESU, as shown in Fig. 26.

Let $(\bar{M}, \bar{g})$ denote the ESU. This is an extension of $(M, \bar{g})$. The boundary ∂M of M in $\bar{M}$ corresponds to "infinity" in Minkowski spacetime. This consists of (i) the points labelled $i^\pm$ i.e. $T = \pm\pi, \chi = 0$ (ii) the point labelled i^0, i.e., $T = 0, \chi = \pi$ (iii) a pair of null hypersurfaces $\mathcal{I}^\pm$ ($\mathcal{I}$ is pronounced "scri") with equations $T = \pm(\pi - \chi)$, which are parameterized by $\chi \in (0, \pi)$ and (θ, ϕ) and hence have the topology of cylinders $\mathbb{R} \times S^2$ (since $(0, \pi)$ is diffeomorphic to $\mathbb{R}$).

It is convenient to project the above diagram onto the (T, χ)-plane to obtain the *Penrose diagram* of Minkowski spacetime shown in Fig. 27. A Penrose diagram is a

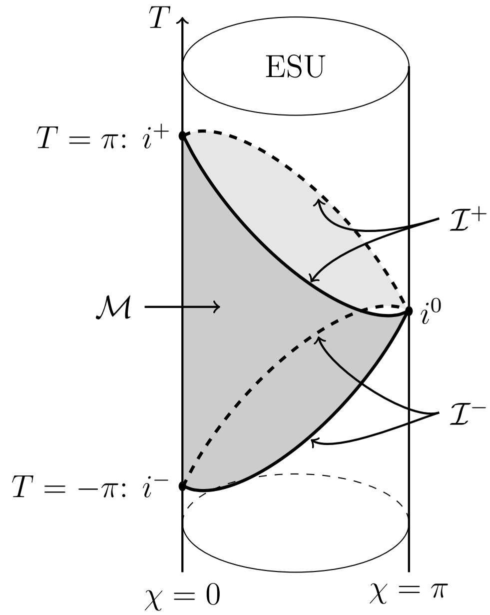


Figure 26. Minkowski spacetime is mapped to a subset of the Einstein static universe.

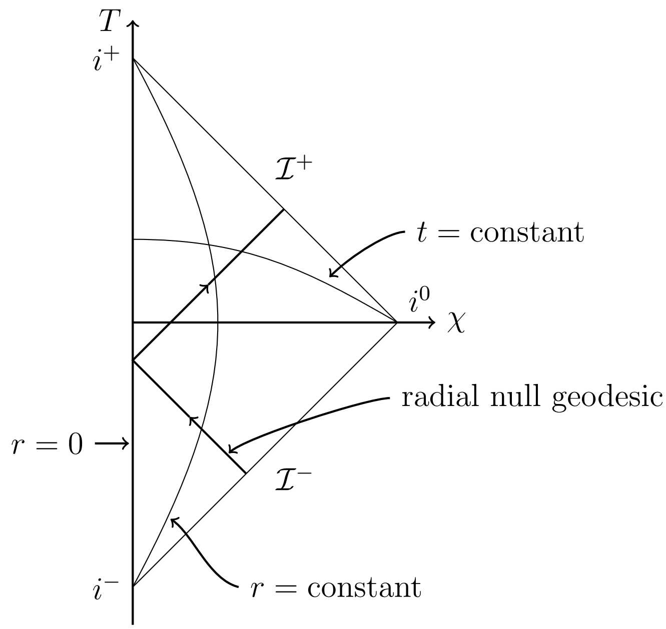


Figure 27. Penrose diagram of Minkowski spacetime.

subset of $\mathbb{R}^2$ endowed with a flat Lorentzian metric (in this case $-dT^2 + d\chi^2$). Each point of the interior of a Penrose diagram represents an S^2. Points of the boundary can represent an axis of symmetry (where $r = 0$), points at "infinity" of our original spacetime with metric g or, as we shall see, singularities.

Let's understand how the geodesics of g look on a Penrose diagram. This is easiest for *radial* geodesics, i.e., constant θ, ϕ. Remember that the causal structure of g and $\bar{g}$ is the same. Hence radial null curves of g are null curves of the flat metric $-dT^2 + d\chi^2$, i.e., straight lines at 45°. These all start at $\mathcal{I}^-$, pass through the origin, and end at $\mathcal{I}^+$. For this reason, $\mathcal{I}^-$ is called *past null infinity* and $\mathcal{I}^+$ is called *future null infinity*. Similarly, radial timelike geodesics start i^- and end at i^+ so i^- is called *past timelike infinity* and i^+ is called *future timelike infinity*. Finally, radial spacelike geodesics start and end at i^0 so i^0 is called *spatial infinity*.

One can also plot the projection of non-radial curves onto the Penrose diagram. This projection makes things look "more timelike" w.r.t. the 2d flat metric (because moving the final term in (5.9) to the LHS gives a negative contribution). Hence a non-radial timelike curve remains timelike when projected and a non-radial null curve looks timelike when projected.

The behaviour of geodesics has an analogue for fields. Roughly speaking, *massless* radiation "comes in from" $\mathcal{I}^-$ and "goes out to" $\mathcal{I}+$. For example, consider a massless scalar field ψ in Minkowski spacetime, i.e., a solution of the wave equation $\nabla^a \nabla_a \psi = 0$. For simplicity, assume it is spherically symmetric $\psi = \psi(t, r)$.

Exercise. By deriving an equation for $r\psi$, show that the general spherically symmetric solution of the wave equation in Minkowski spacetime is

$$\psi(t, r) = \frac{1}{r} \left(f(u) + g(v) \right) = \frac{1}{r} \left(f(t - r) + g(t + r) \right) \tag{5.10}$$

where f and g are arbitrary functions. This is singular at $r = 0$ (and hence not a solution there) unless $g(x) = -f(x)$ which gives

$$\psi(t, r) = \frac{1}{r} \left(f(u) - f(v) \right) = \frac{1}{r} \left(F(p) - F(q) \right) \tag{5.11}$$

where $F(x) = f(\tan x)$. If we let $F_0(q)$ denote the limiting value of $r\psi$ on $\mathcal{I}^-$ (where $p = -\pi/2$) then we have $F(-\pi/2) - F(q) = F_0(q)$ so $F(q) = F(-\pi/2) - F_0(q)$. Hence we can write the solution as

$$\psi = \frac{1}{r} \left(F_0(q) - F_0(p) \right) \tag{5.12}$$

which is uniquely determined by the function F_0 governing the behaviour of the solution at $\mathcal{I}^-$. Similarly it is uniquely determined by the behaviour at $\mathcal{I}^+$.

2d Minkowski

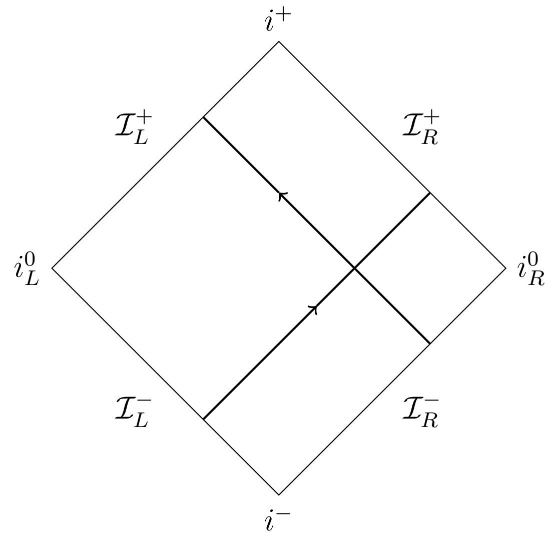


Figure 28. Penrose diagram of 2d Minkowski spacetime, showing left and right moving null geodesics.

As another example of these ideas, consider the Penrose diagram of 2d Minkowski spacetime with metric

$$g = -dt^2 + dr^2 \tag{5.13}$$

Following the same coordinate transformations as before, the only difference is that now we have $-\infty < r < \infty$ hence $-\infty < u, v < \infty$, $-\pi/2 < p, q < \pi/2$ and $T, \chi \in (-\pi, \pi)$. The Penrose diagram is shown in Fig. 28. In this case, we have "left" and "right" portions of spatial infinity and future/past null infinity.

Kruskal spacetime

In this case, we know that the spacetime (M, g) has two asymptotically flat regions. it is natural to expect that "infinity" in each of these regions has the same structure as in (4d) Minkowski spacetime. Hence we expect "infinity" in Kruskal spacetime to correspond to two copies of infinity in Minkowski spacetime. To construct the Penrose diagram for Kruskal we would define new coordinates $P = P(U)$ and $Q = Q(V)$ (so that lines of constant P or Q are radial null geodesics) such that that the range of P, Q is finite, say $(-\pi/2, \pi/2)$, then we would need to find a conformal factor Ω so that the resulting unphysical metric $\bar{g}$ can be extended smoothly onto a bigger manifold $\bar{M}$ (analogous to the Einstein static universe we used for Minkowski spacetime). M is then a subset of $\bar{M}$ with a boundary that has 4 components, corresponding to places where either P or Q is $\pm\pi/2$. We identify these 4 components as future/past null infinity in

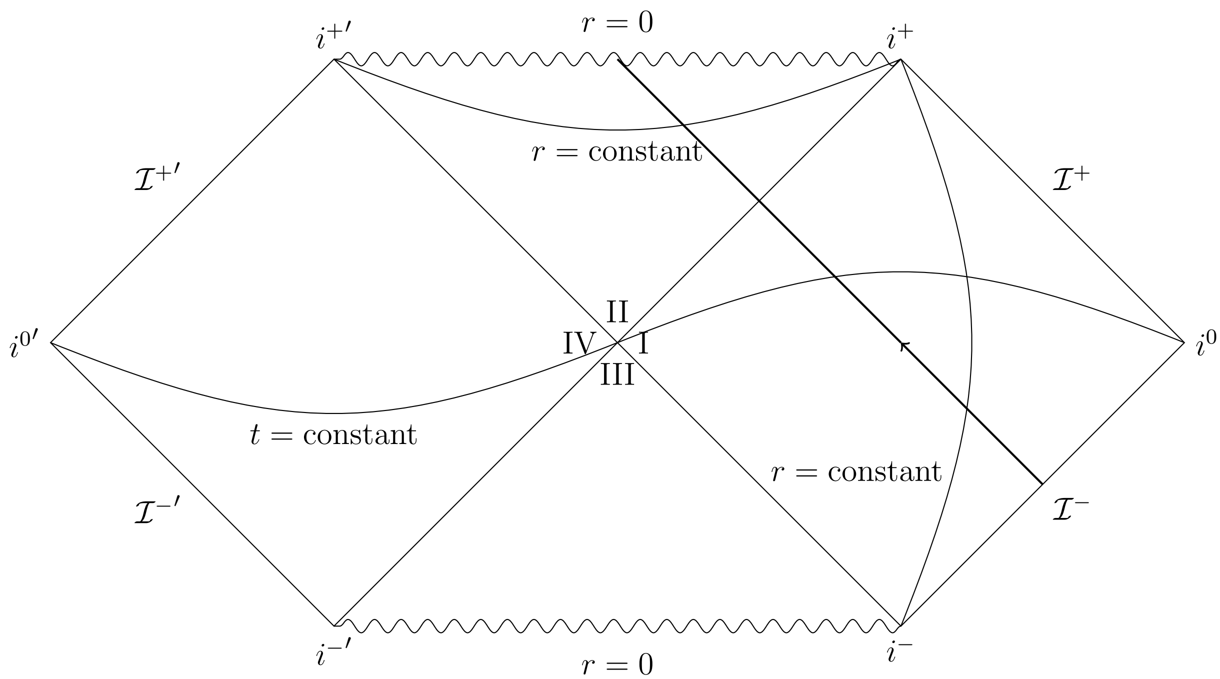


Figure 29. Penrose diagram of the Kruskal spacetime.

region I, which we denote as $\mathcal{I}^{\pm}$ and future/past null infinity in region IV, which we denote as $\mathcal{I}^{\pm\prime}$.

Doing this explicitly is fiddly. Fortunately we don't need to do it: now we've understood the structure of infinity we can deduce the form of the Penrose diagram from the Kruskal diagram. This is because both diagrams show radial null curves as straight lines at 45°. The only important difference is that "infinity" corresponds to a boundary of the Penrose diagram. It is conventional to use the freedom in choosing Ω to arrange that the curvature singularity at $r = 0$ is a horizontal straight line in the Penrose diagram. The result is shown in Fig. 29.

In contrast with the conformal compactification of Minkowski spacetime, it turns out that the unphysical metric is singular at $i^{\pm}$ (and $i^{\pm\prime}$). This can be understood because lines of constant r meet at $i^{\pm}$, and this includes the curvature singularity $r = 0$. Less obviously, it turns out that one can't choose Ω to make the unphysical metric smooth at i^0.

Spherically symmetric collapse

The Penrose diagram for spherically symmetric gravitational collapse is easy to deduce from the form of the Kruskal diagram. It is shown in Fig. 30.

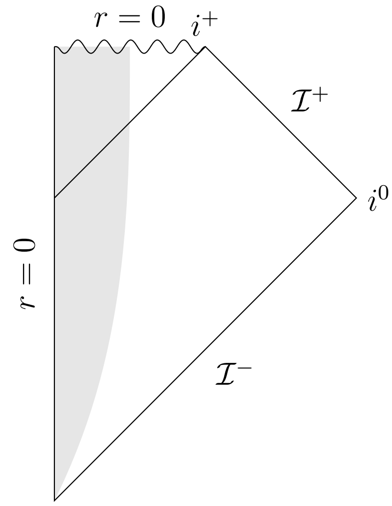


Figure 30. Penrose diagram for spherically symmetric gravitational collapse.

5.2 Asymptotic flatness

An asymptotically flat spacetime is one that "looks like Minkowski spacetime at infinity". In this section we will define this precisely. Infinity in Minkowski spacetime consists of $\mathcal{I}^\pm$, $i^\pm$ and i^0. However, we saw that $i^\pm$ are singular points in the conformal compactification of the Kruskal spacetime. Since we want to regard the latter as asymptotically flat, we cannot include $i^\pm$ in our definition of asymptotic flatness. We also mentioned that i^0 is not smooth in the Kruskal spacetime so we will also not include i^0. (However, it is possible to extend the definition to include i^0, see Wald for details.) So we will define a spacetime to be asymptotically flat if it has the same structure for null infinity $\mathcal{I}^\pm$ as Minkowski spacetime.

First, recall that a manifold-with-boundary is defined in the same way as a manifold except that the charts are now maps $\phi : M \to \mathbb{R}^n/2 \equiv \{(x^1, \ldots, x^n) : x^1 \leq 0\}$. The boundary ∂M of M is defined to be the set of points which have $x^1 = 0$ in some chart.

Definition. A time-orientable spacetime (M, g) is *asymptotically flat at null infinity* if there exists a spacetime $(\bar{M}, \bar{g})$ such that

1. There exists a positive function Ω on M such that $(\bar{M}, \bar{g})$ is an extension of $(M, \Omega^2 g)$ (hence if we regard M as a subset of $\bar{M}$ then $\bar{g} = \Omega^2 g$ on M).

2. Within $\bar{M}$, M can be extended to obtain a manifold-with-boundary $M \cup \partial M$

3. Ω can be extended to a function on $\bar{M}$ such that $\Omega = 0$ and $d\Omega \neq 0$ on ∂M

4. ∂M is the disjoint union of two components $\mathcal{I}^+$ and $\mathcal{I}^-$, each diffeomorphic to $\mathbb{R} \times S^2$

5. No past (future) directed causal curve starting in M intersects $\mathcal{I}^+$ ($\mathcal{I}^-$)

6. $\mathcal{I}^\pm$ are "complete". We'll define this below.

Conditions 1,2,3 are just the requirement that there exist an appropriate conformal compactification. The condition $d\Omega \neq 0$ ensures that Ω has a first order zero on ∂M, as in the examples discussed above. This is needed to ensure that the spacetime metric approaches the Minkowski metric at an appropriate rate near $\mathcal{I}^\pm$. Conditions 4,5,6 ensure that infinity has the same structure as null infinity of Minkowski spacetime. In particular, condition 5 ensures that $\mathcal{I}^+$ lies "to the future of M" and $\mathcal{I}^-$ lies "to the past of M".

Example. Consider the Schwarzschild solution in outgoing EF coordinates (u, r, θ, ϕ), for which $\mathcal{I}^+$ corresponds to $r \to \infty$ with finite u. Let $r = 1/x$ with $x > 0$. This gives

$$g = -\left(1 - 2Mx\right) du^2 + 2\frac{dudx}{x^2} + \frac{1}{x^2}\left(d\theta^2 + \sin^2\theta d\phi^2\right) \tag{5.14}$$

Hence by choosing a conformal factor $\Omega = x$ we obtain the unphysical metric

$$\bar{g} = -x^2\left(1 - 2Mx\right) du^2 + 2dudx + d\theta^2 + \sin^2\theta d\phi^2 \tag{5.15}$$

which can be smoothly extended across $x = 0$. The surface $x = 0$ is $\mathcal{I}^+$. It is parameterized by (u, θ, ϕ) and is hence diffeomorphic to $\mathbb{R} \times S^2$. Of course we've only checked the above definition at $\mathcal{I}^+$ here. But one can do the same at $\mathcal{I}^-$ using ingoing EF coordinates and the same conformal factor $\Omega = 1/r$ (recall that r is the same for both coordinate charts). Hence the Schwarzschild spacetime is asymptotically flat at null infinity. Similarly, the Kruskal spacetime is asymptotically flat (in fact both regions I and IV are asymptotically flat).

Let's now see how the above definition implies that the metric must approach the Minkowski metric near $\mathcal{I}^+$ (of course $\mathcal{I}^-$ is similar).

Exercise (examples sheet 2). Let $\bar{\nabla}$ denote the Levi-Civita connection of $\bar{g}$ and $\bar{R}_{ab}$ the Ricci tensor of $\bar{g}$. Show that on M

$$R_{ab} = \bar{R}_{ab} + 2\Omega^{-1}\bar{\nabla}_a\bar{\nabla}_b\Omega + \bar{g}_{ab}\bar{g}^{cd}\left(\Omega^{-1}\bar{\nabla}_c\bar{\nabla}_d\Omega - 3\Omega^{-2}\partial_c\Omega\partial_d\Omega\right) \tag{5.16}$$

We will consider the case in which (M, g) satisfies the vacuum Einstein equation. This assumption can be weakened: our results will apply also to spacetimes for which the energy-momentum tensor decays sufficiently rapidly near $\mathcal{I}^+$. The vacuum Einstein equation is $R_{ab} = 0$. Multiply by Ω to obtain

$$\Omega \bar{R}_{ab} + 2\bar{\nabla}_a \bar{\nabla}_b \Omega + \bar{g}_{ab} \bar{g}^{cd} \left(\bar{\nabla}_c \bar{\nabla}_d \Omega - 3\Omega^{-1} \partial_c \Omega \partial_d \Omega \right) = 0 \tag{5.17}$$

Since $\bar{g}$ and Ω are smooth at $\mathcal{I}^+$, the first three terms in this equation admit a smooth limit to $\mathcal{I}^+$. It follows that so must the final term which implies that $\Omega^{-1}\bar{g}^{cd}\partial_c\Omega\partial_d\Omega$ can be smoothly extended to $\mathcal{I}^+$. This implies that $\bar{g}^{cd}\partial_c\Omega\partial_d\Omega$ must vanish on $\mathcal{I}^+$ i.e. $d\Omega$ is null (w.r.t $\bar{g}$) on $\mathcal{I}^+$. But $d\Omega$ is normal to $\mathcal{I}^+$ (as $\Omega = 0$ on $\mathcal{I}^+$) hence $\mathcal{I}^+$ must be a null hypersurface in $(\bar{M}, \bar{g})$.

Now the choice of Ω in our definition is far from unique. If Ω satisfies the definition then so will $\Omega' = \omega\Omega$ where ω is any smooth function on $\bar{M}$ that is positive on $M \cup \partial M$. We can use this "gauge freedom" to simplify things further. If we replace Ω with Ω' then we must replace $\bar{g}_{ab}$ with $\bar{g}'_{ab} = \omega^2 \bar{g}_{ab}$. The primed version of the quantity we just showed can be smoothly extended to $\mathcal{I}^+$ is then

$$\begin{aligned} \Omega'^{-1}\bar{g}'^{cd}\partial_c\Omega'\partial_d\Omega' &= \omega^{-3}\bar{g}^{cd}\left(\Omega\partial_c\omega\partial_d\omega + 2\omega\partial_c\Omega\partial_d\omega + \omega^2\Omega^{-1}\partial_c\Omega\partial_d\Omega\right) \\ &= \omega^{-1}\left(2n^a\partial_a \log\omega + \Omega^{-1}\bar{g}^{cd}\partial_c\Omega\partial_d\Omega\right) \qquad \text{on } \mathcal{I}^+ \end{aligned} \tag{5.18}$$

where

$$n^a = \bar{g}^{ab}\partial_b\Omega \tag{5.19}$$

is normal to $\mathcal{I}^+$ and hence also tangent to the null geodesic generators of $\mathcal{I}^+$. We can therefore choose ω to satisfy

$$n^a\partial_a \log\omega = -\frac{1}{2}\Omega^{-1}\bar{g}^{cd}\partial_c\Omega\partial_d\Omega \qquad \text{on } \mathcal{I}^+ \tag{5.20}$$

since this is just an ordinary differential equation along each generator of $\mathcal{I}^+$. Pick an S^2 cross-section of $\mathcal{I}^+$, i.e., an $S^2 \subset \mathcal{I}^+$ which intersects each generator precisely once. There is a unique solution of this differential equation for any (positive) choice of ω on this cross-section. We've now shown that the RHS of (5.18) vanishes on $\mathcal{I}^+$, i.e., that we can choose a gauge for which

$$\Omega^{-1}\bar{g}^{cd}\partial_c\Omega\partial_d\Omega = 0 \qquad \text{on } \mathcal{I}^+ \tag{5.21}$$

Evaluating (5.17) on $\mathcal{I}^+$ now gives

$$2\bar{\nabla}_a\bar{\nabla}_b\Omega + \bar{g}_{ab}\bar{g}^{cd}\bar{\nabla}_c\bar{\nabla}_d\Omega = 0 \qquad \text{on } \mathcal{I}^+ \tag{5.22}$$

Contracting this equation gives $\bar{g}^{cd}\bar{\nabla}_c\bar{\nabla}_d\Omega = 0$. Substituting back in we obtain

$$\bar{\nabla}_a\bar{\nabla}_b\Omega = 0 \qquad \text{on } \mathcal{I}^+ \tag{5.23}$$

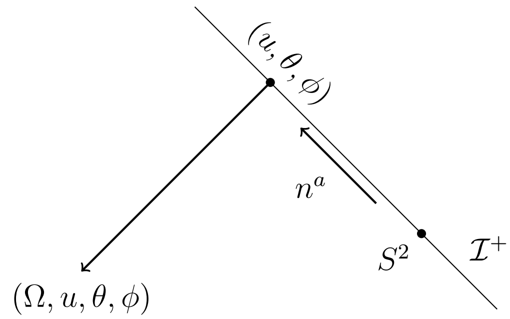


Figure 31. Coordinates near future null infinity.

and hence

$$\bar{\nabla}_a n^b = 0 \qquad \text{on } \mathcal{I}^+ \tag{5.24}$$

In particular we have

$$n^a \bar{\nabla}_a n^b = 0 \qquad \text{on } \mathcal{I}^+ \tag{5.25}$$

so, in this gauge, n^a is tangent to *affinely parameterized* (w.r.t. $\bar{g}$) null geodesic generators of $\mathcal{I}^+$. Furthermore, (5.24) shows that these generators have vanishing expansion and shear.

We introduce coordinates near $\mathcal{I}^+$ as follows (see Fig. 31). In our choice of gauge, we still have the freedom to choose ω on a S^2 cross-section of $\mathcal{I}^+$. A standard result is that any Riemannian metric on S^2 is conformal to the unit round metric on S^2. Hence we can choose ω so that the metric on our S^2 induced by $\bar{g}$ (i.e. the pull-back of $\bar{g}$ to this S^2) is the unit round metric. Introduce coordinates (θ, ϕ) on this S^2 so that the unit round metric takes the usual form $d\theta^2 + \sin^2\theta d\phi^2$. Now assign coordinates (u, θ, ϕ) to the point parameter distance u along the integral curve of n^a through the point on this S^2 with coordinates (θ, ϕ). This defines a coordinate chart on $\mathcal{I}^+$ with the property that the generators of $\mathcal{I}^+$ are lines of constant θ, ϕ with affine parameter u.

On $\mathcal{I}^+$ consider the vectors that are orthogonal (w.r.t. $\bar{g}$) to the 2-spheres of constant u, i.e., orthogonal to $\partial/\partial\theta$ and $\partial/\partial\phi$. Such vectors form a 2d subspace of the tangent space. In 2d, there are only two distinct null directions. Hence there are two distinct null directions orthogonal to the 2-spheres of constant u. One of these is tangent to $\mathcal{I}^+$ so pick the other one, which points into M.

Consider the null geodesics whose tangent at $\mathcal{I}^+$ is in this direction. Extend (u, θ, ϕ) off $\mathcal{I}^+$ by defining them to be constant along these null geodesics. Finally, since $d\Omega \neq 0$ on $\mathcal{I}^+$, we can use Ω as a coordinate near $\mathcal{I}^+$. We now have a coordinate chart $(u, \Omega, \theta, \phi)$ defined in a neighbourhood of $\mathcal{I}^+$, with $\mathcal{I}^+$ given by $\Omega = 0$.

By construction we have a coordinate chart with $n^a = \partial/\partial u$ on $\mathcal{I}^+$. Hence $n^\mu = \delta^\mu_u$. But the definition of n^a implies $\partial_\mu \Omega = \bar{g}_{\mu\nu} n^\nu$ from which we deduce $\bar{g}_{u\mu} = \delta^\Omega_\mu$ at $\Omega = 0$. Since (u, θ, ϕ) don't vary along the null geodesics pointing into M, the tangent vector to these geodesics must be proportional to $\partial/\partial\Omega$. Since the geodesics are null we must therefore have $\bar{g}_{\Omega\Omega} = 0$ for all Ω. We also know that these geodesics are orthogonal to $\partial/\partial\theta$ and $\partial/\partial\phi$ on $\mathcal{I}^+$ hence we have $\bar{g}_{\Omega\theta} = \bar{g}_{\Omega\phi} = 0$ at $\Omega = 0$.

Now consider the gauge condition (5.23). Writing this out in our coordinate chart, it reduces to

$$0 = \bar{\Gamma}^\Omega_{\mu\nu} = \frac{1}{2}\bar{g}^{\Omega\rho}\left(\bar{g}_{\rho\mu,\nu} + \bar{g}_{\rho\nu,\mu} - \bar{g}_{\mu\nu,\rho}\right) = \frac{1}{2}\left(\bar{g}_{u\mu,\nu} + \bar{g}_{u\nu,\mu} - \bar{g}_{\mu\nu,u}\right) \qquad \text{at } \Omega = 0 \tag{5.26}$$

where we used $\bar{g}^{\Omega\rho} = \bar{g}^{\nu\rho}(d\Omega)_\nu = n^\rho = \delta^\rho_u$. Taking μ and ν to be θ or ϕ, we have $\bar{g}_{u\mu,\nu} = \bar{g}_{u\nu,\mu} = 0$ so we learn that $\bar{g}_{\mu\nu,u} = 0$ at $\Omega = 0$, i.e., the θ, ϕ components of the metric $\bar{g}$ on $\mathcal{I}^+$ don't depend on u. Since we know that this metric is the unit round metric when $u = 0$, it must be the unit round metric for all u.

We have now deduced the form of the unphysical metric on $\mathcal{I}^+$:

$$\bar{g}|_{\Omega=0} = 2du d\Omega + d\theta^2 + \sin^2 d\phi^2 \tag{5.27}$$

For small $\Omega \neq 0$, the metric components will differ from the above result by $\mathcal{O}(\Omega)$ terms. However, by setting $\nu = \Omega$ in (5.26) and taking μ to be u, θ or ϕ, we learn that $\bar{g}_{u\mu,\Omega} = 0$ at $\Omega = 0$ so smoothness of $\bar{g}$ implies that $\bar{g}_{u\mu} = \mathcal{O}(\Omega^2)$ for $\mu = u, \theta, \phi$.

Finally we can write down the physical metric $g = \Omega^{-2}\bar{g}$. It is convenient to define a new coordinate $r = 1/\Omega$ so that $\mathcal{I}^+$ corresponds to $r \to \infty$. After a finite shift in r, the metric can be brought to the form

$$g = -2du dr + r^2\left(d\theta^2 + \sin^2\theta d\phi^2\right) + \ldots \tag{5.28}$$

for large r, where the ellipsis refers to corrections that are subleading at large r. The leading terms written above are simply the metric of Minkowski spacetime. If one converts this to inertial frame coordinates (t, x, y, z) so that the leading order metric is $\mathrm{diag}(-1, 1, 1, 1)$ then the correction terms are all of order $1/r$. Hence the metric of an asymptotically flat spacetime does indeed approach the Minkowski metric at $\mathcal{I}^+$.

Finally we can explain condition 6 of our definition of asymptotic flatness. Nothing in the above construction guarantees that the range of u is $(-\infty, \infty)$ as it is in Minkowski spacetime. We would not want to regard a spacetime as asymptotically flat if $\mathcal{I}^+$ "ends" at some finite value of u. Recall that u is the affine parameter along the generators of $\mathcal{I}^+$ so if this happens then the generators of $\mathcal{I}^+$ would be incomplete. Condition 6 eliminates this possibility.

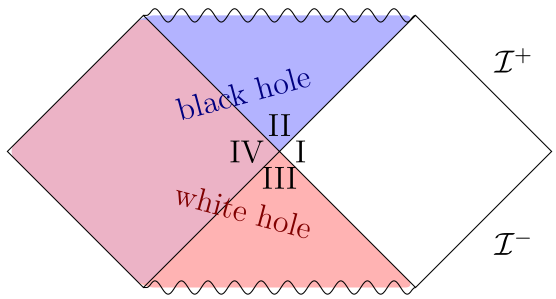


Figure 32. Black hole and white hole regions in the Kruskal spacetime.

Definition. $\mathcal{I}^+$ is *complete* if, in the gauge (5.23), the generators of $\mathcal{I}^+$ are complete (i.e. the affine parameter extends to $\pm\infty$). Similarly for $\mathcal{I}^-$.

This completeness assumption will be important when we discuss weak cosmic censorship.

5.3 Definition of a black hole

We can now make precise our definition of a black hole as a region of an asymptotically flat spacetime from which it is impossible to send a signal to infinity. $\mathcal{I}^+$ is a subset of our unphysical spacetime $(\bar{M}, \bar{g})$ so we can define $J^-(\mathcal{I}^+) \subset \bar{M}$. The set of points of M that can send a signal to $\mathcal{I}^+$ is $M \cap J^-(\mathcal{I}^+)$. We define the black hole region to be the complement of this region, and the future event horizon to be the boundary of the black hole region:

Definition. Let (M, g) be a spacetime that is asymptotically flat at null infinity. The *black hole region* is $\mathcal{B} = M \backslash [M \cap J^-(\mathcal{I}^+)]$ where $J^-(\mathcal{I}^+)$ is defined using the unphysical spacetime $(\bar{M}, \bar{g})$. The *future event horizon* is $\mathcal{H}^+ = \dot{\mathcal{B}}$ (the boundary of $\mathcal{B}$ in M), equivalently $\mathcal{H}^+ = M \cap \dot{J}^-(\mathcal{I}^+)$. Similarly, the *white hole region* is $\mathcal{W} = M \backslash [M \cap J^+(\mathcal{I}^-)]$ and the *past event horizon* is $\mathcal{H}^- = \dot{\mathcal{W}} = M \cap \dot{J}^+(\mathcal{I}^-)$.

One can construct examples of spacetimes with a non-empty black hole region simply by deleting sets of points from Minkowski spacetime. However, we can eliminate such trivial examples by restricting attention to spacetimes that are the maximal development of geodesically complete, asymptotically flat initial data.

In the Kruskal spacetime, no causal curve from region II or IV can reach $\mathcal{I}^+$ hence $\mathcal{B}$ is the union of regions II and IV (including the boundary $U = 0$ where $r = 2M$). $\mathcal{H}^+$ is the surface $U = 0$. $\mathcal{W}$ is the union of regions III and IV (including the boundary $V = 0$). $\mathcal{H}^-$ is the surface $V = 0$. See Fig. 32.

Theorems 2 and 3 of section 4.11 imply that $\mathcal{H}^\pm$ are null hypersurfaces. Theorem 3 (time reversed) implies that the generators of $\mathcal{H}^+$ cannot have future endpoints.

However, they can have past endpoints. This happens in the spacetime describing spherically symmetric gravitational collapse, with Penrose diagram:

The generators of $\mathcal{H}^+$ have a past endpoint at p, which is the point at which the black hole forms. So null generators can enter $\mathcal{H}^+$ but they cannot leave it. Note that the sets $\mathcal{W}$ and $\mathcal{H}^-$ are empty in this spacetime.

We will need a extra technical condition to prove useful things about black holes:

Definition. An asymptotically flat spacetime (M, g) is *strongly asymptotically predictable* if there exists an open region $\bar{V} \subset \bar{M}$ such that $\overline{M \cap J^-(\mathcal{I}^+)} \subset \bar{V}$ and $(\bar{V}, \bar{g})$ is globally hyperbolic.

This definition implies that $(M \cap \bar{V}, g)$ is a globally hyperbolic subset of M. Roughly speaking, there is a globally hyperbolic region $M \cap \bar{V}$ of spacetime consisting of the region not in $\mathcal{B}$ together with a neighbourhood of $\mathcal{H}^+$. It ensures that physics is predictable on, and outside, $\mathcal{H}^+$. A simple consequence of this definition is the result that a black hole cannot bifurcate (split into two):

Theorem. Let (M, g) be strongly asymptotically predictable and let Σ_1, Σ_2 be Cauchy surfaces for $\bar{V}$ with $\Sigma_2 \subset I^+(\Sigma_1)$. Let B be a connected component of $\mathcal{B} \cap \Sigma_1$. Then $J^+(B) \cap \Sigma_2$ is contained within a connected component of $\mathcal{B} \cap \Sigma_2$.

Proof. (See Fig. 33.) Global hyperbolicity implies that every causal curve from Σ_1 intersects Σ_2 and vice-versa. Note that $J^+(B) \subset \mathcal{B}$ hence $J^+(B) \cap \Sigma_2 \subset \mathcal{B} \cap \Sigma_2$. Assume $J^+(B) \cap \Sigma_2$ is not contained within a single connected component of $\mathcal{B} \cap \Sigma_2$. Then we can find disjoint open sets $O, O' \subset \Sigma_2$ such that $J^+(B) \cap \Sigma_2 \subset O \cup O'$ with $J^+(B) \cap O \neq \emptyset$, $J^+(B) \cap O' \neq \emptyset$. Then $B \cap I^-(O)$ and $B \cap I^-(O')$ are non-empty and $B \subset I^-(O) \cup I^-(O')$. Now $p \in B$ cannot lie in both $I^-(O)$ and $I^-(O')$ for then we could divide future-directed timelike geodesics from p into two sets according to whether they intersected O or O', and hence divide the future-directed timelike vectors at p into two disjoint open sets, contradicting connectedness of the future light-cone at p. Hence the open sets $B \cap I^-(O)$ and $B \cap I^-(O')$ are disjoint open sets whose union is B. This contradicts the connectedness of B.

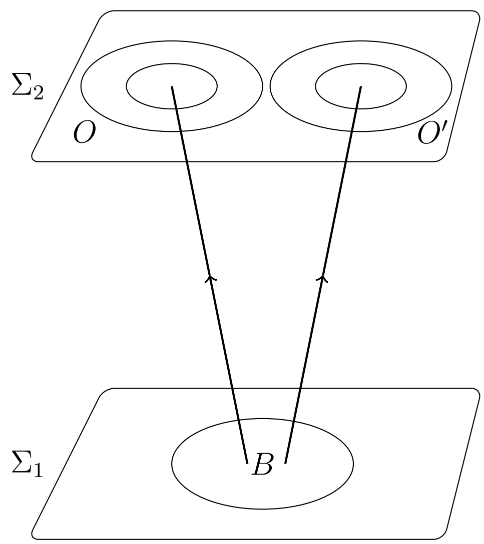


Figure 33. Bifurcation of a black hole.

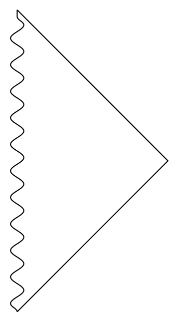


Figure 34. Penrose diagram of $M < 0$ Schwarzschild solution. The curvature singularity at $r = 0$ is naked.

5.4 Weak cosmic censorship

In our Penrose diagram for spherically symmetric gravitational collapse, the singularity at $r = 0$ is hidden behind the event horizon: no signal from the singularity can reach $\mathcal{I}^+$. (More precisely: no inextendible incomplete causal geodesic reaches $\mathcal{I}^+$.) This is not true for the Kruskal spacetime, where a signal from the white hole curvature singularity *can* reach $\mathcal{I}^+$: it is a *naked singularity*. Similarly, the curvature singularity of the $M < 0$ Schwarzschild solution is naked: see Fig. 34.

The singularity theorems tell us that gravitational collapse results in the formation of a singularity (i.e. geodesic incompleteness). But could this singularity be naked?

If we have a spherically symmetric collapsing star then Birkhoff's theorem tells us that the exterior of the star is given by the Schwarzschild solution, with the same (positive) mass as the star. This gives the standard diagram for gravitational collapse to form a black hole. However, this is just a consequence of spherical symmetry and Birkhoff's theorem. With spherical symmetry, the dynamics of the gravitational field is trivial: there are no gravitational waves (and no electromagnetic waves if there is a Maxwell field).

In order to make the dynamics more interesting we will assume that the matter in our spacetime includes a scalar field. This allows us to maintain the convenience of spherical symmetry, i.e., the use of Penrose diagrams, whilst circumventing Birkhoff's theorem. If the scalar field is non-trivial outside the collapsing matter then Birkhoff's theorem doesn't apply. We emphasize that the only reason for including this scalar field is to make the dynamics richer and therefore give us an idea of what is possible in the more general situation without spherical symmetry.

It is now tempting to draw the following diagram describing collapse to form a naked singularity:

(With the scalar field, we can no longer define a sharp boundary to the collapsing matter so the surface of the star is not precisely defined.) Imagine starting from initial data on Σ as shown. This data describes a collapsing star. The initial data is geodesically complete and asymptotically flat. When the star collapses to zero size, a timelike singularity forms. This is naked because it can send a signal to $\mathcal{I}^+$.

This diagram is misleading. Note the presence of a future Cauchy horizon $H^+(\Sigma)$ which bounds the maximal development of Σ. The spacetime beyond $H^+(\Sigma)$ is not determined by data on Σ. Hence we cannot say what happens beyond $H^+(\Sigma)$: one would need extra information (new laws of physics) to do so. So it is incorrect to draw a diagram as above. Instead we should draw just the maximal development of the initial data on Σ:

This spacetime does not have a singularity which can send a signal to $\mathcal{I}^+$. But the spacetime shown is pathological in two respects. First, even though we started from geodesically complete, asymptotically flat initial data, the maximal development is extendible. Hence strong cosmic censorship is violated. Second, the spacetime does not satisfy our definition of asymptotic flatness. This is because $\mathcal{I}^+$ is not complete: only part of it is present. The *weak* cosmic censorship property asserts that the latter behaviour does not occur:

Conjecture (weak cosmic censorship). Let (Σ, h_{ab}, K_{ab}) be a geodesically complete, asymptotically flat, initial data set. Let the matter fields obey hyperbolic equations and satisfy the dominant energy condition. Then generically the maximal development of this initial data is an asymptotically flat spacetime (in particular it has a complete $\mathcal{I}^+$) that is strongly asymptotically predictable.

Just like strong cosmic censorship, this conjecture refers only to the maximal development, i.e., to the region of spacetime that can be predicted uniquely from the initial data. This conjecture captures the idea that a "naked singularity would lead to an incomplete $\mathcal{I}^+$" without referring to any actual singularity.

The word "generically" is included because it is known that there exist examples which violate the conjecture if this word is omitted. However, such examples are "fine-tuned", i.e., if one introduces an appropriate measure on the space of initial data then the set of data which violates the conjecture is of measure zero. For example, consider gravity coupled to a massless scalar field, with spherical symmetry. This system was studied in the early 1990s by Christodoulou (rigorously) and Choptuik (numerically). One can construct a 1-parameter family of initial data labelled by a parameter p with the following property. There exists p_* such that for $p < p_*$, the scalar field simply disperses whereas for $p > p_*$ it collapses to form a black hole. These cases with $p \neq p_*$ respect the weak cosmic censorship conjecture. However, the "critical" solution with $p = p_*$ violates the conjecture. But this solution is fine-tuned and hence non-generic.

In spite of the name, weak cosmic censorship is not implied by strong cosmic censorship: the two conjectures are logically independent. This is shown in the following Penrose diagrams:

The first diagram violates strong but not weak, the second violates weak but not strong and the diagram we drew previously violates both weak and strong.

Historically, a very popular model for gravitational collapse consists of gravity coupled to a pressureless perfect fluid (”dust”), with spherical symmetry. For initial data consisting of a homogeneous ball of dust (i.e. constant density), it is known that gravitational collapse leads to formation of a black hole in the standard way. However, Christodoulou showed that if one considers a spherically symmetric but inhomogenous ball of dust (i.e., the density ρ depends on radius r) then both cosmic censorship conjectures are *false* (if one interprets ”generic” as meaning ”generic within the class of spherically symmetric initial data”). Generically, a singularity forms at the centre of the ball before an event horizon forms. However, it is believed that this model is unphysical because of the neglect of pressure.

For the case of gravity coupled to a massless scalar field, Christodoulou has proved that both cosmic censorship conjectures are *true*, again within the restricted class of spherically symmetric initial data. In this model, generic initial data either disperses (and settles down to flat spacetime at late time), or undergoes gravitational collapse to form a black hole.

Further evidence for the validity of weak cosmic censorship comes from the Penrose inequality (to be discussed later) and many numerical simulations e.g. of gravitational collapse, or black hole collisions.

5.5 Apparent horizon

Note that the definition of $\mathcal{B}$ and $\mathcal{H}^+$ is non-local: to determine whether or not $p \in \mathcal{B}$ we must establish whether there exists a causal curve from p to $\mathcal{I}^+$. This requires knowledge of the behaviour of the spacetime to the future of p, it can't be determined by measurements in a neighbourhood of p. This makes it difficult to determine the location of $\mathcal{H}^+$ e.g. in a numerical simulation. However, determining whether or not a spacelike 2-surface is trapped can be done locally. Furthermore, these must lie inside $\mathcal{B}$ (if weak cosmic censorship is correct):

Theorem. Let T be a trapped surface in a strongly asymptotically predictable spacetime obeying the null energy condition. Then $T \subset \mathcal{B}$.

Proof (sketch). Assume there exists $p \in T$ such that $p \notin \mathcal{B}$, i.e., $p \in J^-(\mathcal{I}^+)$. Then there exists a causal curve from p to $\mathcal{I}^+$. One can use strong asymptotic predictability to show that this implies that $\dot{J}^+(T)$ must intersect $\mathcal{I}^+$, i.e., there exists $q \in \mathcal{I}^+$ with $q \in \dot{J}^+(T)$. Theorem 3 of section 4.10 implies that q lies on a null geodesic γ from $r \in T$ that is orthogonal to T and has no point conjugate to r along γ. Since T is trapped, the expansion of the null geodesics orthogonal to T is negative at r and hence (from section 4.10) $\theta \to \infty$ within finite affine parameter along γ. So there exists a point s conjugate to r along γ, a contradiction. □

In a numerical simulation one considers a foliation of the spacetime by Cauchy surfaces Σ_t labelled by a time function t. Then "the black hole region at time t" is $B_t \equiv \mathcal{B} \cap \Sigma_t$ and the "event horizon at time t" is $H_t \equiv \mathcal{H}^+ \cap \Sigma_t$. We can't determine B_t just from the solution on Σ_t. However, we can investigate whether there exist trapped surfaces on Σ_t. If such surfaces exist then the above theorem implies that B_t is non-empty.

Definition. Let Σ_t be a Cauchy surface in a globally hyperbolic spacetime (M, g). The *trapped region* $\mathcal{T}_t$ of Σ is the set of points $p \in \Sigma_t$ for which there exists a trapped surface S with $p \in S \subset \Sigma_t$. The *apparent horizon* $\mathcal{A}_t$ is the boundary of $\mathcal{T}_t$.

(Note that several different definitions of apparent horizon appear in the literature.) If weak cosmic censorship is correct then $\mathcal{T}_t \subset \mathcal{B}$ which implies that $\mathcal{A}_t \subset \mathcal{B}$ so the apparent horizon always lies *inside* (or on) the event horizon. It is natural to hope that $\mathcal{T}_t$ is a reasonable approximation to B_t, and that $\mathcal{A}_t$ is a reasonable approximation to H_t. Whether or not this is actually true can depend on how the surfaces Σ_t are chosen. For spherically symmetric Cauchy surfaces in the Kruskal spacetime, one has $\mathcal{A}_t = H_t$. However, one can find non-spherically symmetric Cauchy surfaces which enter the black hole region and come arbitrarily close to the singularity but do not contain trapped surfaces (Iyer and Wald 1991).

By continuity, one expects $\mathcal{A}_t$ to be a marginally trapped surface. This is how its location is determined in numerical simulations.

6 Charged black holes

In this chapter, we will discuss the Reissner-Nordstrom solution, which describes a charged, spherically symmetric black hole. Large imbalances of charge don't occur in nature, so matter undergoing gravitational collapse would be expected to be almost neutral. Furthermore, a charged black hole would preferentially attract particles of opposite charge and hence gradually lose its charge. Hence charged black holes are unlikely to be important in astrophysics. However, they have played an important role in quantum gravity, especially in string theory.

6.1 The Reissner-Nordstrom solution

The action for Einstein-Maxwell theory is

$$S = \frac{1}{16\pi} \int d^4x \sqrt{-g} \left(R - F^{ab} F_{ab} \right) \tag{6.1}$$

where $F = dA$ with A a 1-form potential. Note that the normalisation of F used here differs from the standard particle physics normalisation. The Einstein equation is

$$R_{ab} - \frac{1}{2}Rg_{ab} = 2\left(F_a{}^c F_{bc} - \frac{1}{4}g_{ab}F^{cd}F_{cd}\right) \tag{6.2}$$

and the Maxwell equations are

$$\nabla^b F_{ab} = 0 \qquad dF = 0 \tag{6.3}$$

There is a generalisation of Birkhoff's theorem to this theory:

Theorem. The unique spherically symmetric solution of the Einstein-Maxwell equations with non-constant area radius function r is the *Reissner-Nordstrom* solution:

$$\begin{aligned} ds^2 &= -\left(1 - \frac{2M}{r} + \frac{e^2}{r^2}\right)dt^2 + \left(1 - \frac{2M}{r} + \frac{e^2}{r^2}\right)^{-1} dr^2 + r^2 d\Omega^2 \\ A &= -\frac{Q}{r}dt - P\cos\theta d\phi \qquad e = \sqrt{Q^2 + P^2} \end{aligned} \tag{6.4}$$

This solution has 3 parameters: M, Q, P. We will show later that these are the mass, electric charge and magnetic charge respectively (there is no evidence that magnetic charge occurs in nature but it is allowed by the equations).

Several properties are similar to the Schwarzschild solution: the RN solution is static, with timelike Killing vector field $k^a = (\partial/\partial t)^a$. The RN solution is asymptotically flat at null infinity in the same way as the Schwarzschild solution.

If r is constant then the above theorem doesn't apply. In this case, one obtains the Robinson-Bertotti ($AdS_2 \times S^2$) solution discussed on examples sheet 2.

To discuss the properties of this solution, it is convenient to define

$$\Delta = r^2 - 2Mr + e^2 = (r - r_+)(r - r_-) \qquad r_\pm = M \pm \sqrt{M^2 - e^2} \tag{6.5}$$

so the metric is

$$ds^2 = -\frac{\Delta}{r^2}dt^2 + \frac{r^2}{\Delta}dr^2 + r^2 d\Omega^2 \tag{6.6}$$

If $M < e$ then $\Delta > 0$ for $r > 0$ so the above metric is smooth for $r > 0$. There is a curvature singularity at $r = 0$. This is a naked singularity, just like in the $M < 0$ Schwarzschild spacetime. Dynamical formation of such a singularity is excluded by the cosmic censorship hypotheses. If one considers a spherically symmetric ball of charged matter with $M < e$ then electromagnetic repulsion dominates over gravitational attraction so gravitational collapse does not occur. Note that elementary particles (e.g. electrons) can have $M < e$ but these are intrinsically quantum mechanical.

6.2 Eddington-Finkelstein coordinates

The special case $M = e$ will be discussed later so consider the case $M > e$. Δ has simple zeros at $r = r_\pm > 0$. These are coordinate singularities. To see this, we can define Eddington-Finkelstein coordinates in exactly the same way as we did for the Schwarzschild solution. Start with $r > r_+$ and define

$$dr_* = \frac{r^2}{\Delta} dr \tag{6.7}$$

Integrating gives

$$r_* = r + \frac{1}{2\kappa_+} \log \left| \frac{r - r_+}{r_+} \right| + \frac{1}{2\kappa_-} \log \left| \frac{r - r_-}{r_-} \right| + \text{const.} \tag{6.8}$$

where

$$\kappa_\pm = \frac{r_\pm - r_\mp}{2r_\pm^2} \tag{6.9}$$

Now let

$$u = t - r_* \qquad v = t + r_* \tag{6.10}$$

In ingoing EF coordinates (v, r, θ, ϕ), the RN metric becomes

$$ds^2 = -\frac{\Delta}{r^2} dv^2 + 2dvdr + r^2 d\Omega^2 \tag{6.11}$$

This is now smooth for any $r > 0$ hence we can analytically continue the metric into a new region $0 < r < r_+$. There is a curvature singularity at $r = 0$. A surface of constant r has normal $n = dr$ and hence is null when $g^{rr} = \Delta/r^2 = 0$. It follows that the surfaces $r = r_\pm$ are null hypersurfaces.

Exercise. Show that r decreases along any future-directed causal curve in the region $r_- < r < r_+$.

It follows from this that no point in the region $r < r_+$ can send a signal to $\mathcal{I}^+$ (since $r = \infty$ at $\mathcal{I}^+$). Hence this spacetime describes a black hole. The black hole region is $r \leq r_+$ and the future event horizon is the null hypersurface $r = r_+$.

Similarly, if one uses outgoing EF coordinates one obtains the metric

$$ds^2 = -\frac{\Delta}{r^2} du^2 - 2dudr + r^2 d\Omega^2 \tag{6.12}$$

and again one can analytically continue to a new region $0 < r \leq r_+$ and this is a white hole.

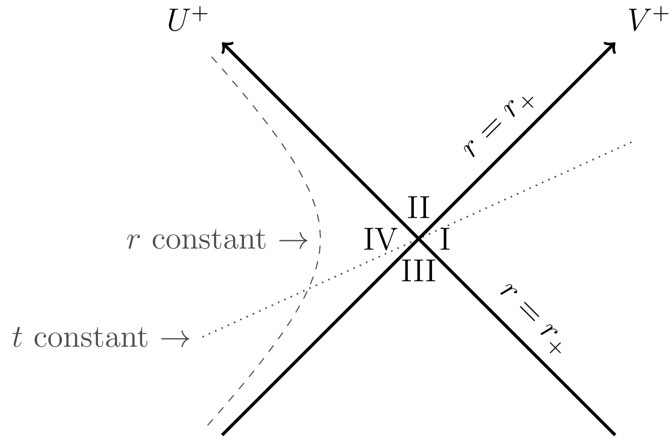


Figure 35. Reissner-Nordstrom solution in (U^+, V^+) coordinates.

6.3 Kruskal-like coordinates

To understand the global structure, define Kruskal-like coordinates

$$U^\pm = -e^{-\kappa_\pm u} \qquad V^\pm = \pm e^{\kappa_\pm v} \tag{6.13}$$

Starting in the region $r > r_+$, use coordinates (U^+, V^+, θ, ϕ) to obtain the metric

$$ds^2 = -\frac{r_+ r_-}{\kappa_+^2 r^2} e^{-2\kappa_+ r} \left(\frac{r - r_-}{r_-}\right)^{1+\kappa_+/|\kappa_-|} dU^+ dV^+ + r^2 d\Omega^2 \tag{6.14}$$

where $r(U^+, V^+)$ is defined implicitly by

$$-U^+ V^+ = e^{2\kappa_+ r} \left(\frac{r - r_+}{r_+}\right) \left(\frac{r_-}{r - r_-}\right)^{\kappa_+/|\kappa_-|} \tag{6.15}$$

The RHS is a monotonically increasing function of r for $r > r_-$. Initially we have $U^+ < 0$ and $V^+ > 0$ which gives $r > r_+$ but now we can analytically continue to $U^+ \geq 0$ or $V^+ \leq 0$. In particular, the metric is smooth and non-degenerate when $U^+ = 0$ or $V^+ = 0$. We obtain a diagram very similar to the Kruskal diagram: see Fig. 35. Just as for Kruskal, we have a pair of null hypersurfaces which intersect in the "bifurcation 2-sphere" $U^+ = V^+ = 0$, where $k^a = 0$. Surfaces of constant t are Einstein-Rosen bridges connection regions I and IV. The major difference with the Kruskal diagram is that we no longer have a curvature singularity in regions II and III because $r(U^+, V^+) > r_-$. However, from our EF coordinates, we know that it is possible to extend the spacetime into a region with $r < r_-$. Hence the above spacetime must be extendible. Phrasing things differently, we know (from the EF coordinates) that radial

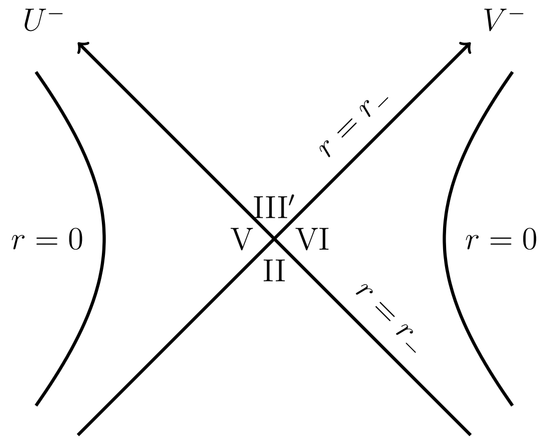


Figure 36. Reissner-Nordstrom solution in (U^-, V^-) coordinates.

null geodesics reach $r = r_-$ in finite affine parameter. Hence such geodesics will reach $U^+V^+ = -\infty$ in finite affine parameter so we have to investigate what happens there.

To do this, start in region II and use ingoing EF coordinates (v, r, θ, ϕ) (as we know these cover regions I and II). We can now define the retarded time coordinate u in region II as follows. First define a time coordinate $t = v - r_*$ in region II with r_* defined by (6.8). The metric in coordinates (t, r, θ, ϕ) takes the static RN form given above, with $r_- < r < r_+$. Now define u by $u = t - r_* = v - 2r_*$. Having defined u in region II we can now define the Kruskal coordinates $U^- < 0$ and $V^- < 0$ in region II using the formula above. In these coordinates, the metric is

$$ds^2 = -\frac{r_+ r_-}{\kappa_-^2 r^2} e^{2|\kappa_-|r} \left(\frac{r_+ - r}{r_+}\right)^{1+|\kappa_-|/\kappa_+} dU^- dV^- + r^2 d\Omega^2 \tag{6.16}$$

where $r(U^-, V^-) < r_+$ is given by

$$U^- V^- = e^{-2|\kappa_-|r} \left(\frac{r - r_-}{r_-}\right) \left(\frac{r_+}{r_+ - r}\right)^{|\kappa_-|/\kappa_+} \tag{6.17}$$

This can now be analytically continued to $U^- > 0$ or $V^- > 0$, giving the diagram shown in Fig. 36. We have new regions V and VI in which $0 < r < r_-$. These regions contain the curvature singularity at $r = 0$ ($U^- V^- = -1$), which is *timelike*. Region III′ is isometric to region III and so, by introducing new coordinates $(U^{+'}, V^{+'})$ this can be analytically to the future to give further new regions I′, II′ and IV′ as shown in Fig. 37. In this diagram, I′ and IV′ are new asymptotically flat regions isometric to I and IV. This procedure can be repeated indefinitely, to the future and past, so the maximal analytic extension of the RN solution contains infinitely many regions. The

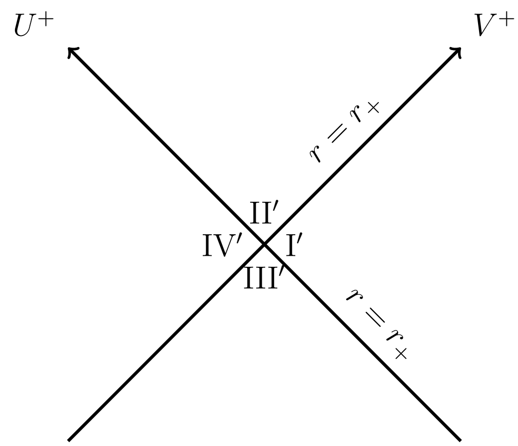


Figure 37. Regions I′, II′ and IV′ of the RN solution.

resulting Penrose diagram is shown in Fig. 38. It extends to infinity in both directions. By an appropriate choice of conformal factor, one can arrange that the singularity is a straight line.

6.4 Cauchy horizons

Consider a partial Cauchy surface Σ extending from spatial infinity in region I to spatial infinity in region IV. Assume that Σ is geodesically complete and asymptotically flat (with 2 ends). (An Einstein-Rosen bridge is an example of such a surface.). See Fig. 39. Note that $D^+(\Sigma)$ is bounded to the future by a Cauchy horizon $H^+(\Sigma)$ and $D^-(\Sigma)$ is bounded to the past by a Cauchy horizon $H^-(\Sigma)$. Both Cauchy horizons have $r = r_-$.

The existence of these Cauchy horizons means that most of the above Penrose diagram is *unphysical*. We should take seriously only the part of the diagram corresponding to $D(\Sigma)$ since this is the part that is uniquely determined by initial data on Σ. The solution outside $D(\Sigma)$ is not determined by this data: to obtain the above Penrose diagram one has to assume analyticity or spherical symmetry. But if we just assume that spacetime is smooth then there are infinitely many ways of extending $D(\Sigma)$.

The extendibility of $D(\Sigma)$ appears to violate strong cosmic censorship. But recall that the latter applies to *generic* initial data: violation of strong cosmic censorship would require that $D(\Sigma)$ is generically extendible for a sufficiently small perturbation of the initial data on Σ. (This could be a perturbation that breaks spherical symmetry or it could be a perturbation that preserves spherical symmetry but introduces a small amount of matter: a popular model is a massless scalar field.)

There is a lot of evidence that $D(\Sigma)$ is *not* extendible when the initial data on Σ is perturbed, i.e., strong cosmic censorship is respected. The physical mechanism

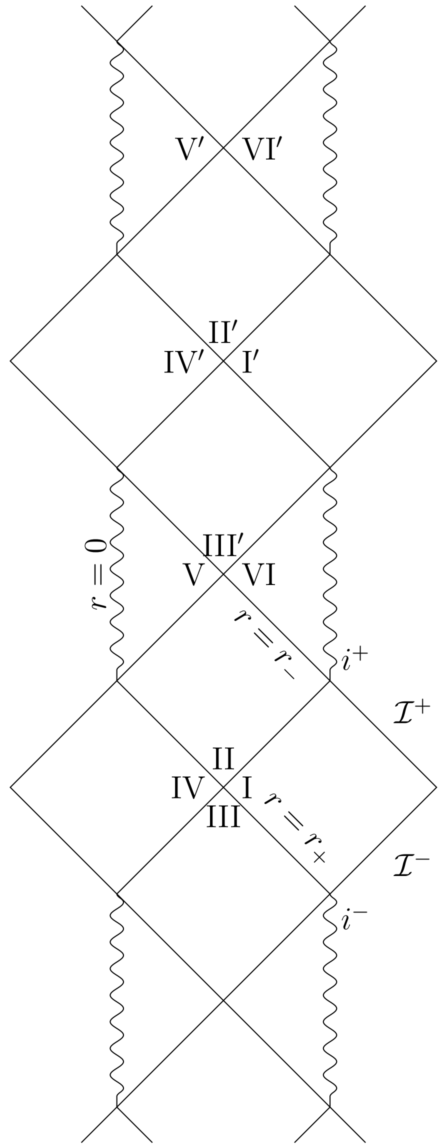


Figure 38. Penrose diagram of the maximally entended Reissner-Nordstrom solution.

for this can be understood as follows. Consider two observers A, B as shown in Fig. 40. A crosses $H^+(\Sigma)$ in region II whereas B stays in region I. Assume that B sends light signals to A at proper time intervals of 1 second. If B lives forever (!) then he sends infinitely many signals. From the Penrose diagram, it is clear that A receives *all* of these signals within a finite proper time as she crosses $H^+(\Sigma)$. Hence signals from region I undergo an infinite blueshift at $H^+(\Sigma)$. Therefore a tiny perturbation in region I will have an enormous energy (as measured by A) at $H^+(\Sigma)$. This suggests that the gravitational back reaction of a tiny perturbation in region I will become large

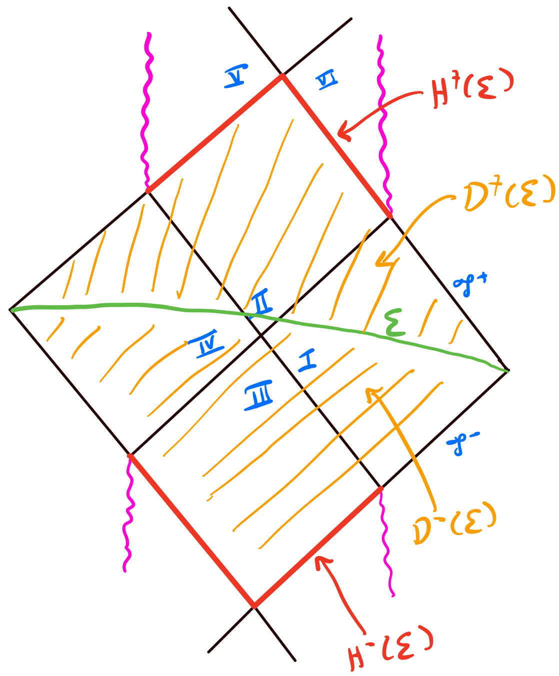


Figure 39. Partial Cauchy surface Σ, its domain of dependence and Cauchy horizons.

in region II. In other words, region II exhibits an instability. The effect of this might be to give a singularity, rather than a Cauchy horizon, in region II, thus rendering $D(\Sigma)$ inextendible in agreement with strong cosmic censorship.

A tractable model for studying this in detail is to consider Einstein-Maxwell theory coupled to a massless scalar field, assuming spherical symmetry. In this case, results of Dafermos (2012) strongly suggest that small perturbations of the initial data on Σ lead to a spacetime in which the Cauchy horizons are replaced by *null* curvature singularities. Hence strong cosmic censorship is respected (at least within the class of spherically symmetric initial data). For a charged black hole formed by gravitational collapse of (almost) spherically symmetric charged matter, it seems likely that the singularity will be partially null (near i^+) and partially spacelike.

6.5 Extreme RN

The RN solution with $M = e$ is called *extreme RN*. The metric is

$$ds^2 = -\left(1 - \frac{M}{r}\right)^2 dt^2 + \left(1 - \frac{M}{r}\right)^{-2} dr^2 + r^2 d\Omega^2 \tag{6.18}$$

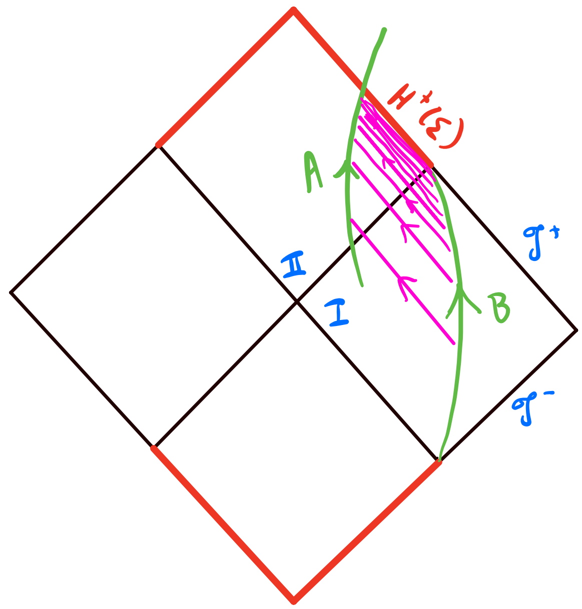


Figure 40. Argument for instability of Cauchy horizon in Reissner-Nordstrom spacetime.

Starting in the region $r > M$ one can define $dr_* = dr/(1 - M/r)^2$, i.e.,

$$r_* = r + 2M \log \left| \frac{r - M}{M} \right| - \frac{M^2}{r - M} \tag{6.19}$$

and introduce ingoing EF coordinates $v = t + r_\star$ so that the metric becomes

$$ds^2 = -\left(1 - \frac{M}{r}\right)^2 dv^2 + 2dvdr + r^2 d\Omega^2 \tag{6.20}$$

which can be analytically extended into the region $0 < r < M$, which is a black hole region. Similarly one can use outgoing EF coordinates to uncover a white hole region. Each of these can be analytically extended across an inner horizon. The Penrose diagram is shown in Fig. 41.

Note that $\mathcal{H}^\pm$ are Cauchy horizons for a surface of constant t. A novel feature of this solution is that a surface of constant t is not an Einstein-Rosen bridge connecting two asymptotically flat ends. Consider the proper length of a line of constant t, θ, ϕ from $r = r_0 > M$ to $r = M$:

$$\int_M^{r_0} \frac{dr}{1 - M/r} = \infty \tag{6.21}$$

Hence a surface of constant t exhibits an "infinite throat" shown in Fig. 42.

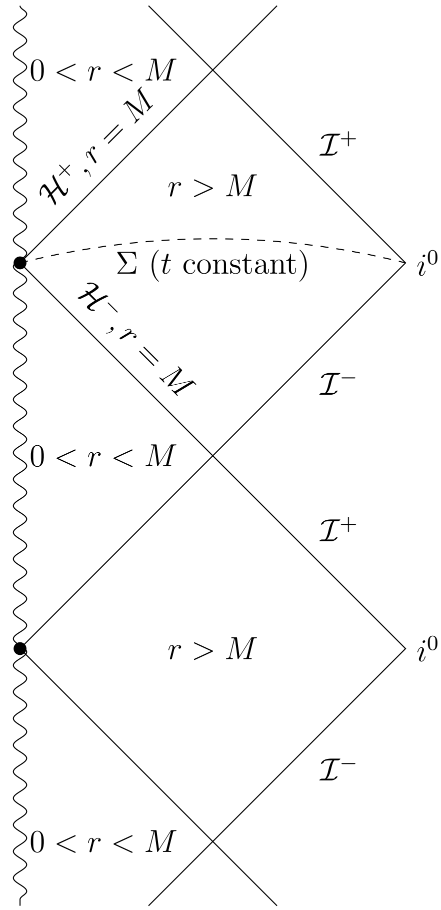


Figure 41. Penrose diagram of the extreme RN solution.

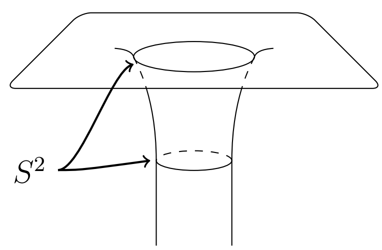


Figure 42. Infinite "throat" on a constant t surface of the extreme RN solution.

To understand the geometry near the horizon, let $r = M(1+\lambda)$. To leading order in λ,

$$ds^2 \approx -\lambda^2 dt^2 + M^2 \frac{d\lambda^2}{\lambda^2} + M^2 d\Omega^2 \tag{6.22}$$

This is the Robinson-Bertotti metric: a product of 2d anti-de Sitter spacetime (AdS_2) with S^2 (see examples sheet 2).

6.6 Majumdar-Papapetrou solutions

Introduce a new radial coordinates $\rho = r - M$ and assume $P = 0$. The extreme RN metric becomes

$$ds^2 = -H^{-2}dt^2 + H^2\left(d\rho^2 + \rho^2 d\Omega^2\right) \qquad H = 1 + \frac{M}{\rho} \tag{6.23}$$

this is a special case of the *Majumdar-Papapetrou* solution:

$$ds^2 = -H(\mathbf{x})^{-2}dt^2 + H(\mathbf{x})^2\left(dx^2 + dy^2 + dz^2\right) \qquad A = H^{-1}dt \tag{6.24}$$

where $\mathbf{x} = (x, y, z)$ and H obeys the Laplace equation in 3d Euclidean space:

$$\nabla^2 H = 0 \tag{6.25}$$

Choosing

$$H = 1 + \sum_{i=1}^{N} \frac{M_i}{|\mathbf{x} - \mathbf{x}_i|} \tag{6.26}$$

gives a static solution describing N extreme RN black holes of masses M_i at positions $\mathbf{x}_i$ (each of these is an S^2, not a point). Physically, such a solution exists because $M_i = Q_i$ for all i hence there is an exact cancellation of gravitational attraction and electromagnetic repulsion between the black holes.

7 Rotating black holes

In this chapter we will discuss the Kerr solution, which describes a stationary rotating black hole. The solution is considerably more complicated than the spherically symmetric solutions that we have discussed so far. We will start by explaining why the Kerr solution is believed to be the unique stationary black hole solution.

7.1 Uniqueness theorems

Black holes form by gravitational collapse, a time-dependent process. However, we would expect an isolated black hole eventually to settle down to a time-independent equilibrium state (this is actually a very fast process, occuring on a time scale set by the radius of the black hole: microseconds for a solar mass black hole). Hence it is desirable to classify all such equilibrium states, i.e., all possible stationary black hole solutions of the vacuum Einstein (or Einstein-Maxwell) equations.

First we will need to weaken slightly our definition of "stationary" to cover rotating black holes:

Definition. A spacetime asymptotically flat at null infinity is *stationary* if it admits a Killing vector field k^a that is timelike in a neighbourhood of $\mathcal{I}^\pm$. It is *static* if it is stationary and k^a is hypersurface-orthogonal.

It is conventional to normalize k^a so that $k^2 \to -1$ at $\mathcal{I}^\pm$. Sometimes the term "strictly stationary/static" is used if k^a is timelike everywhere, not just near $\mathcal{I}^\pm$. So Minkowski spacetime is strictly static. The Kruskal spacetime is static but not strictly static (because k^a is spacelike in regions II, III).

So far, we have discussed only spherically symmetric black holes. But rotating black holes cannot be spherically symmetric. However, they can be axisymmetric, i.e. "symmetric under rotations about an axis". For a stationary spacetime we define this as follows.

Definition. A spacetime asymptotically flat at null infinity is *stationary and axisymmetric* if (i) it is stationary; (ii) it admits a Killing vector field m^a that is spacelike near $\mathcal{I}^\pm$; (iii) m^a generates a 1-parameter group of isometries isomorphic to $U(1)$; (iv) $[k, m] = 0$.

(We can also define the notion of axisymmetry in a non-stationary spacetime by deleting (i) and (iv).) For such a spacetime, one can choose coordinates so that $k = \partial/\partial t$ and $m = \partial/\partial\phi$ with $\phi \sim \phi + 2\pi$.

Now recall that a spherically symmetric vacuum spacetime must be static, by Birkhoff's theorem. The converse of this is untrue: a static vacuum spacetime need not be spherically symmetric e.g. consider the spacetime outside a cube-shaped object. However, if the only object in the spacetime is a black hole then we have:

Theorem (Israel 1967, Bunting & Masood 1987). If (M, g) is a static, asymptotically flat, vacuum black hole spacetime that is suitably regular on, and outside an event horizon, then (M, g) is isometric to the Schwarzschild solution.

We will not attempt to describe precisely what "suitably regular" means here. This theorem establishes that static vacuum *multi* black hole solutions do not exist. There is an Einstein-Maxwell generalisation of this theorem, which states that such a solution is either Reissner-Nordstrom *or* Majumdar-Papapetrou.

For stationary black holes, we have the following:

Theorem (Hawking 1973, Wald 1992). If (M, g) is a stationary, non-static, asymptotically flat *analytic* solution of the Einstein-Maxwell equations that is suitably regular on, and outside an event horizon, then (M, g) is stationary and axisymmetric.

This is sometimes stated as "stationary implies axisymmetric" for black holes. But this theorem has the unsatisfactory assumption that the spacetime be analytic, i.e., the metric components can be expressed as convergent Taylor series about any point. This is unphysical: analyticity implies that the full spacetime is determined by its behaviour in the neighbourhood of a single point. If one accepts the above result, or simply assumes axisymmetry, then

Theorem (Carter 1971, Robinson 1975). If (M, g) is a stationary, axisymmetric, asymptotically flat vacuum spacetime suitably regular on, and outside, a connected event horizon then (M, g) is a member of the 2-parameter Kerr (1963) family of solutions. The parameters are mass M and angular momentum J.

These results lead to the expectation that the final state of gravitational collapse is generically a Kerr black hole. (There is also strong evidence for this coming from numerical simulations.) This implies that the final state is fully characterized by just 2 numbers: M and J. In contrast, the initial state can be arbitrarily complicated. Nearly all information about the initial state is lost during gravitational collapse (either by radiation to infinity, or by falling into the black hole), with just the 2 numbers M, J required to describe the final state on, and outside, the event horizon.

There is an Einstein-Maxwell generalization of the above theorem, which states that (M, g) should belong to the 4-parameter *Kerr-Newman* (1965) solution described in the next section.

7.2 The Kerr-Newman solution

This is a rotating, charged solution of Einstein-Maxwell theory. In *Boyer-Lindquist coordinates*, it is

$$\begin{aligned} ds^2 &= -\frac{\Delta}{\Sigma}\left(dt - a\sin^2\theta d\phi\right)^2 + \frac{\sin^2\theta}{\Sigma}\left((r^2+a^2)d\phi - adt\right)^2 + \frac{\Sigma}{\Delta}dr^2 + \Sigma d\theta^2 \\ A &= \frac{1}{\Sigma}\left[-Qr(dt - a\sin^2\theta d\phi) + P\cos\theta\left((r^2+a^2)d\phi - adt\right)\right] \end{aligned} \tag{7.1}$$

where

$$\Sigma = r^2 + a^2\cos^2\theta \qquad \Delta = r^2 - 2Mr + a^2 + e^2 \qquad e = \sqrt{Q^2+P^2} \tag{7.2}$$

At large r, the coordinates (t, r, θ, ϕ) reduce to spherical polar coordinates in Minkowski spacetime. In particular, (θ, ϕ) have their usual interpretation as angles on S^2 so $0 < \theta < \pi$ and $\phi \sim \phi + 2\pi$. It can be shown that the KN solution is asymptotically flat at null infinity.

The solution is stationary and axisymmetric with two commuting Killing vector fields:

$$k^a = \left(\frac{\partial}{\partial t}\right)^a \qquad m^a = \left(\frac{\partial}{\partial \phi}\right)^a \tag{7.3}$$

k^a is timelike near infinity although, as we will discuss, it is not globally timelike. The solution possesses a discrete isometry $t \to -t$, $\phi \to -\phi$ which simultaneously reverses the direction of time and the sense of rotation.

The solution has 4 parameters: M, a, Q, P. We'll see later that M is the mass, Q the electric charge, P the magnetic charge and $a = J/M$ where J is the angular momentum. When $a = 0$ the KN solution reduces to the RN solution. Note that $\phi \to -\phi$ has the same effect as $a \to -a$ so there is no loss of generality in assuming $a \geq 0$.

7.3 The Kerr solution

Set $Q = P = 0$ in the KN solution to get the Kerr solution of the vacuum Einstein equation. Let's analyze the structure of this solution. As we did for RN, write

$$\Delta = (r - r_+)(r - r_-) \qquad r_\pm = M \pm \sqrt{M^2 - a^2} \tag{7.4}$$

The solution with $M^2 < a^2$ describes a naked singularity so let's assume $M^2 > a^2$ (and discuss $M = a$ later). The metric is singular at $\theta = 0, \pi$ but these are just the usual coordinate singularities of spherical polars. The metric is also singular at $\Delta = 0$ (i.e. $r = r_\pm$) and at $\Sigma = 0$ (i.e. $r = 0, \theta = \pi/2$). Starting in the region $r > r_+$, the first singularity we have to worry about is at $r = r_+$. We will now show that this is a coordinate singularity. To see this, define *Kerr coordinates* (v, r, θ, χ) for $r > r_+$ by

$$dv = dt + \frac{r^2 + a^2}{\Delta} dr \qquad d\chi = d\phi + \frac{a}{\Delta} dr \tag{7.5}$$

which implies that in the new coordinates we have $\chi \sim \chi + 2\pi$ and

$$k^a = \left(\frac{\partial}{\partial v}\right)^a \qquad m^a = \left(\frac{\partial}{\partial \chi}\right)^a \tag{7.6}$$

The metric is (exercise)

$$\begin{aligned} ds^2 = &-\frac{\left(\Delta - a^2 \sin^2\theta\right)}{\Sigma} dv^2 + 2 dv dr - 2a \sin^2\theta \frac{(r^2 + a^2 - \Delta)}{\Sigma} dv d\chi \\ &- 2a \sin^2\theta d\chi dr + \left(\frac{(r^2 + a^2)^2 - \Delta a^2 \sin^2\theta}{\Sigma}\right) \sin^2\theta d\chi^2 + \Sigma d\theta^2 \end{aligned} \tag{7.7}$$

This metric is smooth and non-degenerate at $r = r_+$. It can be analytically continued through the surface $r = r_+$ into a new region with $0 < r < r_+$.

Proposition. The surface $r = r_+$ is a null hypersurface with normal

$$\xi^a = k^a + \Omega_H m^a \tag{7.8}$$

where

$$\Omega_H = \frac{a}{r_+^2 + a^2} \tag{7.9}$$

Proof. Exercise: Determine ξ_μ and show that $\xi_\mu dx^\mu|_{r=r_+}$ is proportional to dr. Hence (i) ξ_a is normal to the surface $r = r_+$ and (ii) $\xi^\mu \xi_\mu|_{r=r_+} = 0$ because $\xi^r = 0$.

Just as for RN, the region $r \leq r_+$ is (part of) the black hole region of this spacetime with $r = r_+$ (part of) the future event horizon $\mathcal{H}^+$.

In BL coordinates we have $\xi = \partial/\partial t + \Omega_H \partial/\partial\phi$. Hence $\xi^\mu \partial_\mu (\phi - \Omega_H t) = 0$ so $\phi = \Omega_H t + \text{const.}$ on orbits (integral curves) of ξ^a. Note that $\phi = \text{const.}$ on orbits of k^a. Hence particles moving on orbits of ξ^a rotate with angular velocity Ω_H with respect to a stationary observer (i.e. someone on an orbit of k^a). In particular, they rotate with this angular velocity w.r.t. a stationary observer at infinity. Since ξ^a is tangent to the generators of $\mathcal{H}^+$, it follows that these generators rotate with angular velocity Ω_H w.r.t. a stationary observer at infinity, so we interpret Ω_H as the angular velocity of the black hole.

7.4 Maximal analytic extension

The Kerr coordinates are analogous to the ingoing EF coordinates we used for RN. One can similarly define coordinates analogous to outgoing EF coordinates and use these to construct an analytic extension into a white hole region. Then, just as for RN, one can define Kruskal-like coordinates that cover all of these regions, as well as a new asymptotically flat region, i.e., there are regions analogous to regions I to IV of the analytically extended RN solution.

Just as for RN, the spacetime can be analytically extended across null hypersurfaces at $r = r_-$ in regions II and III. The resulting maximal analytic extension is similar to that of RN except for the behaviour near the singularity. In the Kerr case, it turns out that the curvature singularity has the structure of a ring and by passing through the ring one can enter a new asymptotically flat region with $r < 0$. One also finds that m^a becomes timelike near the singularity. The orbits of m^a are closed (because $\phi \sim \phi + 2\pi$) hence there are closed timelike curves near the singularity.

The Kerr solution is not spherically symmetric so one can't draw a Penrose diagram for it. However, if one considers the submanifold of the spacetime corresponding to the

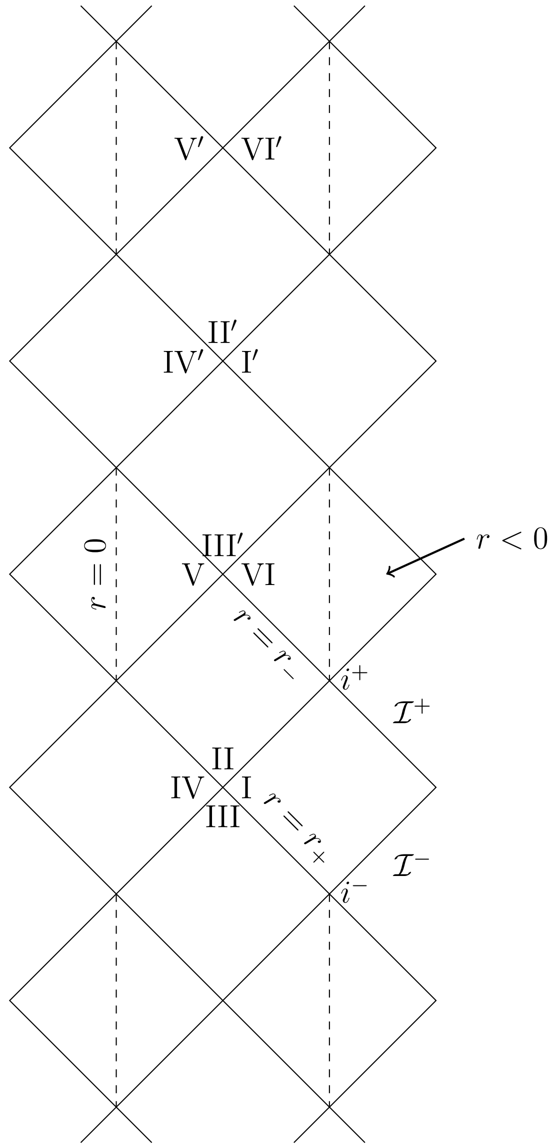


Figure 43. Penrose diagram of the maximally extended Kerr solution. The dotted lines denote the "ring" singularity at $r = 0$, $\theta = \pi/2$.

axis of symmetry ($\theta = 0$ or $\theta = \pi$) then, since this submanifold is two-dimensional, one can draw a Penrose diagram for it. Note that this submanifold is "totally geodesic", i.e., a geodesic initially tangent to it will remain tangent. (The same is true for the "equatorial plane" $\theta = \pi/2$.) The resulting diagram takes the following form shown in Fig. 43.

Most of this diagram is unphysical because, just as for RN, the null hypersurfaces $r = r_-$ are Cauchy horizons for a geodesically complete, asymptotically flat (with 2 ends) surface Σ. Hence the spacetime beyond $r = r_-$ is not determined uniquely by the data on Σ (unless one makes the unphysical assumption of analyticity). By the same argument as for RN, these Cauchy horizons are expected to be unstable against small perturbations in region I (or IV), with the perturbed spacetime exhibiting null or spacelike singularities instead of Cauchy horizons, in agreement with strong cosmic censorship.

When we studied the Schwarzschild solution, we saw that it describes the metric outside a spherical star. This was a consequence of Birkhoff's theorem. In contrast, the Kerr solution does *not* describe the spacetime outside a rotating star. This solution is expected to describe only the "final state" of gravitational collapse. One can't obtain a solution describing gravitational collapse to form a Kerr black hole simply by "gluing in" a ball of collapsing matter as we did for Schwarzschild. In particular, the spacetime during such collapse would be non-stationary because the collapse would lead to emission of gravitational waves.

Finally, the special case $M = a$ is called the *extreme Kerr* solution. It is a black hole solution with several properties similar to those of the extreme RN solution. In particular, surfaces of constant t exhibit an "infinite throat" and $\mathcal{H}^\pm$ are Cauchy horizons for surfaces of constant t.

7.5 Geodesics of Kerr-Newman

The Schwarzschild or Reissner-Nordstrom spacetimes have a lot of symmetry, which is the reason that the geodesic equation in these spacetimes is integrable, i.e., it can be reduced to first order ODEs. The Kerr-Newman solution only has two commuting Killing vectors, which is usually not enough to render the geodesic equation integrable. Remarkably, there is a "hidden symmetry" which ensures integrability.

Let's start with the Hamiltonian version of the geodesic equation in a general spacetime with coordinates x^μ. The Lagrangian for geodesics is $L(x, \dot{x}) = (1/2)g_{\mu\nu}(x)\dot{x}^\mu\dot{x}^\nu$ where a dot denotes a derivative w.r.t. an affine parameter τ. The momentum conjugate to x^μ is

$$p_\mu = \frac{\partial L}{\partial \dot{x}^\mu} = g_{\mu\nu}\dot{x}^\nu \tag{7.10}$$

so the Hamiltonian is $H = p_\mu \dot{x}^\mu - L$ which simplifies to

$$H(x, p) = \frac{1}{2} g^{\mu\nu}(x) p_\mu p_\nu \tag{7.11}$$

Hamilton's equations of motion are

$$\dot{x}^\mu = \frac{\partial H}{\partial p_\mu} = g^{\mu\nu} p_\nu \tag{7.12}$$

(recovering the definition of p_μ) and

$$\dot{p}_\mu = -\frac{\partial H}{\partial x^\mu} = -\frac{1}{2} g^{\nu\rho}{}_{,\mu} p_\nu p_\rho \tag{7.13}$$

If we eliminate p_μ using (7.10) then it is easy to see that these equations are equivalent to the geodesic equation for $x^\mu(\tau)$ (exercise). The Lagrangian is independent of τ so H is conserved along a geodesic (it is proportional to $g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu$).

The Kerr-Newman metric is independent of t, ϕ and so p_t and p_ϕ are constant along a geodesic. These are the usual constants associated with the Killing vectors k, m. Hence we have 3 conserved quantities along a geodesic in any stationary and axisymmetric spacetime. What is special about Kerr-Newman is the existence of fourth conserved quantity, the *Carter constant*:

$$Q(x,p) \equiv p_\theta^2 + \frac{1}{\sin^2\theta}\left(p_\phi + a p_t \sin^2\theta\right)^2 - 2Ha^2\cos^2\theta \tag{7.14}$$

If we have a function $f(x,p)$ then its time derivative along a geodesic is given by its Poisson bracket with H:

$$\frac{df}{d\tau} = \frac{\partial f}{\partial x^\mu}\dot{x}^\mu + \frac{\partial f}{\partial p_\mu}\dot{p}_\mu = \frac{\partial f}{\partial x^\mu}\frac{\partial H}{\partial p_\mu} - \frac{\partial f}{\partial p_\mu}\frac{\partial H}{\partial x^\mu} \equiv \{f, H\} \tag{7.15}$$

On examples sheet 3 you will prove $\{Q, H\} = 0$ hence Q is constant along any geodesic! Carter discovered this conserved quantity via Hamilton-Jacobi theory, which we'll describe below.

Notice that Q is quadratic in p_μ. Hence, using (7.10) we can write Q as an expression quadratic in $\dot{x}^\mu$: $Q = K_{\mu\nu}\dot{x}^\mu\dot{x}^\nu$ where $K_{\mu\nu}$ is a symmetric tensor. Our conservation law is

$$0 = \frac{dQ}{d\tau} = \dot{x}^\rho Q_{;\rho} = K_{\mu\nu;\rho}\dot{x}^\mu \dot{x}^\nu \dot{x}^\rho \tag{7.16}$$

where we used the geodesic equation in the final step. Now, at any point p, consider a geodesic with tangent vector V^μ at p. From the above we have $K_{\mu\nu;\rho}V^\mu V^\nu V^\rho = 0$ at p. This holds for any V^μ, which implies (e.g. differentiate 3 times w.r.t. V^μ)

$$K_{(\mu\nu;\rho)} = 0 \tag{7.17}$$

A tensor satisfying this equation is called a *Killing tensor*. Examples of Killing tensors are obtained by taking outer products of Killing vectors and/or the metric tensor. The Kerr-Newman spacetime admits a non-trivial Killing tensor, i.e., one that is not obtained this way. The existence of the Carter constant arises from this Killing tensor.

To write out the geodesic equations we'll need the inverse metric, which you will determine on examples sheet 3:

$$g^{-1} = \frac{1}{\Sigma}\left\{ -\frac{1}{\Delta}\left((r^2+a^2)\frac{\partial}{\partial t} + a\frac{\partial}{\partial \phi}\right)^2 + \Delta\left(\frac{\partial}{\partial r}\right)^2 + \frac{1}{\sin^2\theta}\left(\frac{\partial}{\partial \phi} + a\sin^2\theta\frac{\partial}{\partial t}\right)^2 + \left(\frac{\partial}{\partial \theta}\right)^2 \right\} \tag{7.18}$$

We now use $\dot{x}^\mu = g^{\mu\nu}p_\nu$ and write $p_t = -E$, $p_\phi = h$ to obtain

$$\Sigma\frac{dt}{d\tau} = \left[\frac{(r^2+a^2)^2}{\Delta} - a\sin^2\theta\right]E - \frac{2Mrah}{\Delta} \tag{7.19}$$

$$\Sigma\frac{d\phi}{d\tau} = \frac{2MraE}{\Delta} + \left(\frac{1}{\sin^2\theta} - \frac{a^2}{\Delta}\right)h \tag{7.20}$$

We'll write the conserved Hamiltonian as $H = -(1/2)\sigma$ where σ is constant and conventionally the affine parameter is normalized so that $\sigma \in \{1, 0, -1\}$ for timelike/null/spacelike geodesics. Writing out this conservation law and using (7.14) to eliminate p_θ^2 gives

$$\frac{1}{\Delta}\left(\Sigma\frac{dr}{d\tau}\right)^2 - \frac{1}{\Delta}\left((r^2+a^2)E - ah\right)^2 + \sigma r^2 = -Q \tag{7.21}$$

where we substituted $p_r = g_{rr}\dot{r}$ from (7.10). Similarly substituting $p_\theta = g_{\theta\theta}\dot{\theta}$ in (7.14) gives

$$\left(\Sigma\frac{d\theta}{d\tau}\right)^2 + \frac{1}{\sin^2\theta}\left(h - aE\sin^2\theta\right)^2 + \sigma a^2\cos^2\theta = Q \tag{7.22}$$

This equation suggests a physical interpretation for Q. Recall that, in Newtonian theory, motion in a spherically symmetric potential admits a conserved angular momentum $\mathbf{L}$. The equations of motion can be written in terms of L_z and $\mathbf{L}^2$. We already know that for a massive particle h is analogous to L_z/μ (where μ is rest mass). The Schwarzschild and Reissner-Nordstrom spacetimes are spherically symmetric, so they should admit a conserved quantity analogous to $\mathbf{L}^2$. In these spacetimes, by comparing the above equation with the corresponding Newtonian equation for θ, one sees that Q corresponds to $\mathbf{L}^2/\mu^2$. In the Kerr-Newman spacetime we therefore also interpret Q as analogous to "squared total angular momentum per unit rest mass", even though there is no isometry that gives rise to this conservation law.

Equations (7.19) to (7.22) are a set of first order ODEs that determine Kerr-Newman geodesics. They are straightforward to solve numerically on a computer.

Note that the factors of Σ (which depends on both r and θ) imply that (7.21) and (7.22) are coupled together. There are two common approaches to decoupling them. One is to switch to using r as a parameter along the geodesic. We do this by writing these equations in the form

$$\left(\Sigma\frac{dr}{d\tau}\right)^2 = R(r) \qquad \left(\Sigma\frac{d\theta}{d\tau}\right)^2 = \Theta(\theta) \tag{7.23}$$

Taking square roots and dividing the equations then gives

$$\frac{dr}{\sqrt{R(r)}} = \pm\frac{d\theta}{\sqrt{\Theta(\theta)}} \tag{7.24}$$

which can be integrated to determine θ in terms of r along the geodesic. The integrals are rather complicated. They can be written in terms of elliptic functions but often it is more useful just to evaluate them numerically. One must also be careful near turning points in r. Using r as a parameter also enables (7.19) and (7.20) to be integrated.

An alternative approach is to introduce a non-affine parameter $\tilde{\tau}$ defined by $d\tilde{\tau}/d\tau = \Sigma^{-1}$ along the geodesic. This gives $dx^\mu/d\tilde{\tau} = \Sigma dx^\mu/d\tau$. So now our (r,θ) equations are decoupled: $(dr/d\tilde{\tau})^2 = R(r)$ and $(d\theta/d\tilde{\tau})^2 = \Theta(\theta)$. With suitable initial conditions, these can be integrated, at least numerically, to determine $r(\tilde{\tau})$ and $\theta(\tilde{\tau})$. Plugging $r(\tilde{\tau})$ into the RHS of (7.19) and (7.20) one can then integrate to determine $t(\tilde{\tau})$ and $\phi(\tilde{\tau})$. Finally we can solve $d\tau/d\tilde{\tau} = \Sigma$ to determine $\tau(\tilde{\tau})$.

Carter discovered his constant using Hamilton-Jacobi theory, which we'll review briefly. In a general spacetime, given any point q we can find a convex normal neighbourhood of q, within which there exists a unique geodesic connecting any pair of points. Given such a neighbourhood, introduce a coordinate chart x^μ and define a function $S(\tau, x^\mu; q)$ as

$$S(\tau, x; q) = \int_0^\tau L(x(\tau'), \dot{x}(\tau'))d\tau' \tag{7.25}$$

where the integral is evaluated along the unique geodesic $x^\mu(\tau')$ with $x^\mu(0) = x^\mu_q$ (the coordinates of q) and $x^\mu(\tau) = x^\mu$ (NB $x^\mu(\tau')$ denotes the geodesic and x^μ its endpoint). Now consider how S changes if we vary x^μ with τ held fixed. This implies that the geodesic $x^\mu(\tau)$ must vary. But we know that varying the integral on the RHS just gives the Euler-Lagrange equations plus a surface term. Since we're considering a geodesic, the EL equations vanish, leaving just the surface term:

$$\delta S = [g_{\mu\nu}(x(\tau'))\delta x^\mu(\tau')\dot{x}^\nu(\tau')]_0^\tau = g_{\mu\nu}(x)\delta x^\mu \dot{x}^\nu(\tau) \tag{7.26}$$

using $\delta x^\mu(0) = 0$ since p is fixed and $\delta x^\mu(\tau) = \delta x^\mu$. Hence we have

$$\frac{\partial S}{\partial x^\mu} = g_{\mu\nu}(x)\dot{x}^\nu(\tau) \qquad \Rightarrow \qquad \dot{x}^\mu(\tau) = g^{\mu\nu}S_{,\nu} \tag{7.27}$$

from which we can read off $\dot{x}^\mu(\tau)$ if we know S. We can also vary τ with x^μ held fixed. We do this simply by rescaling the affine parameter along the geodesic: $x^\mu_{\rm new}(\tau') = x^\mu(\tau'/(1+\delta\tau/\tau))$ satisfies the geodesic equation and the boundary conditions $x^\mu_{\rm new}(0) = x^\mu_q$ and $x^\mu_{\rm new}(\tau+\delta\tau) = x^\mu$. This gives $\delta x^\mu(\tau') = -(\tau'\delta\tau/\tau)\dot{x}^\mu(\tau')$. Hence when we vary τ in (7.25) we get two terms, one from varying the upper limit of integration and the other from varying the integrand. As before, the latter gives a term proportional to the EL equations, and so vanishes, plus a surface term of the form (7.26). Hence

$$\delta S = L(x(\tau),\dot{x}(\tau))\delta\tau + [g_{\mu\nu}(x(\tau'))(-\tau'\delta\tau/\tau)\dot{x}^\mu(\tau')\dot{x}^\nu(\tau')]_0^\tau = -\frac{1}{2}g_{\mu\nu}(x(\tau))\dot{x}^\mu(\tau)\dot{x}^\nu(\tau)\delta\tau$$

Eliminating $\dot{x}^\mu(\tau)$ with (7.27) gives the *Hamilton-Jacobi equation*:

$$\frac{\partial S}{\partial \tau} = -\frac{1}{2}g^{\mu\nu}S_{,\mu}S_{,\nu} \tag{7.28}$$

Note that S depends on the choice of q. A different choice of q will give a different solution of this equation.

Now we reverse the argument. Let $S(\tau, x)$ be a solution of the Hamilton-Jacobi equation in some region. Given a point q we define a curve $x^\mu(\tau)$ through q by solving the ODE

$$\dot{x}^\mu(\tau) = g^{\mu\nu}(x(\tau))S_{,\nu}(\tau, x(\tau)) \tag{7.29}$$

with the initial condition $x^\mu(0) = x^\mu_q$. We claim that this curve is a geodesic. To see this, first note that (7.10) gives the conjugate momentum along the curve as

$$p_\mu(\tau) = S_{,\mu}(\tau, x(\tau)) \tag{7.30}$$

and differentiating w.r.t. τ gives

$$\dot{p}_\mu = \frac{d}{d\tau}\left[S_{,\mu}(\tau, x(\tau))\right] = S_{,\tau\mu} + S_{,\mu\nu}\dot{x}^\nu = S_{,\tau\mu} + g^{\nu\rho}S_{,\mu\nu}p_\rho \tag{7.31}$$

but we can differentiate the HJ equation to obtain

$$S_{,\tau\mu} = -\frac{1}{2}g^{\nu\rho}{}_{,\mu}S_{,\nu}S_{,\rho} - g^{\nu\rho}S_{,\nu\mu}S_{,\rho} = -\frac{1}{2}g^{\nu\rho}{}_{,\mu}\,p_\nu p_\rho - g^{\nu\rho}S_{,\nu\mu}p_\rho \tag{7.32}$$

so $\dot{p}_\mu = -\frac{1}{2}g^{\nu\rho}{}_{,\mu}\,p_\nu p_\rho$ and we've shown that our curve satisfies equations (7.10) and (7.13) so it is a geodesic, as claimed.

Now consider the HJ equation in the Kerr-Newman spacetime. Using (7.18) this is

$$\begin{aligned} 0 = 2\frac{\partial S}{\partial\tau} &- \frac{1}{\Sigma\Delta}\left((r^2+a^2)\frac{\partial S}{\partial t} + a\frac{\partial S}{\partial\phi}\right)^2 + \frac{1}{\Sigma\sin^2\theta}\left(\frac{\partial S}{\partial\phi} + a\sin^2\theta\frac{\partial S}{\partial t}\right)^2 \\ &+ \frac{\Delta}{\Sigma}\left(\frac{\partial S}{\partial r}\right)^2 + \frac{1}{\Sigma}\left(\frac{\partial S}{\partial\theta}\right)^2 \end{aligned} \tag{7.33}$$

Remarkably, this can be solved by separation of variables: make the Ansatz

$$S = \frac{1}{2}\sigma\tau - Et + L\phi + S_r(r) + S_\theta(\theta) \tag{7.34}$$

where σ, E and L are constants. Note from (7.30) that E, L are the usual conserved quantities along the geodesic associated with the Killing vectors, and the HJ equation gives $g^{\mu\nu}p_\mu p_\nu = -\sigma$, so σ is the usual conserved quantity coming from the metric. After multiplying by Σ, the HJ equation separates into

$$\Delta\left(\frac{dS_r}{dr}\right)^2 - \frac{1}{\Delta}\left(-(r^2+a^2)E + aL\right)^2 + \sigma r^2 = -Q \tag{7.35}$$

$$\left(\frac{dS_\theta}{d\theta}\right)^2 + \frac{1}{\sin^2\theta}\left(L - aE\sin^2\theta\right)^2 + \sigma a^2\cos^2\theta = Q \tag{7.36}$$

where the separation constant Q is the Carter constant. Using $p_r = dS_r/dr$ and $p_\theta = dS_\theta/d\theta$ these are equations (7.21) and (7.22).

7.6 The ergosphere and Penrose process

In BL coordinates, consider the norm of the Killing vector field k^a:

$$k^2 = g_{tt} = -\frac{\left(\Delta - a^2\sin^2\theta\right)}{\Sigma} = -\left(1 - \frac{2Mr}{r^2 + a^2\cos^2\theta}\right) \tag{7.37}$$

Hence k^a is timelike in region I if and only if $r^2 - 2Mr + a^2\cos^2\theta > 0$ i.e. if, and only if $r > M + \sqrt{M^2 - a^2\cos^2\theta}$. Hence k^a is *spacelike* in the following region outside $\mathcal{H}^+$

$$r_+ = M + \sqrt{M^2 - a^2} < r < M + \sqrt{M^2 - a^2\cos^2\theta} \tag{7.38}$$

This region is called the *ergosphere*. Its surface is called the *ergosurface*. The latter intersects $\mathcal{H}^+$ at the poles $\theta = 0, \pi$, as shown in Fig. 44.

A *stationary observer* is someone with 4-velocity parallel to k^a. Such observers do not exist in the ergosphere because k^a is spacelike there. Any causal curve in the ergosphere must rotate (relative to observers at infinity) in the same direction as the black hole.

Consider a particle with 4-momentum $P^a = \mu u^a$ (where μ is rest mass and u^a is 4-velocity). Let the particle approach a Kerr black hole along a geodesic. The energy of the particle according to a stationary observer at infinity is the conserved quantity along the geodesic

$$E = -k \cdot P \tag{7.39}$$

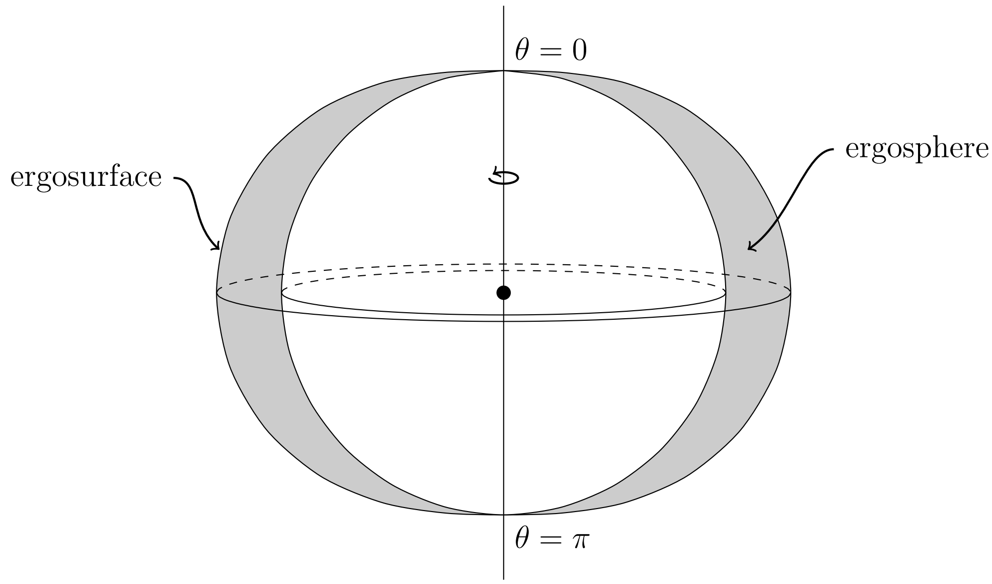


Figure 44. Ergosphere of a Kerr black hole.

Suppose that the particle decays at a point p inside the ergosphere into two other particles with 4-momenta P_1^a and P_2^a (Fig. 45). From the equivalence principle, we know that the decay must conserve 4-momentum (because we can use special relativity in a local inertial frame at p) hence

$$P^a = P_1^a + P_2^a \qquad \Rightarrow \qquad E = E_1 + E_2 \tag{7.40}$$

where $E_i = -k \cdot P_i$. Since k^a is spacelike within the ergoregion, it is possible that $E_1 < 0$. We must then have $E_2 = E + |E_1| > E$. It can be shown that the first particle must fall into the black hole and the second one can escape to infinity. This particle emerges from the ergoregion with greater energy than the particle that was sent in! Energy is conserved because the particle that falls into the black hole carries in negative energy, so the energy (mass) of the black hole decreases. This *Penrose process* is a method for extracting energy from a rotating black hole.

How much energy can be extracted in this process? A particle crossing $\mathcal{H}^+$ must have $-P \cdot \xi \geq 0$ because both P^a and ξ^a are future-directed causal vectors. But $\xi^a = k^a + \Omega_H m^a$ hence

$$E - \Omega_H L \geq 0 \tag{7.41}$$

where E is the energy of the particle and

$$L = m \cdot P \tag{7.42}$$

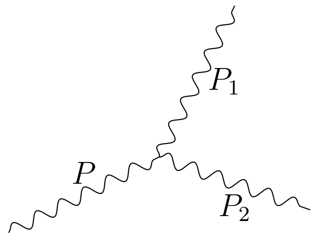


Figure 45. Decay of a particle inside the ergosphere.

is its conserved angular momentum. Hence we have $L \leq E/\Omega_H$ (recall our convention $a > 0$ so $\Omega_H > 0$). The particle carries energy E and angular momentum L into the black hole. If the black hole now settles down to a Kerr solution then this new Kerr solution will have slightly different mass and angular momentum: $\delta M = E$ and $\delta J = L$. Therefore

$$\delta J \leq \frac{\delta M}{\Omega_H} = \frac{2M(M^2 + \sqrt{M^4 - J^2})}{J} \delta M \tag{7.43}$$

Exercise. Show that this is equivalent to $\delta M_{\rm irr} \geq 0$ where the *irreducible mass* is

$$M_{\rm irr} = \left[\frac{1}{2} \left(M^2 + \sqrt{M^4 - J^2} \right) \right]^{1/2} \tag{7.44}$$

Inverting this expression gives

$$M^2 = M_{\rm irr}^2 + \frac{J^2}{4M_{\rm irr}^2} \geq M_{\rm irr}^2 \tag{7.45}$$

Hence in the Penrose process it is not possible to reduce the mass of the black hole below the initial value of $M_{\rm irr}$: there is a limit to the amount of energy that can be extracted.

Exercise. Show that $A = 16\pi M_{\rm irr}^2$ is the "area of the event horizon" of a Kerr black hole, i.e., the area of the intersection of $\mathcal{H}^+$ with a partial Cauchy surface (e.g. a surface $v = \text{const}$ in Kerr coordinates).

Hence $\delta A \geq 0$ in the Penrose process: the area of the event horizon is non-decreasing. This is a special case of the second law of black hole mechanics. The explicit expression for A is

$$A = 8\pi \left(M^2 + \sqrt{M^4 - J^2} \right) \tag{7.46}$$

8 Mass, charge and angular momentum

8.1 Charges in curved spacetime

On an orientable n-dimensional manifold with a metric, we denote the volume form by $\epsilon_{a_1 \dots a_n}$. This can be shown to obey

$$\epsilon^{a_1 \dots a_p c_{p+1} \dots c_n} \epsilon_{b_1 \dots b_p c_{p+1} \dots c_n} = \pm p!(n-p)! \delta^{a_1}_{[b_1} \dots \delta^{a_p}_{b_p]} \tag{8.1}$$

where the upper (lower) sign holds for Riemannian (Lorentzian) signature.

Definition. The *Hodge dual* of a p-form X is the $(n-p)$-form $\star X$ defined by

$$(\star X)_{a_1 \dots a_{n-p}} = \frac{1}{p!} \epsilon_{a_1 \dots a_{n-p} b_1 \dots b_p} X^{b_1 \dots b_p} \tag{8.2}$$

Lemma. For a p-form X

$$\star (\star X) = \pm (-1)^{p(n-p)} X \tag{8.3}$$

$$(\star d \star X)_{a_1 \dots a_{p-1}} = \pm (-1)^{p(n-p)} \nabla^b X_{a_1 \dots a_{p-1} b} \tag{8.4}$$

where the upper (lower) sign holds for Riemannian (Lorentzian) signature.

Proof. Use (8.1).

For example, in 3d Euclidean space, the usual operations of vector calculus can be written using differential forms as

$$\nabla f = df \qquad \text{div } X = \star d \star X \qquad \text{curl } X = \star dX \tag{8.5}$$

where f is a function and X denotes the 1-form X_a obtained from a vector field X^a. The final equation shows that the exterior derivative can be thought of as a generalization of the curl operator.

Another example is Maxwell's equations

$$\nabla^a F_{ab} = -4\pi j_b \qquad \nabla_{[a} F_{bc]} = 0 \tag{8.6}$$

where j^a is the current density vector. These can be written as

$$d \star F = -4\pi \star j, \qquad dF = 0 \tag{8.7}$$

The first of these implies $d \star j = 0$, which is equivalent to $\nabla_a j^a = 0$, i.e., j^a is a conserved current. The second of these implies (by the Poincaré lemma) that *locally* there exists a 1-form A such that $F = dA$.

Now consider a spacelike hypersurface Σ. We define the total electric charge on Σ to be

$$Q = -\int_\Sigma \star j \tag{8.8}$$

(The orientation of Σ is fixed by regarding Σ as a boundary of $J^-(\Sigma)$ and choosing the orientation used in Stokes' theorem.) Using Maxwell's equations we can write

$$Q = \frac{1}{4\pi}\int_\Sigma d\star F \tag{8.9}$$

Hence if Σ is a manifold with boundary $\partial\Sigma$ then Stokes' theorem gives

$$Q = \frac{1}{4\pi}\int_{\partial\Sigma} \star F \tag{8.10}$$

This expresses the total charge on Σ in terms of an integral of $\star F$ over $\partial\Sigma$. It is the curved space generalisation of Gauss' law $Q \sim \int \mathbf{E}\cdot d\mathbf{S}$.

For example, consider Minkowski spacetime in spherical polar coordinates, choosing the orientation so that the volume form is $r^2 \sin\theta dt \wedge dr \wedge d\theta \wedge d\phi$. Let Σ be the surface $t = 0$. If we regard this as the boundary of the region $t \le 0$ then Stokes' theorem fixes the orientation of Σ as $dr \wedge d\theta \wedge d\phi$. Now let Σ_R be the region $r \le R$ of Σ, whose boundary is S^2_R: the sphere $t = 0$, $r = R$. Stokes tells us to pick the orientation of S^2_R to be $d\theta \wedge d\phi$. Consider a Coulomb potential

$$A = -\frac{q}{r}dt \qquad \Rightarrow \qquad F = -\frac{q}{r^2}dt \wedge dr \tag{8.11}$$

Taking the Hodge dual gives

$$(\star F)_{\theta\phi} = r^2 \sin\theta F^{tr} = q\sin\theta \tag{8.12}$$

and hence the charge on Σ_R is

$$Q[\Sigma_R] = \frac{1}{4\pi}\int_{S^2_R} \star F = \frac{1}{4\pi}\int d\theta d\phi\, q\sin\theta = q \tag{8.13}$$

so our definition of Q indeed gives the correct result.

For an asymptotically flat hypersurface in Minkowski spacetime we can take the limit $R \to \infty$ to express the total charge on Σ as an integral at infinity. Motivated by this, we now define the total charge for any asymptotically flat end:

Definition. Let (Σ, h_{ab}, K_{ab}) be an asymptotically flat end. Then the electric and magnetic charges associated to this end are

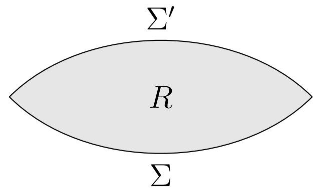


Figure 46. Region R bounded by Σ, Σ'.

$$Q = \frac{1}{4\pi} \lim_{r\to\infty} \int_{S_r^2} \star F \qquad\qquad P = \frac{1}{4\pi} \lim_{r\to\infty} \int_{S_r^2} F \tag{8.14}$$

where S_r^2 is a sphere $x^i x^i = r^2$ where x^i are the coordinates used in the definition of an asymptotically flat end.

Exercise (examples sheet 3). Show that these definitions agree with Q, P used in the Kerr-Newman solution.

Hence the charges can be non-zero even when no charged matter is present in the spacetime (i.e. $j^a = 0$). Consider a surface of constant t in Kerr-Newman (or Reissner-Nordstrom). The total charge on this surface is zero. But when we convert it to a surface integral at infinity, we get two terms because the surface has two asymptotically flat ends. Hence the charges of these two ends must be equal in magnitude with opposite sign.

8.2 Komar integrals

If (M, g) is stationary then there exists a conserved energy-momentum current

$$J_a = -T_{ab} k^b \qquad d \star J = 0 \tag{8.15}$$

Hence one can define the total energy of matter on a spacelike hypersurface Σ as

$$E[\Sigma] = -\int_\Sigma \star J \tag{8.16}$$

This is conserved: if Σ, Σ' bound a spacetime region R (Fig. 46) then

$$E[\Sigma'] - E[\Sigma] = -\int_{\partial R} \star J = -\int_R d \star J = 0 \tag{8.17}$$

Note that we need not require that the energy-momentum tensor T_{ab} used above is the one appearing on the RHS of the Einstein equation. It could be the time-dependent energy momentum tensor of a test field in a stationary vacuum spacetime.

Now if we had $\star J = dX$ for some 2-form X then we could convert $E[\Sigma]$ to an integral over $\partial\Sigma$ as we did in the previous section. We could then define the total energy for a general asymptotically flat end. Unfortunately, this is not possible. However, consider

$$(\star d \star dk)_a = -\nabla^b (dk)_{ab} = -\nabla^b \nabla_a k_b + \nabla^b \nabla_b k_a = 2\nabla^b \nabla_b k_a \tag{8.18}$$

where we using Killing's equation. Now recall

Lemma. A Killing vector field k^a obeys

$$\nabla_a \nabla_b k^c = R^c{}_{bad} k^d \tag{8.19}$$

Hence we have

$$(\star d \star dk)_a = -2R_{ab} k^b = 8\pi J'_a \tag{8.20}$$

where we used Einstein's equation (so henceforth T_{ab} must be the one appearing in Einstein's equation) and

$$J'_a = -2\left(T_{ab} - \frac{1}{2} T g_{ab}\right) k^b \tag{8.21}$$

Therefore

$$d \star dk = 8\pi \star J' \tag{8.22}$$

So $\star J'$ is *exact* (and conserved: $d \star J' = 0$). It follows that

$$-\int_\Sigma \star J' = -\frac{1}{8\pi}\int_\Sigma d \star dk = -\frac{1}{8\pi}\int_{\partial\Sigma} \star dk \tag{8.23}$$

The LHS appears to be a measure of the energy-momentum content of spacetime.

Exercise. Consider a static, spherically symmetric, perfect fluid star. Let Σ be the region $r \leq r_0$ of a surface of constant t where $r_0 > R$. Show that the RHS of (8.23) is the Schwarzschild parameter M. Show that, in the Newtonian limit, ($p \ll \rho$, $|\Phi| \ll 1$, $|\Psi| \ll 1$), the LHS of (8.23) is the total mass of the fluid.

Hence M is the mass of the star in the Newtonian limit. This motivates the following definition:

Definition. Let (Σ, h_{ab}, K_{ab}) be an asymptotically flat end in a stationary spacetime. The *Komar mass* (or Komar energy) is

$$M_{\text{Komar}} = -\frac{1}{8\pi}\lim_{r\to\infty}\int_{S_r^2} \star dk \tag{8.24}$$

with S_r^2 defined as above.

The Komar mass is a measure of the *total* energy of the spacetime. This energy comes both from matter and from the gravitational field. For example, the first part of the above exercise shows that the Komar mass of a Schwarzschild *black hole* is non-zero, even when no matter is present in the spacetime.

The only property of k^a that we used above is the Killing property. In an axisymmetric spacetime we have a Killing vector field m^a that generates rotations about the axis of symmetry. Using this we can define the angular momentum of an axisymmetric spacetime:

Definition. Let (Σ, h_{ab}, K_{ab}) be an asymptotically flat end in an axisymmetric spacetime. The *Komar angular momentum* is

$$J_{\text{Komar}} = \frac{1}{16\pi} \lim_{r\to\infty} \int_{S_r^2} \star dm \tag{8.25}$$

Exercise (examples sheet 3). Show that $M_{\text{Komar}} = M$ and $J_{\text{Komar}} = J$ for the Kerr-Newman solution.

8.3 Hamiltonian formulation of GR

The Komar mass can be defined only in a stationary spacetime. How do we define energy in a non-stationary spacetime? Energy is defined as the value of the Hamiltonian. So we need to consider the Hamiltonian formulation of GR. For simplicity we'll work in vacuum, i.e., no matter fields present. It is also convenient to change our units. Previously we have set $G = 1$. But in this section we will set $16\pi G = 1$ instead.

Recall that in the 3+1 decomposition of spacetime, we consider a spacetime foliated with surfaces of constant t, so that the metric takes the form

$$ds^2 = -N^2 dt^2 + h_{ij}(dx^i + N^i dt)(dx^j + N^j dt) \tag{8.26}$$

where N is the lapse function and N^i the shift vector. If one substitutes this into the Einstein-Hilbert action then the resulting action is, neglecting surface terms,

$$S = \int dt d^3x \mathcal{L} = \int dt d^3x \sqrt{h} N \left({}^{(3)}R + K_{ij}K^{ij} - K^2 \right) \tag{8.27}$$

where ${}^{(3)}R$ is the Ricci scalar of h_{ij}, K_{ij} is the extrinsic curvature of a surface of constant t, with trace K, and ij indices on the RHS are raised with h^{ij}, the inverse of h_{ij}. The extrinsic curvature can be written

$$K_{ij} = \frac{1}{2N} \left(\dot{h}_{ij} - D_i N_j - D_j N_i \right) \tag{8.28}$$

where a dot denotes a t-derivative and D_i is the covariant derivative associated to h_{ij} on a surface of constant t.

The action S is a functional of N, N^i and h_{ij}. Note that it does not depend on time derivatives of N or N^i. Varying N gives the hamiltonian constraint for a surface of constant t. Similarly, varying N^i gives the momentum constraint. Varying h_{ij} gives the evolution equation for h_{ij}. There are no evolution equations for N, N^i: these functions are not dynamical but can be freely specified, which amounts to a choice of coordinates.

To introduce the Hamiltonian formulation of GR, we need to determine the momenta conjugate to N, N^i and h_{ij}. Since the action does not depend on time derivatives of N and N^i, it follows that their conjugate momenta are identically zero. The momentum conjugate to h_{ij} is

$$\pi^{ij} \equiv \frac{\delta S}{\delta \dot{h}_{ij}} = \sqrt{h}\left(K^{ij} - K h^{ij}\right) \tag{8.29}$$

Note that the factor of $\sqrt{h}$ means that π^{ij} is not a tensor, it is an example of a *tensor density*. (A tensor density of weight p transforms under a coordinate transformation in the same way as h^p times a tensor.)

Now we define the Hamiltonian as the Legendre transform of the Lagrangian:

$$H = \int d^3x \left(\pi^{ij}\dot{h}_{ij} - \mathcal{L}\right) \tag{8.30}$$

If we integrate by parts and neglect surface terms, this gives

$$H = \int d^3x \sqrt{h}\left(N\mathcal{H} + N^i \mathcal{H}_i\right) \tag{8.31}$$

where

$$\mathcal{H} = -{}^{(3)}R + h^{-1}\pi^{ij}\pi_{ij} - \frac{1}{2}h^{-1}\pi^2 \tag{8.32}$$

$$\mathcal{H}_i = -2h_{ik}D_j\left(h^{-1/2}\pi^{jk}\right) \tag{8.33}$$

with $\pi \equiv h^{ij}\pi_{ij}$. In the Hamiltonian formalism, h_{ij} and π^{ij} are the dynamical variables. N and N^i play the role of Lagrange multipliers, i.e., we demand $\delta H/\delta N = \delta H/\delta N^i = 0$, which gives $\mathcal{H} = \mathcal{H}_i = 0$. These are simply the Hamiltonian and momentum constraints. The equations of motion are given by Hamilton's equations:

$$\dot{h}_{ij} = \frac{\delta H}{\delta \pi^{ij}} \qquad \dot{\pi}^{ij} = -\frac{\delta H}{\delta h_{ij}} \tag{8.34}$$

The first of these just reproduces the definition of π^{ij}. The second equation is quite lengthy.

Now we've determined the Hamiltonian for GR, we can define the energy of a solution as the value of the Hamiltonian. But (8.31) *vanishes* for any solution of the constraint equations!

The resolution of this puzzle is that we need to add a boundary term to the Hamiltonian. To calculate the variational derivatives in (8.34) we need to integrate by parts in order to remove derivatives from $\delta\pi^{ij}$ and δh_{ij}. This generates surface terms. We need to investigate whether neglecting these terms is legitimate. If the constant t surfaces are *compact* then there won't be any surface terms. So in this case, referred to as a *closed universe*, the Hamiltonian really does evaluate to zero on a solution. This remains true when matter is included. Hence, in GR, the total energy of a closed universe is exactly zero. (This leads to speculation about quantum creation of a closed universe from nothing...)

Now consider the case in which the surfaces constant t are not spatially compact. Let's assume that each of these surfaces is asymptotically flat with 1 end. Hence we can introduce "almost Cartesian" coordinates so that as $r \to \infty$ we have $h_{ij} = \delta_{ij} + \mathcal{O}(1/r)$ and $\pi^{ij} = \mathcal{O}(1/r^2)$. Hence the natural boundary conditions on the variations of h_{ij} and π^{ij} are $\delta h_{ij} = \mathcal{O}(1/r)$ and $\delta\pi^{ij} = \mathcal{O}(1/r^2)$. We also assume our time foliation is chosen so that t, x^i approach "inertial" coordinates in Minkowski spacetime at large r. More precisely, assume that $N = 1 + \mathcal{O}(1/r)$ and $N^i \to 0$ as $r \to \infty$.

Consider the region of our constant t surface contained within a sphere of constant r, with boundary S^2_r. When we vary π^{ij}, the resulting surface term on S^2_r is

$$\int_{S^2_r} dA \left(-2N^i h_{ik} n_j h^{-1/2} \delta\pi^{jk}\right) \tag{8.35}$$

where dA is the area element, and n^j the outward unit normal, of S^2_r. Now $dA = \mathcal{O}(r^2)$ but our boundary conditions imply that the expression in brackets decays faster than $1/r^2$ as $r \to \infty$ hence the whole expression vanishes as $r \to \infty$. So we don't need to worry about the surface term that arises when we vary π^{ij}.

When we vary h_{ij}, surface terms arise in two ways. First, the variation of $h^{-1/2}$ in $\mathcal{H}_i$ is within a derivative so we need to integrate by parts, generating a surface term. This is very similar to the surface term above and vanishes as $r \to \infty$. Second, we have the variation of the term ${}^{(3)}R$ in $\mathcal{H}$. You know the variation of the Ricci scalar because this is what you need to calculate when you derive the Einstein equation from the Einstein-Hilbert action. The only difference is that we are now varying a 3d, rather than a 4d, Ricci scalar:

$$\delta^{(3)}R = -R^{ij}\delta h_{ij} + D^i D^j \delta h_{ij} - D^k D_k \left(h^{ij}\delta h_{ij}\right) \tag{8.36}$$

When we calculate δH, we need to integrate by parts twice to eliminate these derivatives on δh_{ij}. The first integration by parts gives the surface term

$$S_1 = -\int_{S^2_r} dA\, N\left[n^i D^j \delta h_{ij} - n^k D_k (h^{ij}\delta h_{ij})\right] \tag{8.37}$$

and the second integration by parts gives another surface term

$$S_2 = \int_{S^2_r} dA \left(n^j \delta h_{ij} D^i N - h^{ij}\delta h_{ij} n^k D_k N\right) \tag{8.38}$$

Our boundary conditions implies that $S_2 \to 0$ as $r \to \infty$. On the other hand, we have

$$\lim_{r\to\infty} S_1 = -\lim_{r\to\infty}\int_{S^2_r} dA\, n_i \left(\partial_j \delta h_{ij} - \partial_i \delta h_{jj}\right) \tag{8.39}$$

Here we have used the fact that $h_{ij} \to \delta_{ij}$ so (a) $D_k \to \partial_k$ as $r \to \infty$ and (b) we don't need to distinguish between "upstairs" and "downstairs" indices. But we can rewrite this as

$$\lim_{r\to\infty} S_1 = -\delta E_{ADM} \tag{8.40}$$

where

$$E_{ADM} = \lim_{r\to\infty}\int_{S^2_r} dA\, n_i \left(\partial_j h_{ij} - \partial_i h_{jj}\right) \tag{8.41}$$

In general, δE_{ADM} will be non-zero. But now consider

$$H' = H + E_{ADM} \tag{8.42}$$

Since H' and H differ by a surface term, they will give the same equations of motion. However, when we vary h_{ij} in H', the boundary term S_1 coming from the variation of H will be cancelled by the variation of the surface term E_{ADM}. Hence no surface terms arise in the variation of H' so H' must be the Hamiltonian for General Relativity with asymptotically flat initial data. The need for this surface term was first pointed out by Regge and Teitelboim (1974).

8.4 ADM energy

Now that we have a satisfactory variational principle, we can evaluate the Hamiltonian on a solution. As before, we have that $H = 0$ so the value of H' is the value of the surface term E_{ADM}. Hence E_{ADM} must be the energy of our initial data set. This is the *Arnowitt-Deser-Misner energy* (1962). We now return to $G = 1$ units to obtain the following

Definition. The *ADM energy* of an asymptotically flat end is

$$E_{ADM} = \frac{1}{16\pi} \lim_{r\to\infty} \int_{S_r^2} dA \, n_i \left(\partial_j h_{ij} - \partial_i h_{jj}\right) \tag{8.43}$$

If we have asymptotically flat initial data with several asymptotically flat ends then one can define a separate ADM energy for each asymptotic end. In a stationary, asymptotically flat spacetime, it can be shown that $E_{ADM} = M_{\rm Komar}$ if one chooses the surfaces of constant t to be orthogonal to the timelike Killing vector field as $r \to \infty$.

Exercise (examples sheet 3). Show that $E_{ADM} = M$ for a constant t surface in the Kerr-Newman solution.

There is also a notion of the total *3-momentum* of an asymptotically flat end:

Definition. The *ADM 3-momentum* of an asymptotically flat end is

$$P_i = \frac{1}{8\pi} \lim_{r\to\infty} \int_{S_r^2} dA \left(K_{ij} n_j - K n_i\right) \tag{8.44}$$

In Newtonian gravity, the energy density of the gravitational field is *negative*. So one might wonder whether the ADM energy in GR could also be negative. Since $E_{ADM} = M$ for a surface of constant t in the Schwarzschild spacetime, it follows that $E_{ADM} < 0$ for $M < 0$ Schwarzschild. But in this case, the surface of constant t is singular (not geodesically complete). We could also arrange that $E_{ADM} < 0$ if we included matter with negative energy density. But if we exclude these unphysical possibilities then we have the *positive energy theorem*:

Theorem (Schoen & Yau 1979, Witten 1981). Let (Σ, h_{ab}, K_{ab}) be an initial data set that is geodesically complete and asymptotically flat. Assume that the energy-momentum tensor satisfies the dominant energy condition. Then $E_{ADM} \geq \sqrt{P_i P_i}$, with equality if, and only if, (Σ, h_{ab}, K_{ab}) arises from a surface in Minkowski spacetime.

In the case of a spacetime containing black holes, one might not want to assume anything about the black hole interior. In this case, one can allow Σ to have an inner boundary corresponding to an apparent horizon and the result still holds (Gibbons, Hawking, Horowitz & Perry 1983).

There is a natural way of regarding (E_{ADM}, P_i) as a 4-vector defined at spatial infinity i^0. We then define the *ADM mass* by

$$M_{ADM} = \sqrt{E_{ADM}^2 - P_i P_i} \geq 0 \tag{8.45}$$

9 Black hole mechanics

9.1 Killling horizons and surface gravity

Definition. A null hypersurface $\mathcal{N}$ is a *Killing horizon* if there exists a Killing vector field ξ^a defined in a neighbourhood of $\mathcal{N}$ such that ξ^a is normal to $\mathcal{N}$.

Theorem (Hawking 1972). In a stationary, analytic, asymptotically flat vacuum black hole spacetime, $\mathcal{H}^+$ is a Killing horizon.

Proof. See Hawking and Ellis.

The result extends to Einstein-Maxwell theory or theories where the matter fields obey hyperbolic equations. As mentioned previously, it would be desirable to eliminate the assumption of analyticity because analyticity implies that the full spacetime is determined by its behaviour in a neighbourhood of a single point.

Note that $\mathcal{H}^+$ is not necessarily a Killing horizon of the stationary Killing vector field k^a. For example, in the Kerr solution, we have $\xi^a = k^a + \Omega_H m^a$ where m^a is the Killing field corresponding to axisymmetry. One can show (see Hawking and Ellis) that this behaviour is general: if ξ^a is not tangent to k^a then one can construct a linear combination m^a of ξ^a and k^a so that the spacetime is stationary and axisymmetric.

If $\mathcal{N}$ is a Killing horizon w.r.t. a Killing vector field ξ^a then it is also a Killing horizon w.r.t. the Killing vector field $c\xi^a$ where c is any non-zero constant. In a stationary, asymptotically flat spacetime, it is conventional to normalise the generator of time translations so that $k^a k_a \to -1$ at infinity. We then normalize ξ^a so that so that $\xi^a = k^a + \Omega_H m^a$.

Since $\xi^a \xi_a = 0$ on $\mathcal{N}$, it follows that the gradient of $\xi^a \xi_a$ is normal to $\mathcal{N}$, i.e., proportional to ξ_a. Hence there exists a function κ on $\mathcal{N}$ such that

$$\nabla_a(\xi^b \xi_b)|_{\mathcal{N}} = -2\kappa \xi_a \tag{9.1}$$

The function κ is called the *surface gravity* of the Killing horizon. The LHS can be rearranged to give $2\xi^b \nabla_a \xi_b = -2\xi^b \nabla_b \xi_a$ using Killing's equation. Hence we have

$$\xi^b \nabla_b \xi^a|_{\mathcal{N}} = \kappa \xi^a \tag{9.2}$$

which shows that κ measures the failure of integral curves of ξ^a to be affinely parameterized. If we let n^a be the tangent to the affinely parameterized generators of $\mathcal{N}$ then we have $\xi^a = f n^a$ for some function f on $\mathcal{N}$. Then using $n \cdot \nabla n^a = 0$ we have, on $\mathcal{N}$, $\xi^b \nabla_b \xi^a = f n^a n^b \partial_b f = f^{-1} \xi^a \xi^b \partial_b f$ and hence

$$\kappa = \xi^a \partial_a \log|f| \tag{9.3}$$

Example. The Reissner-Nordstrom solution in ingoing EF coordinates is

$$ds^2 = -\frac{\Delta}{r^2}dv^2 + 2dvdr + r^2 d\Omega^2 \tag{9.4}$$

where $\Delta = (r - r_+)(r - r_-)$ and $r_\pm = M \pm \sqrt{M^2 - e^2}$. The stationary Killing vector field is $k = \partial/\partial v$. At $r = r_\pm$ we have $\Delta = 0$ so $k_a = (dr)_a$, which is normal to the null hypersurfaces $r = r_\pm$. Hence these surfaces are Killing horizons. To calculate the surface gravity we use

$$d(k^b k_b) = d(-\Delta/r^2) = (-\Delta'/r^2 + 2\Delta/r^3)dr \tag{9.5}$$

Evaluating at $r = r_\pm$ gives

$$d(k^b k_b)|_{r=r_\pm} = -\frac{(r_\pm - r_\mp)}{r_\pm^2}dr = -\frac{(r_\pm - r_\mp)}{r_\pm^2}k|_{r=r_\pm} \tag{9.6}$$

hence the surface gravities are

$$\kappa = \kappa_\pm = \frac{(r_\pm - r_\mp)}{2r_\pm^2} \tag{9.7}$$

For Schwarzschild we have $e = 0$ so $r_+ = 2M$, $r_- = 0$ and hence $\kappa = 1/4M$ is the surface gravity of $\mathcal{H}^+$. For extreme RN we have $r_+ = r_-$ and $\kappa = 0$.

Exercise. In the Kruskal spacetime, $\mathcal{H}^+$ is the surface $U = 0$ and $\mathcal{H}^-$ the surface $V = 0$. Use (2.34) to show that these are Killing horizons of k^a (the time translation Killing vector field). Calculate the LHS of (9.1). Use (2.31) to relate dr to $d(UV)$. Hence show that the surface gravity of $\mathcal{H}^\pm$ is $\pm 1/(4M)$.

This is an example of a *bifurcate Killing horizon* i.e. a pair of intersecting null hypersurfaces $\mathcal{N}^\pm$ that are each Killing horizons with respect to the same Killing vector field. At the *bifurcation surface* $B = \mathcal{N}^+ \cap \mathcal{N}^-$, the Killing field can't be normal to both $\mathcal{N}^+$ and $\mathcal{N}^-$ so it must vanish on B. Any vector X^a tangent to B is tangent to both $\mathcal{N}^+$ and $\mathcal{N}^-$, which implies that X^a must be spacelike so B is a spacelike surface. For the Kruskal spacetime this is the 2-sphere $\{U = V = 0\}$.

9.2 Interpretation of surface gravity

The main reason that κ is important is because $\hbar\kappa/(2\pi)$ is the Hawking temperature of the hole (see later). There is also a classical interpretation of κ.

In a static, asymptotically flat spacetime, consider a particle of unit mass that is "at rest", i.e., following an orbit of k^a. Such orbits are not geodesics so the particle is accelerating. This acceleration requires a force, let's assume it is provided by a massless inelastic string attached to the particle, with the other end of the string held by an observer at infinity. Let F be the force in the string (i.e. the tension) measured at infinity. Then $F \to \kappa$ as we consider orbits closer and closer to a Killing horizon of k^a (for the Schwarzschild solution this is proved on examples sheet 3). Hence κ is the force per unit mass required *at infinity* to hold a test particle at rest near the horizon.

The *local* force on the particle is certainly not κ. In a general stationary spacetime, the 4-velocity of a particle on an orbit of k^a is

$$u^a = \frac{k^a}{\sqrt{-k^2}} \tag{9.8}$$

where the normalisation is fixed by the condition $u^2 = -1$. The proper acceleration of the particle is therefore

$$A^a = u \cdot \nabla u^a = \frac{k \cdot \nabla k^a}{-k^2} + \frac{k^a}{2(-k^2)^2} k \cdot \nabla(k^2) \tag{9.9}$$

In the first term, Killing's equation gives $k^b \nabla_b k_a = -k^b \nabla_a k_b = -(1/2)\partial_a(k^2)$. In the second term $k \cdot \nabla(k^2) = 2k^a k^b \nabla_a k_b = 0$. Hence we have

$$A_a = \frac{\partial_a(-k^2)}{2(-k^2)} = \frac{1}{2}\partial_a \log(-k^2) \tag{9.10}$$

Since $k^2 \to 0$ at a Killing horizon, it follows that A_a must diverge at the horizon. For Schwarzschild we have (viewing A_a as a 1-form)

$$A = \frac{1}{2} d \log\left(1 - \frac{2M}{r}\right) = \frac{M}{r^2\,(1 - 2M/r)} dr \tag{9.11}$$

and so the norm of A is (using $g^{rr} = (1 - 2M/r)$)

$$|A| \equiv \sqrt{g^{ab} A_a A_b} = \sqrt{\frac{M^2}{r^4(1 - 2M/r)}} = \frac{M}{r^2\sqrt{1 - 2M/r}} \tag{9.12}$$

which diverges as $r \to 2M$. Hence the *local* tension (i.e. the force exerted on the particle by the string) is very large if the particle is near the horizon. A physical string would break if the particle were too near the horizon.

9.3 Zeroth law of black holes mechanics

Proposition. Consider a null geodesic congruence that contains the generators of a Killing horizon $\mathcal{N}$. Then $\theta = \hat{\sigma} = \hat{\omega} = 0$ on $\mathcal{N}$.

Proof. $\hat{\omega} = 0$ on $\mathcal{N}$ because the generators are hypersurface orthogonal.

Let ξ^a be a Killing field normal to $\mathcal{N}$. On $\mathcal{N}$ we can write $\xi^a = hU^a$ where U^a is tangent to the (affinely parameterized) generators of $\mathcal{N}$ and h is a function on $\mathcal{N}$. Let $\mathcal{N}$ be specified by an equation $f = 0$. Then we can write $U^a = h^{-1}\xi^a + fV^a$ where V^a is a smooth vector field. We can then calculate

$$B_{ab} = \nabla_b U_a = (\partial_b h^{-1})\xi_a + h^{-1}\nabla_b \xi_a + (\partial_b f)V_a + f\nabla_b V_a \tag{9.13}$$

so evaluating on $\mathcal{N}$ and using Killing's equation gives

$$B_{(ab)}|_{\mathcal{N}} = \left(\xi_{(a}\partial_{b)}h^{-1} + V_{(a}\partial_{b)}f\right)_{\mathcal{N}} \tag{9.14}$$

But both ξ_a and $\partial_a f$ are parallel to U_a on $\mathcal{N}$. Hence when we project onto $T_\perp$, both terms are eliminated:

$$\hat{B}_{(ab)}|_{\mathcal{N}} = P_a^c B_{(cd)} P_b^d = 0 \tag{9.15}$$

Hence θ and $\hat{\sigma}$ vanish on $\mathcal{N}$.

Theorem (zeroth law of black hole mechanics). κ is constant on the future event horizon of a stationary black hole spacetime obeying the dominant energy condition.

Proof. Note that Hawking's theorem implies that $\mathcal{H}^+$ is a Killing horizon w.r.t some Killing vector field ξ^a. From the above result we know that $\theta = 0$ along the generators of $\mathcal{H}^+$ hence $d\theta/d\lambda = 0$ along these generators. We also have $\hat{\sigma} = \hat{\omega} = 0$ so Raychaudhuri's equation gives

$$0 = R_{ab}\xi^a\xi^b|_{\mathcal{H}^+} = 8\pi T_{ab}\xi^a\xi^b|_{\mathcal{H}^+} \tag{9.16}$$

where we used Einstein's equation and $\xi^2|_{\mathcal{H}^+} = 0$ in the second equality. This implies

$$J \cdot \xi|_{\mathcal{H}^+} = 0 \tag{9.17}$$

where $J_a = -T_{ab}\xi^b$. Now ξ^a is a future-directed causal vector field hence (by the dominant energy condition), so is J_a (unless $J_a = 0$). Hence the above equation implies J^a is parallel to ξ^a on $\mathcal{H}^+$. Therefore

$$0 = \xi_{[a}J_{b]}|_{\mathcal{H}^+} = -\xi_{[a}T_{b]c}\xi^c|_{\mathcal{H}^+} = -\frac{1}{8\pi}\xi_{[a}R_{b]c}\xi^c|_{\mathcal{H}^+} \tag{9.18}$$

where we used Einstein's equation in the final equality. On examples sheet 4, it is shown that this implies

$$0 = \frac{1}{8\pi}\xi_{[a}\partial_{b]}\kappa \tag{9.19}$$

Hence $\partial_a\kappa$ is proportional to ξ_a so $t \cdot \partial\kappa = 0$ for any vector field t^a that is tangent to $\mathcal{H}^+$. Hence κ is constant on $\mathcal{H}^+$ (assuming $\mathcal{H}^+$ is connected).

9.4 First law of black hole mechanics

The Kerr solution is specified by two parameters M, a. Consider a small variation of these parameters. This will induce small changes in J and A (the horizon area). Using the formula for A one can check that, to first order (exercise)

$$\frac{\kappa}{8\pi}\delta A = \delta M - \Omega_H \delta J \tag{9.20}$$

We can define a linearized metric perturbation to be the difference of the Kerr metric with parameters $(M + \delta M, a + \delta a)$ and the Kerr metric with parameters (M, a). The above formula tells us how this linearized perturbation of the Kerr solution changes A etc. Remarkably, it turns out that this formula holds for *any* linearized perturbation of the metric of the Kerr solution. Consider a hypersurface Σ which extends from the bifurcation surface B to infinity and, near infinity, is asymptotically orthogonal to the timelike Killing vector field. Σ is actually a manifold with boundary because it includes B. Let h_{ab} be the induced metric and K_{ab} the extrinsic curvature of Σ. Then $(\Sigma \backslash B, h_{ab}, K_{ab})$ is an asymptotically flat end. Now consider a linearized perturbation $h_{ab} \to h_{ab} + \delta h_{ab}$, $K_{ab} \to K_{ab} + \delta K_{ab}$ and assume that this obeys the constraint equations to linear order. Then the perturbed initial data satisfies equation (9.20) where δA is the change in the area of B, δM is the change in the ADM energy and δJ is the change in the ADM angular momentum (we have not defined the latter but for an axisymmetric spacetime it agrees with the Komar angular momentum).

This result was proved by Sudasky and Wald in 1992. (A more restricted version, applying only to stationary axisymmetric perturbations, was obtained by Bardeen, Carter and Hawking in 1973.) The proof can be extended to any stationary black hole solution, not just Kerr. For example, it holds for stationary black holes in theories containing matter fields even when one cannot write down the solution explicitly. The result even holds for more general diffeomorphism-covariant theories of gravity involving higher derivatives of the metric. In the particular case of Einstein-Maxwell theory, there is an additional term $-\Phi_H \delta Q$ on the RHS where Q is the electric charge and Φ_H is the electrostatic potential difference between the event horizon and infinity (examples sheet 4).

In this version of the first law of black hole mechanics, we are comparing two different spacetimes: a stationary black hole and a perturbed stationary black hole. There is a another version of the first law, due to Hartle and Hawking (1972) in which we perturb a black hole by throwing in a small amount of matter and wait for it to settle down to a stationary solution again. In this case, (9.20) relates the change in horizon area to the energy and angular momentum of the matter that crosses the event horizon, rather than to a change in the ADM energy and angular momentum (indeed the latter don't change, they are conserved). We will prove this "physical process" version of the first law. (The other version is sometimes called the "equilibrium state" version when restricted to stationary perturbations.)

We treat the matter as a small perturbation of a Kerr black hole, i.e., the energy momentum tensor is $\mathcal{O}(\epsilon)$. We can define energy and angular momentum 4-vectors for the matter

$$J^a = -T^a{}_b k^b \qquad L^a = T^a{}_b m^b \tag{9.21}$$

If we treat the matter as a test field then these are exactly conserved. However, we want to include the gravitational backreaction of the matter, which induces an $\mathcal{O}(\epsilon)$ change in the metric, which will not be stationary and axisymmetric in general, hence J^a and L^a will not be exactly conserved. However, this is a second order effect so $\nabla_a J^a$ and $\nabla_a L^a$ will be $\mathcal{O}(\epsilon^2)$. We will work to linear order in ϵ so we can assume that J^a and L^a are conserved.

Assume that the matter crosses $\mathcal{H}^+$ to the future of the bifurcation sphere B. Let $\mathcal{N}$ be the portion of $\mathcal{H}^+$ to the future of B:

The energy and angular momentum of the matter that crosses $\mathcal{N}$ are (examples sheet 3)

$$\delta M = -\int_{\mathcal{N}} \star J \qquad \delta J = -\int_{\mathcal{N}} \star L \tag{9.22}$$

(Do not confuse angular momentum J in δJ with the energy momentum current J_a appearing in the first integral!) We can introduce Gaussian null coordinates (r, λ, y^i) on $\mathcal{H}^+$ as described in section 4.6, taking the surface S used there to be B. We choose the affine parameter λ of the generators of $\mathcal{H}^+$ to vanish on B, so $\mathcal{N}$ is the portion $\lambda > 0$ of $\mathcal{H}^+$. In these coordinates, $\mathcal{H}+$ is the surface $r = 0$ and the metric on $\mathcal{H}$ is

$$ds^2|_{\mathcal{H}^+} = 2dr d\lambda + h_{ij}(\lambda, y) dy^i dy^j \tag{9.23}$$

We order (y^1, y^2) so that the volume form on $\mathcal{H}^+$ is

$$\eta = \sqrt{h}\, d\lambda \wedge dr \wedge dy^1 \wedge dy^2 \tag{9.24}$$

using $\sqrt{-g} = \sqrt{h}$. The orientation of $\mathcal{N}$ used in (9.22) is the one used in Stokes' theorem, viewing $\mathcal{N}$ as the boundary of the region $r > 0$. This is $d\lambda \wedge dy^1 \wedge dy^2$. We then have, on $\mathcal{N}$

$$(\star J)_{\lambda 12} = \sqrt{h} J^r = \sqrt{h} J_\lambda = \sqrt{h} U \cdot J \tag{9.25}$$

where $U = \partial/\partial\lambda$ is tangent to the generators of $\mathcal{N}$. Hence

$$\delta M = -\int_{\mathcal{N}} d\lambda d^2 y \sqrt{h}\, U \cdot J \tag{9.26}$$

and similarly

$$\delta J = -\int_{\mathcal{N}} d\lambda d^2 y \sqrt{h}\, U \cdot L \tag{9.27}$$

Since J^a and L^a are $\mathcal{O}(\epsilon)$, the perturbation to the spacetime metric contributes to these integrals only at $\mathcal{O}(\epsilon^2)$ hence we can evaluate the integrals by working in the Kerr spacetime. Hence $\mathcal{N}$ is a Killing horizon of $\xi = k + \Omega_H m$ so on $\mathcal{N}$ we have $\xi = fU$ for some function f and we have equation (9.3)

$$\xi \cdot \partial \log |f| = \kappa \qquad \Rightarrow \qquad U \cdot \partial f = \kappa \qquad \Rightarrow \qquad \frac{\partial f}{\partial \lambda} = \kappa \tag{9.28}$$

hence $f = \kappa\lambda + f_0(y)$. But we know that $\xi = 0$ on B hence $f = 0$ at $\lambda = 0$ so $f_0 = 0$. We have shown that

$$\xi^a = \kappa \lambda U^a \qquad \text{on } \mathcal{N} \tag{9.29}$$

From the definition of J^a we have

$$\begin{aligned} \delta M &= \int_{\mathcal{N}} d\lambda d^2 y \sqrt{h}\, T_{ab} U^a k^b = \int_{\mathcal{N}} d\lambda d^2 y \sqrt{h}\, T_{ab} U^a \left(\xi^b - \Omega_H m^b\right) \\ &= \int_{\mathcal{N}} d\lambda d^2 y \sqrt{h}\, T_{ab} U^a U^b \kappa \lambda - \Omega_H \int_{\mathcal{N}} d\lambda d^2 y \sqrt{h}\, U \cdot L \end{aligned} \tag{9.30}$$

The final integral is $-\delta J$. In the first integral the Einstein equation gives $8\pi T_{ab} U^a U^b = R_{ab} U^a U^b$ (as U^a is null). Here R_{ab} is the $\mathcal{O}(\epsilon)$ Ricci tensor of the perturbed spacetime. Hence we have

$$\delta M - \Omega_H \delta J = \frac{\kappa}{8\pi} \int_{\mathcal{N}} d\lambda d^2 y \sqrt{h}\, \lambda R_{ab} U^a U^b \tag{9.31}$$

Raychaudhuri's equation gives

$$\frac{d\theta}{d\lambda} = -R_{ab} U^a U^b \tag{9.32}$$

where we have used the fact that generators of $\mathcal{N}$ have $\hat{\omega} = 0$ and neglected θ^2, $\hat{\sigma}^2$ because these are $\mathcal{O}(\epsilon^2)$ (since θ and $\hat{\sigma}$ vanish for the unperturbed spacetime). Hence

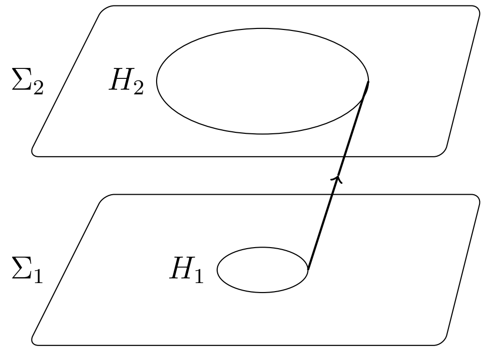


Figure 47. Second law of black hole mechanics, showing a horizon generator.

we have

$$\begin{aligned}\delta M - \Omega_H \delta J &= -\frac{\kappa}{8\pi}\int d^2y \int_0^\infty \sqrt{h}\,\lambda \frac{d\theta}{d\lambda} d\lambda \\ &= -\frac{\kappa}{8\pi}\int d^2y \left\{ \left[\sqrt{h}\,\lambda\theta\right]_0^\infty - \int_0^\infty \left(\sqrt{h} + \lambda \frac{d\sqrt{h}}{d\lambda}\right)\theta d\lambda \right\}\end{aligned} \tag{9.33}$$

Now recall that $d\sqrt{h}/d\lambda = \theta\sqrt{h} = \mathcal{O}(\epsilon)$. This is multiplied by θ in the final integral, giving a negligible $\mathcal{O}(\epsilon^2)$ contribution. If we assume that the black hole settles down to a new stationary solution at late time then $\sqrt{h}$ must approach a finite limit as $\lambda \to \infty$. We have

$$\int_0^\infty \sqrt{h}\,\theta d\lambda = \int_0^\infty \frac{d\sqrt{h}}{d\lambda} d\lambda = \delta\sqrt{h} \tag{9.34}$$

the RHS is finite hence the integral on the LHS must converge so $\theta = o(1/\lambda)$ as $\lambda \to \infty$. This implies that the boundary term on the RHS of (9.33) vanishes, leaving

$$\delta M - \Omega_H \delta J = \frac{\kappa}{8\pi}\int d^2y\, \delta\sqrt{h} = \frac{\kappa}{8\pi}\delta \int d^2y\,\sqrt{h} = \frac{\kappa}{8\pi}\delta A \tag{9.35}$$

9.5 Second law of black hole mechanics

Theorem (Hawking 1972). Let (M, g) be a strongly asymptotically predictable spacetime satisfying the Einstein equation with the null energy condition. Let $U \subset M$ be a globally hyperbolic region for which $\overline{J^-(\mathcal{I}^+)} \subset U$ (such U exists because the spacetime is strongly asymptotically predictable). Let Σ_1, Σ_2 be spacelike Cauchy surfaces for U with $\Sigma_2 \subset J^+(\Sigma_1)$. Let $H_i = \mathcal{H}^+ \cap \Sigma_i$. Then $\text{area}(H_2) \geq \text{area}(H_1)$. (See Fig. 47.)

Proof. We will make the additional assumption that inextendible generators of $\mathcal{H}^+$ are future complete, i.e., $\mathcal{H}^+$ is "non-singular". (This assumption can be eliminated with a bit more work.) First we will show $\theta \geq 0$ on $\mathcal{H}^+$. So assume $\theta < 0$ at $p \in \mathcal{H}^+$. Let γ be the (inextendible) generator of $\mathcal{H}^+$ through p and let q be slightly to the future of p along γ. By continuity we have $\theta < 0$ at q. But then we know from section 4.10 that there exists a point r (to the future of q) conjugate to p on γ (here we use the assumption that γ is future-complete). Theorem 2 of section 4.10 then tells us that we can deform γ to obtain a timelike curve from p to r, violating achronality of $\mathcal{H}^+$. Hence $\theta \geq 0$ on $\mathcal{H}^+$.

Let $p \in H_1$. The generator of $\mathcal{H}^+$ through p cannot leave $\mathcal{H}^+$ (as generators can't have future endpoints) so it must intersect H_2 (as Σ_2 is a Cauchy surface). This defines a map $\phi : H_1 \to H_2$. Now $\text{area}(H_2) \geq \text{area}(\phi(H_1)) \geq \text{area}(H_1)$ where the first inequality follows because $\phi(H_1) \subset H_2$ and the second inequality follows from $\theta \geq 0$.
□

For example, consider the formation of a Schwarzschild black hole in spherically symmetric gravitational collapse. We can draw a Finkelstein diagram:

Now consider two well-separated non-rotating black holes such that the metric near each is well approximated by the Schwarzschild solution. Let the mass parameters be M_1 and M_2. Assume that these black holes collide and merge into a single black hole which settles down to a Schwarzschild black hole of mass M_3. The above theorem implies that the horizon areas obey

$$A_3 \geq A_1 + A_2 \qquad \Rightarrow \qquad 16\pi M_3^2 \geq 16\pi M_1^2 + 16\pi M_2^2 \tag{9.36}$$

hence

$$M_3 \geq \sqrt{M_1^2 + M_2^2} \tag{9.37}$$

The energy radiated as gravitational radiation in this process is $M_1 + M_2 - M_3$. In principle, this energy could be used to do work. The efficiency of this process is limited by the second law because

$$\text{efficiency} = \frac{M_1 + M_2 - M_3}{M_1 + M_2} \leq 1 - \frac{\sqrt{M_1^2 + M_2^2}}{M_1 + M_2} \leq 1 - \frac{1}{\sqrt{2}} \tag{9.38}$$

with the final inequality arising from dividing the numerator and denominator by M_1 and then maximising w.r.t M_2/M_1.

Finally we can discuss the *Penrose inequality*. Consider initial data which is asymptotically flat and contains a trapped surface behind an apparent horizon of area $A_{\rm app}$. Let E_i denote the ADM energy of this data ("i" for initial). If weak cosmic censorship is correct, the spacetime resulting from this data will be a strongly asymptotically predictable black hole spacetime. We would expect this to "settle down" to a stationary black hole at late time. By the uniqueness theorems, this should be described by a Kerr solution with mass M_f and angular momentum J_f ("f" for final). Now since the apparent horizon must lie inside the event horizon we expect $A_{\rm app} \le A_i$ where A_i is the area of the intersection of $\mathcal{H}^+$ with the initial surface Σ. The second law tells us that $A_i \le A_{\rm Kerr}(M_f, J_f)$ (the horizon area of the final Kerr black). But from (7.46) we have

$$A_{\rm Kerr}(M_f, J_f) = 8\pi \left(M_f^2 + \sqrt{M_f^4 - J_f^2} \right) \le 16\pi M_f^2 \tag{9.39}$$

Finally, we have $M_f \le E_i$ because gravitational radiation carries away energy in this process. Putting this together gives

$$A_{\rm app} \le 16\pi E_i^2 \qquad \Rightarrow \qquad E_i \ge \sqrt{\frac{A_{\rm app}}{16\pi}} \tag{9.40}$$

This refers only to quantities that can be calculated from the initial data! If standard beliefs about the gravitational collapse process are correct then this inequality must be satisfied by any initial data set. If one could find initial data that violated this inequality then some aspect of the above argument (e.g. weak cosmic censorship) must be false. No counterexample has been found. Indeed, in the case of *time-symmetric* initial data ($K_{ab} = 0$) with matter obeying the weak energy condition, the above inequality has been proved (Huisken and Ilmanen 1997). Note that the inequality can be regarded as a stronger version of the positive mass theorem.

10 Quantum field theory in curved spacetime

10.1 Introduction

The laws of black hole mechanics have a remarkable similarity to the laws of thermodynamics. At rest, a black hole has energy $E = M$. Consider a thermodynamic system with the same energy and angular momentum as the black hole. This is governed by the first law of thermodynamics

$$dE = TdS + \mu dJ \tag{10.1}$$

where μ is the chemical potential that enforces conservation of angular momentum. This is identical to the first law of black hole mechanics if we make the identifications

$$T = \lambda\kappa \qquad S = A/(8\pi\lambda) \qquad \mu = \Omega_H \tag{10.2}$$

for some constant λ. Furthermore, if we do this then the zeroth law of thermodynamics (the temperature is constant in a body in thermodynamic equilibrium) becomes the zeroth law of black hole mechanics. The second law of thermodynamics (the entropy is non-decreasing in time) becomes the second law of black hole mechanics.

This similarity suggests that black holes might be thermodynamic objects. Another reason for believing this is that if black holes do not have entropy then one could violate the second law of thermodynamics simply by throwing some matter into a black hole: the total entropy of the universe would effectively decrease according to an observer who remains outside the hole. This led Bekenstein (1972) to suggest that black holes have an entropy proportional to their area, as above.

There is a serious problem with this proposal: if (10.2) is correct then a black hole has a temperature and hence must emit radiation just like any other hot body in empty space. But, by definition, a black hole cannot emit anything!

These different ideas were all drawn together into a consistent picture by Hawking's famous discovery (1974) that, if one treats matter quantum mechanically then a black hole *does* emit radiation, with a blackbody spectrum at the *Hawking temperature*

$$T_H = \frac{\hbar\kappa}{2\pi} \tag{10.3}$$

Hence black holes are indeed thermodynamic objects, and the laws of black hole mechanics *are* the laws of thermodynamics applied to these objects. Hawking's calculation determines the correct value of λ to use in (10.2).

In this chapter, we will explain Hawking's result. In order to do this we need to study quantum field theory in curved spacetime. QFT is usually studied in Minkowski spacetime and the standard approach relies heavily on the symmetries of Minkowski spacetime. We will see that several familiar features of flat spacetime QFT are absent, or ambiguous in curved spacetime.

10.2 Quantization of the free scalar field

Let (M, g) be a globally hyperbolic spacetime. Perform a $3+1$ decomposition of the metric as explained in section 3.1:

$$ds^2 = -N^2 dt^2 + h_{ij}(dx^i + N^i dt)(dx^j + N^j dt) \tag{10.4}$$

Let Σ_t denote a (Cauchy) surface of constant t. The future-directed unit normal to this is $n_a = -N(dt)_a$. The metric on Σ_t is h_{ij} and we have $\sqrt{-g} = N\sqrt{h}$.

Consider a massive real Klein-Gordon field with action

$$S = \int_M dt d^3x \sqrt{-g} \left(-\frac{1}{2} g^{ab} \partial_a \Phi \partial_b \Phi - \frac{1}{2} m^2 \Phi^2 \right) \tag{10.5}$$

and equation of motion

$$g^{ab} \nabla_a \nabla_b \Phi - m^2 \Phi = 0 \tag{10.6}$$

The canonical momentum conjugate to Φ is obtained by varying the action:

$$\Pi(x) = \frac{\delta S}{\delta(\partial_t \Phi(x))} = -\sqrt{-g} g^{t\mu} \partial_\mu \Phi = -N\sqrt{h}(dt)_\nu g^{\nu\mu} \partial_\mu \Phi = \sqrt{h} n^\mu \partial_\mu \Phi \tag{10.7}$$

To quantize, we promote Φ and Π to *operators* and impose the canonical commutation relations (units: $\hbar = 1$)

$$[\Phi(t,x), \Pi(t,x')] = i\delta^{(3)}(x - x') \qquad [\Phi(t,x), \Phi(t,x')] = 0 \qquad [\Pi(t,x), \Pi(t,x')] = 0 \tag{10.8}$$

We now want to introduce a Hilbert space of states that these operators act on. Let $\mathcal{S}$ be the space of *complex* solutions of the KG equation. Global hyperbolicity implies that a point of $\mathcal{S}$ is specified uniquely by initial data $\Phi, \partial_t \Phi$ on Σ_0. For $\alpha, \beta \in \mathcal{S}$ we can define

$$(\alpha, \beta) = -\int_{\Sigma_0} d^3x \sqrt{h}\, n_a j^a(\alpha, \beta) \tag{10.9}$$

where j_a is defined by

$$j(\alpha, \beta) = -i \left(\bar{\alpha} d\beta - \beta d\bar{\alpha} \right) \tag{10.10}$$

Note that this can be calculated just from the initial data on Σ_0. Now

$$\nabla^a j_a = -i \left(\bar{\alpha} \nabla^2 \beta - \beta \nabla^2 \bar{\alpha} \right) = -im^2 (\bar{\alpha}\beta - \beta\bar{\alpha}) = 0 \tag{10.11}$$

so j is conserved. It follows that we can replace Σ_0 by any surface Σ_t in (10.9) and get the same result. Note the following properties:

$$(\alpha, \beta) = \overline{(\beta, \alpha)} \tag{10.12}$$

which implies that $(,)$ is a Hermitian form. It is non-degenerate: if $(\alpha, \beta) = 0$ for all $\beta \in \mathcal{S}$ then $\alpha = 0$. However,

$$(\alpha, \beta) = -(\bar{\beta}, \bar{\alpha}) \tag{10.13}$$

so $(\alpha, \alpha) = -(\bar{\alpha}, \bar{\alpha})$ so $(,)$ is *not* positive definite.

In Minkowski spacetime, $(,)$ is positive definite on the subspace $\mathcal{S}_p$ of $\mathcal{S}$ consisting of *positive frequency* solutions. A basis for $\mathcal{S}_p$ are the positive frequency plane waves:

$$\psi_{\mathbf{p}}(x) = \frac{1}{(2\pi)^{3/2}(2p^0)^{1/2}} e^{ip\cdot x} \qquad p^0 = \sqrt{\mathbf{p}^2 + m^2} \tag{10.14}$$

where x denotes inertial frame coordinates $(t, \mathbf{x})$. These modes (solutions) are positive frequency in the sense that, if $k = \partial/\partial t$ then they have negative imaginary eigenvalue w.r.t. $\mathcal{L}_k$:

$$\mathcal{L}_k \psi_{\mathbf{p}} = -ip^0 \psi_{\mathbf{p}} \tag{10.15}$$

The complex conjugate of $\psi_{\mathbf{p}}$ is a negative frequency plane wave. These are orthogonal to the positive frequency plane waves so we have the orthogonal decomposition

$$\mathcal{S} = \mathcal{S}_p \oplus \bar{\mathcal{S}}_p \tag{10.16}$$

where $(,)$ is positive definite on $\mathcal{S}_p$ and negative definite on $\bar{\mathcal{S}}_p$.

In curved spacetime, we do not have a definition of "positive frequency" except when the spacetime is stationary (see below). Hence there is no preferred way to decompose $\mathcal{S}$ as above. Instead, we simply choose a subspace $\mathcal{S}_p$ for which $(,)$ is positive definite and (10.16) holds. In general there will be many ways to do this.

In the quantum theory, we define the creation and annihilation operators associated to a mode $f \in \mathcal{S}_p$ of a real scalar field ($\Phi^\dagger = \Phi$) by

$$a(f) = (f, \Phi) \qquad a(f)^\dagger = -(\bar{f}, \Phi) \tag{10.17}$$

e.g. taking $f = \psi_{\mathbf{p}}$ in Minkowski spacetime gives the usual $a(f) = a_{\mathbf{p}}$. The canonical commutation relations imply (examples sheet 4)

$$[a(f), a(g)^\dagger] = (f, g) \qquad [a(f), a(g)] = [a(f)^\dagger, a(g)^\dagger] = 0 \tag{10.18}$$

e.g. in Minkowski spacetime with $f = \psi_{\mathbf{p}}$ and $g = \psi_{\mathbf{q}}$, the first condition gives $[a_{\mathbf{p}}, a^\dagger_{\mathbf{q}}] = \delta^{(3)}(\mathbf{p} - \mathbf{q})$.

We define a vacuum state $|0\rangle$ by the conditions

$$a(f)|0\rangle = 0 \qquad \forall f \in \mathcal{S}_p \qquad \langle 0|0\rangle = 1 \tag{10.19}$$

Given a basis $\{\psi_i\}$ for $\mathcal{S}_p$, we define the N-particle states as

$$a^\dagger_{i_1} \dots a^\dagger_{i_N} |0\rangle \tag{10.20}$$

where

$$a_i = a(\psi_i) \tag{10.21}$$

(Here the index i might be continuous e.g. in flat spacetime, basis elements are usually labelled by 3-momentum $\mathbf{p}$.) We then choose the Hilbert space to be the Fock space spanned by the vacuum state, the 1-particle states, the 2-particles states etc. The fact that elements of $\mathcal{S}_p$ have positive Klein-Gordon norm implies that this Hilbert space has a positive definite inner product e.g.

$$||a(f)^\dagger|0\rangle||^2 = \langle 0|a(f)a(f)^\dagger|0\rangle = \langle 0|[a(f), a(f)^\dagger]|0\rangle = (f, f) > 0 \tag{10.22}$$

In a general curved spacetime there is no preferred choice of $\mathcal{S}_p$, instead there will be many inequivalent choices. Let $\mathcal{S}'_p$ be another choice of positive frequency subspace. Then any $f' \in \mathcal{S}'_p$ can be decomposed uniquely as $f' = f + \bar{g}$ with $f, g \in \mathcal{S}_p$. Hence

$$a(f') = (f, \Phi) + (\bar{g}, \Phi) = a(f) - a(g)^\dagger \tag{10.23}$$

so $a(f')|0\rangle \neq 0$ hence $|0\rangle$ is not the vacuum state if one uses $\mathcal{S}'_p$ as the positive frequency subspace. In fact it can be shown that the vacuum state defined using $\mathcal{S}'_p$ does not even belong to the Hilbert space that one defines using $\mathcal{S}_p$! Since the vacuum state depends on the choice of $\mathcal{S}_p$, so does the definition of 1-particle states etc. So there is no natural notion of particles in a general curved spacetime.

Why doesn't this issue arise in Minkowski spacetime? In a *stationary* spacetime, one can use the time translation symmetry to identify a preferred choice of $\mathcal{S}_p$. Let k^a be the (future-directed) time-translation Killing vector field. Since this generates a symmetry, it follows that $\mathcal{L}_k$ (the Lie derivative w.r.t. k) commutes with $\nabla^2 - m^2$ and therefore maps $\mathcal{S}$ to $\mathcal{S}$. It can be shown that $\mathcal{L}_k$ is anti-hermitian w.r.t $(,)$ (examples sheet 4) and hence has purely imaginary eigenvalues. We say that an eigenfunction has positive frequency if the eigenvalue is negative imaginary:

$$\mathcal{L}_k u = -i\omega u \qquad \omega > 0 \tag{10.24}$$

(The flat spacetime solutions (10.14) have positive frequency.) Such solutions have positive KG norm (examples sheet 4) so we define $\mathcal{S}_p$ to be the space spanned by these positive frequency eigenfunctions. Complex conjugation shows that the solution $\bar{u}$ is a negative frequency eigenfunction. The anti-hermitian property implies that eigenfunctions with distinct eigenvalues are orthogonal so we indeed have an orthogonal decomposition as in (10.16).

10.3 Bogoliubov transformations

Let $\{\psi_i\}$ be an orthonormal basis for $\mathcal{S}_p$:

$$(\psi_i, \psi_j) = \delta_{ij} \qquad \Rightarrow \qquad (\bar{\psi}_i, \bar{\psi}_j) = -\delta_{ij} \tag{10.25}$$

The orthogonality of the decomposition (10.16) implies

$$(\psi_i, \bar{\psi}_j) = 0 \tag{10.26}$$

Expanding the quantum field in this basis gives

$$\Phi = \sum_j \left(c_j \psi_j + d_j \bar{\psi}_j\right) \tag{10.27}$$

We define the annihilation operators a_i by (10.21) then (10.17) gives $a_i = c_i$ and $a_i^\dagger = d_i$ so

$$\Phi = \sum_i \left(a_i \psi_i + a_i^\dagger \bar{\psi}_i\right) \tag{10.28}$$

For such a basis we have

$$[a_i, a_j^\dagger] = \delta_{ij} \qquad [a_i, a_j] = 0 \tag{10.29}$$

Let $\mathcal{S}_p'$ be a different choice for the positive frequency subspace, with orthonormal basis $\{\psi_i'\}$. This will be related to the first basis by a *Bogoliubov transformation*:

$$\psi_i' = \sum_j \left(A_{ij}\psi_j + B_{ij}\bar{\psi}_j\right) \qquad \bar{\psi}_i' = \sum_j \left(\bar{B}_{ij}\psi_j + \bar{A}_{ij}\bar{\psi}_j\right) \tag{10.30}$$

A, B are called *Bogoliubov coefficients*. For $\mathcal{S}_p'$ we define annihilation operators $a_i' = a(\psi_i')$.

Exercise. Substitute (10.30) into $a_i' = (\psi_i', \Phi)$ to obtain

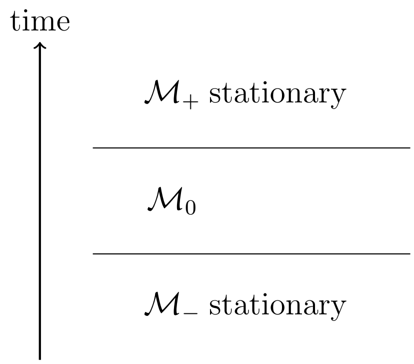
$$a_i' = \sum_j \left(\bar{A}_{ij} a_j - \bar{B}_{ij} a_j^\dagger\right) \tag{10.31}$$

Show also that the requirement that the second basis obeys the conditions (10.25) and (10.26) implies that

$$\sum_k \left(\bar{A}_{ik} A_{jk} - \bar{B}_{ik} B_{jk}\right) = \delta_{ij} \qquad \text{i.e.} \qquad AA^\dagger - BB^\dagger = 1 \tag{10.32}$$

$$\sum_k \left(A_{ik} B_{jk} - B_{ik} A_{jk}\right) = 0 \qquad \text{i.e.} \qquad AB^T - BA^T = 0 \tag{10.33}$$

10.4 Particle production in a non-stationary spacetime



Consider a globally hyperbolic spacetime (M, g) which is stationary at early time, then becomes non-stationary, and finally becomes stationary again. Write $M = M_- \cup M_0 \cup M_+$ where $(M_\pm, g)$ are stationary but (M_0, g) is non-stationary.

In the spacetimes $(M_\pm, g)$, stationarity implies that there is a preferred choice of positive frequency subspace $\mathcal{S}_p^\pm$ and hence the notion of particles is well-defined at early time and again at late time. Global hyperbolicity implies that any solution of the KG equation in $(M_\pm, g)$ extends uniquely to (M, g). Hence we have two choices of positive frequency subspace for (M, g): $\mathcal{S}_p^+$ and $\mathcal{S}_p^-$.

Let $\{u_i^\pm\}$ denote an orthonormal basis for $\mathcal{S}_p^\pm$ and let $a_i^\pm$ be the associated annihilation operators. The bases are related by a Bogoliubov transformation:

$$u_i^+ = \sum_j \left(A_{ij} u_j^- + B_{ij} \bar{u}_j^- \right) \tag{10.34}$$

from (10.31) we have

$$a_i^+ = \sum_j \left(\bar{A}_{ij} a_j^- - \bar{B}_{ij} a_j^{-\dagger} \right) \tag{10.35}$$

Denote the vacua defined w.r.t. $\mathcal{S}_p^\pm$ as $|0\pm\rangle$ i.e. $a_i^\pm|0\pm\rangle = 0$. Assume that no particles are present at early time so the state is $|0-\rangle$. The particle number operator for the ith late-time mode is $N_i^+ = a_i^{+\dagger} a_i^+$, so the expected number of such particles present is

$$\begin{aligned}\langle 0-|N_i^+|0-\rangle = \langle 0-|a_i^{+\dagger} a_i^+|0-\rangle &= \sum_{j,k} \langle 0-|a_k^-(-B_{ik})(-\bar{B}_{ij}) a_j^{-\dagger}|0-\rangle \\ &= \sum_{j,k} B_{ik}\bar{B}_{ij}\langle 0-|a_k^- a_j^{-\dagger}|0-\rangle = \sum_j B_{ij}\bar{B}_{ij} = (BB^\dagger)_{ii}\end{aligned} \tag{10.36}$$

using the expression for the commutator in the penultimate step. The expected *total* number of particles present at late time is $\mathrm{tr}(BB^\dagger) = \mathrm{tr}(B^\dagger B)$, which vanishes iff $B = 0$ i.e. iff $\mathcal{S}_p^+ = \mathcal{S}_p^-$, which will not be true generically. In this example, one can say that a time-dependent gravitational field results in particle production. But we emphasise that this interpretation is possible here only because of the assumed stationarity at early and late times.

10.5 Rindler spacetime

Consider the geometry near the event horizon of a Schwarzschild black hole. Define a new radial coordinate x by

$$r = 2M + \frac{x^2}{8M} \tag{10.37}$$

then the metric becomes (exercise)

$$ds^2 = -\kappa^2 x^2 dt^2 + dx^2 + (2M)^2 d\Omega^2 + \ldots \tag{10.38}$$

where $\kappa = 1/(4M)$ is the surface gravity and the ellipsis denotes terms that are sub-leading near $x = 0$. The first two terms of the above metric are

$$ds^2 = -\kappa^2 x^2 dt^2 + dx^2 \qquad x > 0 \tag{10.39}$$

This is called *Rindler spacetime*. It is a popular toy model for understanding physics near a black hole horizon. There is a coordinate singularity at $x = 0$ which can be removed by introducing Kruskal-like coordinates

$$U = -xe^{-\kappa t} \qquad V = xe^{\kappa t} \tag{10.40}$$

with the result

$$ds^2 = -dUdV = -dT^2 + dX^2 \tag{10.41}$$

where (T, X) are defined by

$$U = T - X \qquad V = T + X \tag{10.42}$$

so Rindler spacetime is flat. But it corresponds to just part of Minkowski spacetime because $U < 0$ and $V > 0$: see Fig. 48.

This is analogous to region I of the Kruskal spacetime. There is another Rindler region analogous to region IV of Kruskal. We will refer to these two Rindler regions as R and L respectively. The lines $U = 0$ and $V = 0$ correspond to a bifurcate Killing horizon of $k = \partial/\partial t$ with surface gravity $\pm\kappa$. In (U, V) coordinate we have

$$k = \kappa \left(V \frac{\partial}{\partial V} - U \frac{\partial}{\partial U} \right) \tag{10.43}$$

Orbits of k (i.e. lines of constant x) are worldlines of observers whose proper acceleration (9.10) is $A_a = (1/x)(dx)_a$ with norm $|A| = 1/x$. Such a "Rindler observer" would naturally regard k as the generator of time translations, and use it to define "positive frequency". However, this differs from the conventional definition of positive

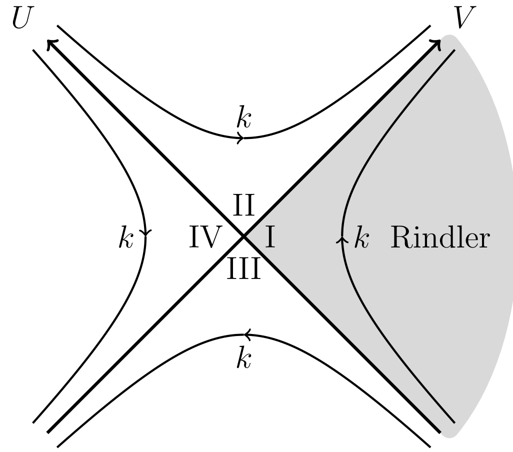


Figure 48. Rindler spacetime is the shaded subset of Minkowski spacetime.

frequency in Minkowski spacetime, which uses $\partial/\partial T$. Let's investigate how the standard Minkowski vacuum state appears to a Rindler observer. We will use $\mathcal{S}_p$ to denote the usual Minkowski definition of positive frequency.

Consider the massless Klein-Gordon equation (wave equation). In inertial coordinates this is

$$\left(-\frac{\partial^2}{\partial T^2} + \frac{\partial^2}{\partial X^2}\right)\Phi = 0 \tag{10.44}$$

The general solution consists of a "right-moving" part and and a "left-moving" part:

$$\Phi = f(U) + g(V) \tag{10.45}$$

The standard Minkowski basis of positive frequency solutions is

$$u_p(T, X) = c_p e^{-i(\omega T - pX)} \qquad \omega = |p| \tag{10.46}$$

where c_p is a normalization constant. This can also be written as

$$u_p = \begin{cases} c_p e^{-i\omega U} \text{ if } p > 0 \text{ (right movers)} \\ c_p e^{-i\omega V} \text{ if } p < 0 \text{ (left movers)} \end{cases} \tag{10.47}$$

We now want to find a basis of positive frequency solutions for Rindler spacetime. A solution with frequency σ w.r.t. k has time dependence $e^{-i\sigma t}$ so the wave equation is

$$0 = \nabla^a \nabla_a \Phi = \frac{1}{\sqrt{-g}}\partial_\mu\left(\sqrt{-g}g^{\mu\nu}\partial_\nu\Phi\right) = \frac{1}{x^2}\left[x\partial_x\left(x\partial_x\Phi\right) + \frac{\sigma^2}{\kappa^2}\Phi\right] \tag{10.48}$$

with solutions $\Phi \propto e^{-i\sigma t}x^{iP}$ where $P = \pm\sigma/\kappa$. If $\sigma > 0$ then the $P > 0$ solution is a right-moving mode because x increases with t along lines of constant phase. Similarly the $P < 0$ solution is a left-moving mode. We can now define a basis of positive frequency solutions in R by

$$u_P^R = C_P e^{-i(\sigma t - P \log x)} \qquad \sigma = \kappa |P| \tag{10.49}$$

for some normalisation constant C_P.

We will want to relate these to the standard Minkowski modes. To do this, it is useful to extend the definition of the Rindler modes to the whole of Minkowski spacetime. We do this by defining $u_P^R = 0$ in L. The solution is then uniquely determined throughout Minkowski spacetime. Converting to the Kruskal-like coordinates gives

$$u_P^R = \begin{cases} \left.\begin{cases} C_P e^{i\frac{\sigma}{\kappa}\log(-U)} & U < 0 \\ 0 & U > 0 \end{cases}\right\} & P > 0 \text{ (right movers)} \\ \left.\begin{cases} 0 & V < 0 \\ C_P e^{-i\frac{\sigma}{\kappa}\log(V)} & V > 0 \end{cases}\right\} & P < 0 \text{ (left movers)} \end{cases} \tag{10.50}$$

(These are solutions everywhere since they have the form (10.45).) We would like to choose the constant C_P so that the above modes have unit norm w.r.t. the KG inner product in Rindler spacetime. However, there is a problem here, which also arises for the Minkowski modes (10.46): these modes are not normalizable. To deal with this problem one can instead consider *wavepackets* constructed as superpositions of positive frequency modes and work with a basis of such wave packets. We won't do this but it means we will encounter certain integrals below that do not converge. We will manipulate them as if they did converge, a more rigorous treatment would use the wavepacket basis. We also won't need to choose a value of C_P here.

The modes u_P^R do not supply a basis for solutions in Minkowski spacetime (e.g. because they vanish in L). We can obtain a second set of modes, which is non-vanishing in L and vanishing in R, by applying the isometry $(U, V) \to (-U, -V)$:

$$\bar{u}_P^L = \begin{cases} \left.\begin{cases} C_P e^{i\frac{\sigma}{\kappa}\log(U)} & U > 0 \\ 0 & U < 0 \end{cases}\right\} & P > 0 \\ \left.\begin{cases} 0 & V > 0 \\ C_P e^{-i\frac{\sigma}{\kappa}\log(-V)} & V < 0 \end{cases}\right\} & P < 0 \end{cases} \tag{10.51}$$

The reason for the overbar on the LHS is that the isometry preserves k^a hence these modes will be positive frequency w.r.t. k^a. But k^a is past-directed in L. Hence it is more natural to use $-k^a$ to define the notion of positive frequency in L. The above modes are negative frequency w.r.t. $-k^a$ hence the overbar. (However, nothing will depend on how we define positive frequency in L.) Now $\{u_P^R, \bar{u}_P^R, u_P^L, \bar{u}_P^L\}$ is a basis for solutions in Minkowski spacetime.

We now discuss a useful condition which ensures that a mode is positive frequency w.r.t. $\partial/\partial T$. To decompose a right-moving mode $f(U)$ into Minkowski modes of frequency ω we perform a Fourier transform:

$$f(U) = \int_{-\infty}^{\infty} \frac{d\omega}{2\pi} e^{-i\omega U} \tilde{f}(\omega) \tag{10.52}$$

where

$$\tilde{f}(\omega) = \int_{-\infty}^{\infty} dU e^{i\omega U} f(U) \tag{10.53}$$

Assume that, in the lower half of the complex U-plane, $f(U)$ is analytic with $\max_{\theta\in[-\pi,0]} |f(Re^{i\theta})| \to 0$ as $R \to \infty$. Then, for $\omega < 0$, we can close the contour in the lower half-plane to deduce that $\tilde{f}(\omega) = 0$ (Jordan's lemma). Hence such $f(U)$ is positive frequency w.r.t. $\partial/\partial T$, i.e., an element of $\mathcal{S}_p$.

To apply this result, consider for $P > 0$ and $U > 0$:

$$\bar{u}_P^L = C_P e^{i\frac{\sigma}{\kappa}\log U} = C_P e^{i\frac{\sigma}{\kappa}[\log(-U) - i\pi]} = C_P e^{\frac{\pi\sigma}{\kappa}} e^{i\frac{\sigma}{\kappa}\log(-U)} \tag{10.54}$$

where we define the logarithm in the complex plane by taking a branch cut along the negative imaginary axis:

$$\log z = \log|z| + i\arg z \qquad \arg z \in (-\pi/2, 3\pi/2) \tag{10.55}$$

Hence we have

$$u_P^R + e^{-\frac{\pi\sigma}{\kappa}} \bar{u}_P^L = C_P e^{i\frac{\sigma}{\kappa}\log(-U)} \qquad P > 0 \tag{10.56}$$

for all U. This is analytic in the lower half U-plane. It does not decay as $|U| \to \infty$ but this is a consequence of working with non-normalizable modes (the integral (10.53) does not converge). Modulo this technicality, we deduce that the above combination of Rindler modes is an element of $\mathcal{S}_p$. For $P < 0$ we have

$$u_P^R + e^{-\frac{\pi\sigma}{\kappa}} \bar{u}_P^L = C_P e^{-\frac{\pi\sigma}{\kappa}} e^{-i\frac{\sigma}{\kappa}\log(-V)} \qquad P < 0 \tag{10.57}$$

which is similarly analytic in the lower half V-plane and therefore a superposition of the positive frequency left-moving Minkowski modes. Similarly

$$u_P^L + e^{-\frac{\pi\sigma}{\kappa}} \bar{u}_P^R = \begin{cases} C_P e^{-\frac{\pi\sigma}{\kappa}} e^{-i\frac{\sigma}{\kappa}\log(-U)} & P > 0 \\ C_P e^{i\frac{\sigma}{\kappa}\log(-V)} & P < 0 \end{cases} \tag{10.58}$$

which is also analytic in the lower half U, V planes and therefore an element of $\mathcal{S}_p$. So we have a new set of positive frequency (w.r.t. $\partial/\partial T$) modes

$$v_P^{(1)} = D_P^{(1)} \left(u_P^R + e^{-\frac{\pi\sigma}{\kappa}} \bar{u}_P^L \right) \qquad v_P^{(2)} = D_P^{(2)} \left(u_P^L + e^{-\frac{\pi\sigma}{\kappa}} \bar{u}_P^R \right) \tag{10.59}$$

where $D_P^{(i)}$ are normalization constants. Notice that u_P^R can be expressed as linear combinations of $v_P^{(1)}$ and $\bar{v}_P^{(2)}$. Since the latter has *negative* frequency, it follows that u_P^R is a mixture of both positive and negative Minkowski space modes (and similarly for u_P^L).

This new set of modes, together with their complex conjugates, forms a basis for $\mathcal{S}$. Since $v_P^{(i)}$ are positive frequency w.r.t. $\partial/\partial T$ it follows that $\{v_P^{(1)}, v_P^{(2)}\, \forall P\}$ is a basis for $\mathcal{S}_p$. Hence the vacuum state defined using annihilation operators $a_P^{(1)}$ and $a_P^{(2)}$ for this basis will agree with that defined using the usual Minkowski modes:

$$a_P^{(i)}|0\rangle = 0 \tag{10.60}$$

where $|0\rangle$ is the standard Minkowski vacuum state.

To fix the normalisation, we use the orthogonality of u_P^R and $\bar{u}_P^L$, and the properties of the KG norm to obtain

$$\begin{aligned}(v_P^{(1)}, v_P^{(1)}) &= |D_P^{(1)}|^2 \left[(u_P^R, u_P^R) - e^{-2\frac{\pi\sigma}{\kappa}}(u_P^L, u_P^L)\right] \\ &= 2|D_P^{(1)}|^2 e^{-\frac{\pi\sigma}{\kappa}} \sinh(\pi\sigma/\kappa)(u_P^R, u_P^R)\end{aligned} \tag{10.61}$$

using the fact that the L modes have the same norm as the R modes. A similar result holds for $v_P^{(2)}$. So we normalize by choosing

$$D_P^{(i)} = \frac{e^{\frac{\pi\sigma}{2\kappa}}}{\sqrt{2\sinh(\pi\sigma/\kappa)}} \tag{10.62}$$

We then have (exercise)

$$u_P^R = \frac{1}{\sqrt{2\sinh(\pi\sigma/\kappa)}}\left(e^{\frac{\pi\sigma}{2\kappa}} v_P^{(1)} - e^{-\frac{\pi\sigma}{2\kappa}} \bar{v}_P^{(2)}\right) \tag{10.63}$$

and hence, using (10.17), the annihilation operators for the R Rindler modes are

$$\begin{aligned}b_P^R \equiv (u_P^R, \Phi) &= \frac{1}{\sqrt{2\sinh(\pi\sigma/\kappa)}}\left[e^{\frac{\pi\sigma}{2\kappa}}(v_P^{(1)}, \Phi) - e^{-\frac{\pi\sigma}{2\kappa}}(\bar{v}_P^{(2)}, \Phi)\right] \\ &= \frac{1}{\sqrt{2\sinh(\pi\sigma/\kappa)}}\left[e^{\frac{\pi\sigma}{2\kappa}} a_P^{(1)} + e^{-\frac{\pi\sigma}{2\kappa}} a_P^{(2)\dagger}\right]\end{aligned} \tag{10.64}$$

In R, the number operator for Rindler particles of momentum P is $N_P^R = b_P^{R\dagger} b_P^R$. How many such particles does a Rindler observer see in the Minkowski vacuum state? The expected number is (using (10.60))

$$\begin{aligned}\langle 0|N_P^R|0\rangle &= \frac{e^{-\frac{\pi\sigma}{\kappa}}}{2\sinh(\pi\sigma/\kappa)}\langle 0|a_P^{(2)} a_P^{(2)\dagger}|0\rangle = \frac{1}{e^{\frac{2\pi\sigma}{\kappa}} - 1}\langle 0|[a_P^{(2)}, a_P^{(2)\dagger}]|0\rangle \\ &= \frac{1}{e^{\frac{2\pi\sigma}{\kappa}} - 1}(v_P^{(2)}, v_P^{(2)})\end{aligned} \tag{10.65}$$

using (10.18). The RHS involves the norm of the mode $v_P^{(2)}$ which, by (10.61) and (10.62), is the same as that of the mode u_P^R. Although this mode is not normalizable, we will assume that it is, with the justification that this can be made rigorous by using a basis of wavepackets. Hence we have

$$\langle 0|N_P^R|0\rangle = \frac{1}{e^{\frac{2\pi\sigma}{\kappa}} - 1} \tag{10.66}$$

Consider a Rindler observer at fixed x. Her 4-velocity is

$$\frac{1}{\kappa x}\frac{\partial}{\partial t} = \frac{A}{\kappa}\frac{\partial}{\partial t} \tag{10.67}$$

where $A = 1/x$ is the magnitude of her proper acceleration. Hence, according to her, the frequency of a R mode is $\hat{\sigma} = A\sigma/\kappa$. So

$$\langle 0|N_P^R|0\rangle = \frac{1}{e^{\frac{2\pi\hat{\sigma}}{A}} - 1} \tag{10.68}$$

This is the Planck spectrum of thermal radiation at the *Unruh temperature*

$$T_U = \frac{A}{2\pi} \tag{10.69}$$

in units where Boltzmann's constant $k_B = 1$. A uniformly accelerating observer perceives the Minkowski vacuum state as a thermal state at the temperature T_U. This is a physical effect: if the observer carries a sufficiently sensitive particle detector then it will detect particles! However, for plausible values of a, the effect is very small. In physical units we have

$$T_U \approx \left(\frac{A}{10^{19}\text{ms}^{-2}}\right)\text{K} \tag{10.70}$$

10.6 Wave equation in Schwarzschild spacetime

To discuss Hawking radiation we first need to understand the behaviour of solutions of the wave equation in the Schwarzschild spacetime. Work in Schwarzschild coordinates.We can decompose a KG field Φ into spherical harmonics $Y_{lm}(\theta, \phi)$:

$$\Phi = \sum_{l=0}^{\infty}\sum_{m=-l}^{l}\frac{1}{r}\phi_{lm}(t,r)Y_{lm}(\theta,\phi) \tag{10.71}$$

The wave equation $\nabla^a\nabla_a\Phi = 0$ reduces to (examples sheet 4)

$$\left[\frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial r_*^2} + V_l(r_*)\right]\phi_{lm} = 0 \tag{10.72}$$

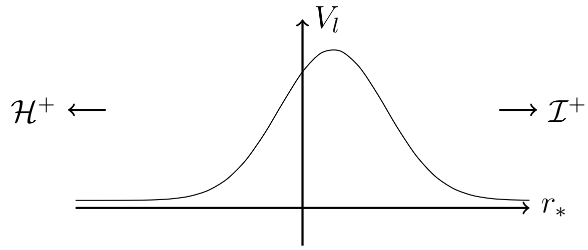


Figure 49. Effective potential for the wave equation in the Schwazschild spacetime.

where

$$V_l(r_*) = \left(1 - \frac{2M}{r}\right)\left(\frac{l(l+1)}{r^2} + \frac{2M}{r^3}\right) \tag{10.73}$$

where on the RHS we view r as a function of r_*. This has the form of a 2d wave equation with a potential $V_l(r_*)$ sketched in Fig. 49.

Note that $V_l(r_*)$ vanishes as $r_* \to \infty$ ($r \to \infty$, i.e., $\mathcal{I}^\pm$) and as $r_* \to -\infty$ ($r \to 2M+$, i.e., $\mathcal{H}^\pm$). Consider a solution describing a wavepacket localized at some finite value of r_* at time t_0. At late time $t \to \infty$ we expect the solution to consist of a superposition of two wavepackets, propagating to the "left" ($r_* \to -\infty$) and to the "right" ($r_* \to \infty$). Time reversal implies that at early time $t \to -\infty$ the solution consists of a superposition of wavepackets propagating in from the left and the right. Hence we expect

$$\phi_{lm} \approx f_\pm(t - r_*) + g_\pm(t + r_*) = f_\pm(u) + g_\pm(v) \qquad \text{as } t \to \pm\infty \tag{10.74}$$

where $f_\pm$ and $g_\pm$ are each localized around some particular value of u or v and hence vanish for $|u| \to \infty$ or $|v| \to \infty$. The full solution is uniquely determined by its behaviour for $t \to \infty$ or $t \to -\infty$ i.e. by either f_+, g_+ or by f_-, g_-.

At late time the term $f_+(u)$ describes an outgoing wavepacket propagating to $\mathcal{I}^+$ whereas $g_+(v)$ describes an ingoing wavepacket propagating to $\mathcal{H}^+$. More precisely, if we evaluate the above solution on $\mathcal{I}^+$ (where $v \to \infty$ with finite u) we obtain the result $f_+(u)$. Similarly we can evaluate on $\mathcal{H}^+$ (where $u \to \infty$ with finite v) to obtain the result $g_+(v)$. Hence the solution is uniquely determined (for all t) by specifying its behaviour on $\mathcal{I}^+ \cup \mathcal{H}^+$.

We will define an "out" mode to be a solution which vanishes on $\mathcal{H}^+$ and a "down" mode to be a solution which vanishes on $\mathcal{I}^+$. From what we have just said, any solution of (10.72) can be written uniquely as a superposition of an out mode and a down mode. Out modes and down modes are orthogonal since we can evaluate the integral defining the KG inner product at late time, when the out modes are non-zero only near $r_* = \infty$ and the down modes are non-zero only near $r_* = -\infty$.

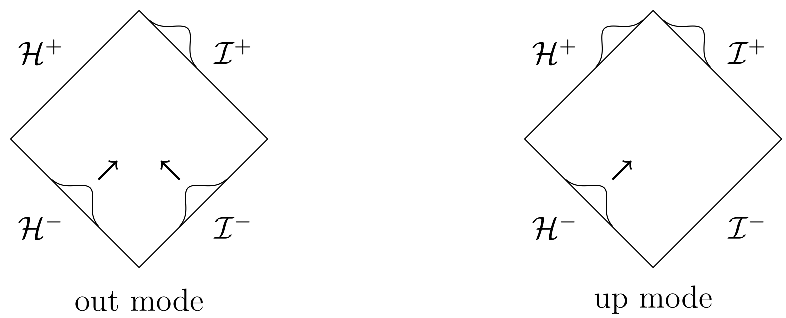


Figure 50. Out modes vanish on $\mathcal{H}^+$. Up modes vanish on $\mathcal{I}^-$.

Similarly, at early time, the solution is a superposition of a wavepacket $g_-(v)$ propagating in from $\mathcal{I}^-$ and a wavepacket $f_-(u)$ propagating out from $\mathcal{H}^-$. So the solution is uniquely determined by its behaviour on $\mathcal{I}^- \cup \mathcal{H}^-$. We define an "in" mode to be a solution which vanishes on $\mathcal{H}^-$ and an "up" mode to be a solution which vanishes on $\mathcal{I}^-$. Any solution can be written uniquely as a superposition of an in mode and an up mode.

The late time modes can be written in terms of the early time modes and vice versa. For example, an out mode is a superposition of an in mode and an up mode; an up mode is a superposition of an out mode and a down mode, see Fig. 50.

This spacetime is stationary so we can consider modes with definite frequency i.e. eigenfunctions of $\mathcal{L}_k$ with eigenvalue $-i\omega$. Such modes have time dependence $e^{-i\omega t}$. A mode with frequency $\omega > 0$ has the form

$$\Phi_{\omega lm} = \frac{1}{r} e^{-i\omega t} R_{\omega lm}(r) Y_{lm}(\theta, \phi) \qquad \omega > 0 \tag{10.75}$$

More generally, we say that a solution has positive frequency if it can be written as a superposition of such modes. Setting $\phi_{lm} = e^{-i\omega t} R_{\omega lm}$ above gives the "radial equation"

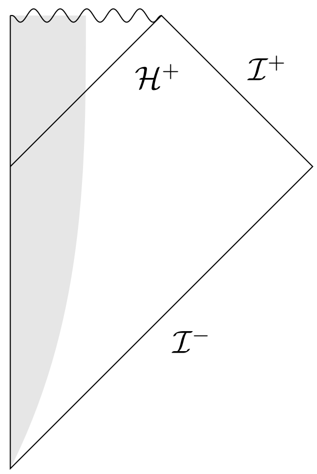
$$\left[-\frac{d^2}{dr_*^2} + V_l(r_*)\right] R_{\omega lm} = \omega^2 R_{\omega lm} \tag{10.76}$$

This has the form of a Schrödinger equation with potential $V_l(r_*)$. Since $V_l(r_*)$ vanishes as $|r_*| \to \infty$ we expect the solutions to behave for $|r_*| \to \infty$ as

$$R_{\omega lm} \sim e^{\pm i\omega r_*} \qquad \Rightarrow \qquad \Phi_{\omega lm} \propto e^{-i\omega(t \mp r_*)} = \begin{Bmatrix} e^{-i\omega u} \\ e^{-i\omega v} \end{Bmatrix} \tag{10.77}$$

The upper (lower) choice of sign corresponds to outgoing (ingoing) waves.

10.7 Hawking radiation



Consider a massless scalar field in the spacetime describing spherically symmetric gravitational collapse, with the Penrose diagram shown. Outside the collapsing matter, the spacetime is described by the Schwarzschild solution, which is static. However, the spacetime is not stationary because the geometry inside the collapsing matter is not stationary. Hence we expect particle creation. The surprising result is that this particle creation is not a transient effect, but there is a steady flux of particles from the black hole at late time.

We will introduce bases analogous to those used above. At early time, there is no past event horizon so there is no analogue of the "up" modes, we have just the "in" modes, i.e., wavepackets propagating in from $\mathcal{I}^-$. The geometry near $\mathcal{I}^-$ is static so there is a natural notion of "positive frequency" there. Let f_i be a basis of "in" modes that are positive frequency near $\mathcal{I}^-$.

At late time, we can define "out" and "down" modes as before, i.e., as wavepackets that vanish on $\mathcal{H}^+$ and $\mathcal{I}^+$ respectively. The geometry near $\mathcal{I}^+$ is static so we can define a notion of "positive frequency" there. Let p_i be a basis of positive frequency out modes. The geometry is not static everywhere on $\mathcal{H}^+$ so there is no natural notion of positive frequency for the down modes. We pick an arbitrary basis $\{q_i, \bar{q}_i\}$ for these modes.

We have two different bases for $\mathcal{S}$, i.e., $\{f_i, \bar{f}_i\}$ and $\{p_i, q_i, \bar{p}_i, \bar{q}_i\}$. We will assume that both bases are orthonormal, i.e., $(f_i, f_j) = \delta_{ij}$ and

$$(p_i, p_j) = (q_i, q_j) = \delta_{ij} \qquad (p_i, q_j) = 0 \tag{10.78}$$

where the orthogonality of the out and down modes was discussed above. Let a_i, b_i be annihilation operators for the "in" and "out" modes respectively:

$$a_i = (f_i, \Phi) \qquad b_i = (p_i, \Phi) \tag{10.79}$$

We can expand

$$p_i = \sum_j \left(A_{ij} f_j + B_{ij} \bar{f}_j \right) \tag{10.80}$$

so from (10.31)

$$b_i = (p_i, \Phi) = \sum_j \left(\bar{A}_{ij} a_j - \bar{B}_{ij} a_j^\dagger \right) \tag{10.81}$$

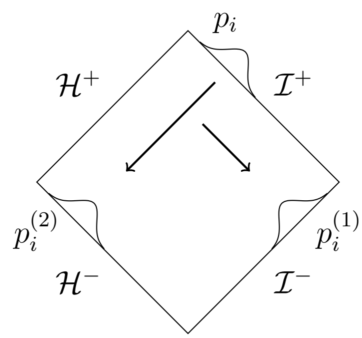


Figure 51. Backwards propagation of an out mode in Kruskal spacetime.

We assume that there are no particles present at early time, i.e., that the state is the vacuum state defined using the modes f_i:

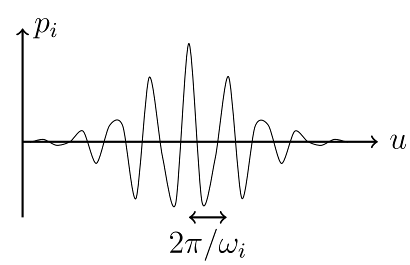
$$a_i|0\rangle = 0 \tag{10.82}$$

The expected number of particles present in the ith "out" mode is then

$$\langle 0|b_i^\dagger b_i|0\rangle = (BB^\dagger)_{ii} \tag{10.83}$$

To calculate this we need to determine the Bogoliubov coefficients B_{ij}.

We will choose our "out" basis elements p_i so that at $\mathcal{I}^+$ they are wavepackets localized around some particular retarded time u_i and containing only positive frequencies localized around some value ω_i, as shown.



We define the "in" basis element f_i to be a (positive frequency) wavepacket on $\mathcal{I}^-$ whose dependence on v is the same as the dependence of p_i on u at $\mathcal{I}^+$.

Consider first Kruskal spacetime. Imagine propagating the wavepacket p_i backwards in time from $\mathcal{I}^+ \cup \mathcal{H}^+$. Part of the wavepacket would be "reflected" to give a wavepacket on $\mathcal{I}^-$ (an in mode) and part would be "transmitted" to give a wavepacket crossing $\mathcal{H}^-$ (an up mode) as shown in Fig. 51. So we can write

$$p_i = p_i^{(1)} + p_i^{(2)} \tag{10.84}$$

where $p_i^{(1)}$ is the "in" part and $p_i^{(2)}$ the "up" part. Let

$$R_i = \sqrt{(p_i^{(1)}, p_i^{(1)})} \qquad T_i = \sqrt{(p_i^{(2)}, p_i^{(2)})} \tag{10.85}$$

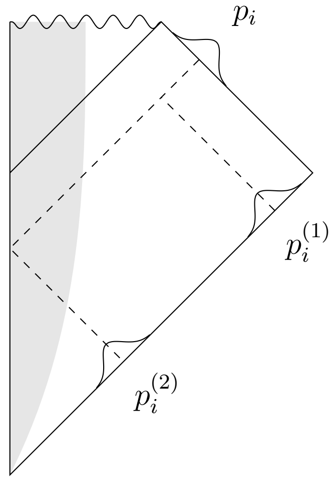


Figure 52. Backwards propagation of an out mode in collapse spacetime.

(Both KG norms are positive because there is no mixing of frequencies in Kruskal spacetime.) Then from the normalisation of p_i and the fact that "in" and "up" modes are orthogonal, we have

$$R_i^2 + T_i^2 = 1 \tag{10.86}$$

R_i is called the reflection coefficient, i.e., the fraction of the wavepacket that is reflected to $\mathcal{I}^-$ and T_i is called the transmission coefficient, i.e., the fraction that crosses $\mathcal{H}^-$. The time reversal symmetry of the Schwarzschild spacetime implies that R_i, T_i are also the reflection and transmission coefficients for the "in" wavepacket f_i propagating in from $\mathcal{I}^-$. Specifically, T_i is the fraction of f_i that crosses $\mathcal{H}^+$ and R_i is the fraction reflected to $\mathcal{I}^+$.

Let's now include the collapsing matter in our spacetime. We will be interested in the case of a wavepacket p_i that is localized around a *late* retarded time u_i. See Fig. 52. The reflected wavepacket will be localized around a late advanced time v_i. In this case, the scattering of the wavepacket occurs outside the collapsing matter and hence behaves just as in Kruskal spacetime. So we can write (10.84) as above, where $p_i^{(1)}$ is defined to be the part of the wavepacket that is scattered outside the collapsing matter. This does not experience the time-dependent geometry of the collapsing matter and so just gives a positive frequency mode at $\mathcal{I}^-$. From the above arguments we know that the norm of $p_i^{(1)}$ is R_i which is the same as the fraction of the mode f_i that is reflected to $\mathcal{I}^+$ in the Kruskal spacetime.

On the other hand, the part of the wavepacket that would have entered $\mathcal{H}^-$ in the Kruskal spacetime now enters the collapsing matter. This is the part $p_i^{(2)}$ in (10.84). It propagates through the collapsing matter and out to $\mathcal{I}^-$. Since it has travelled through a time-dependent geometry, the resulting solution will be a mixture of positive and negative frequency modes at $\mathcal{I}^-$. Hence it is $p_i^{(2)}$ that determines B_{ij}. We can decompose both $p_i^{(1)}$ and $p_i^{(2)}$ as in (10.80) hence we have (as $B_{ij}^{(1)} = 0$)

$$A_{ij} = A_{ij}^{(1)} + A_{ij}^{(2)} \qquad B_{ij} = B_{ij}^{(2)} \tag{10.87}$$

At early time it is clear that $p_i^{(1)}$ and $p_i^{(2)}$ are well-separated wavepackets and hence they are orthogonal w.r.t. the KG inner product. Hence (since p_i has unit norm and $R_i^2 + T_i^2 = 1$) the norm of $p_i^{(2)}$ must be T_i, which is the same as the fraction of the mode f_i which crosses $\mathcal{H}^+$ in the Kruskal spacetime.

To calculate B_{ij} we must determine the behaviour of $p_i^{(2)}$ on $\mathcal{I}^-$. On $\mathcal{I}^+$, the wavepacket p_i has oscillations with characteristic frequency near to ω_i, modulated by a smooth profile (e.g. a Gaussian function) localized around some retarded time u_i. There will be infinitely many of these oscillations along $\mathcal{I}^+$. When these are propagated backwards in time, there will be infinitely many oscillations between the line $u = u_i$ and the event horizon at $u = \infty$. See Fig. 53. This means that an observer who crosses $\mathcal{H}^+$ would observe infinitely many oscillations of the field in a finite affine time, i.e., the proper frequency of the field measured by the observer would diverge at $\mathcal{H}^+$.

Let γ denote a generator of $\mathcal{H}^+$ and extend γ to the past until it intersects $\mathcal{I}^-$. We can define our advanced time coordinate v so that γ intersects $\mathcal{I}^-$ at $v = 0$. Our wavepacket will be localized around some value $v_0 < 0$ on $\mathcal{I}^-$, with infinitely many oscillations in $v_0 < v < 0$. Hence the arguments just given imply that the field oscillates very rapidly near γ all the way back to $\mathcal{I}^-$. Since the field is oscillating so rapidly near γ, we can use the *geometric optics* approximation.

In geometric optics we write the scalar field as $\Phi(x) = A(x)e^{i\lambda S(x)}$ and assume that $\lambda \gg 1$. To leading order in λ the wave equation reduces to $(\nabla S)^2 = 0$, i.e., surfaces of constant phase S are null hypersurfaces. The generators of these hypersurfaces are null geodesics.

Consider a null geodesic congruence containing the generators of these surfaces of constant S, and also the generators of $\mathcal{H}^+$ (which is the surface $S = \infty$). We can introduce a null vector N^a as in section 4.4 such that $N \cdot U = -1$ where U^a is the tangent vector to the geodesics and $U \cdot \nabla N^a = 0$. We can decompose a deviation vector for this congruence into the sum of a part orthogonal to U^a and a term βN^a parallelly transported along the geodesics (equation (4.17)). On $\mathcal{H}^+$, the former is tangent to $\mathcal{H}^+$ but the latter points off $\mathcal{H}^+$ and hence towards a generator of a surface of constant S. Choose $\beta = -\epsilon$ where $\epsilon > 0$ is small. Then $-\epsilon N^a$ is a deviation vector from γ to a generator γ' of a surface of constant S.

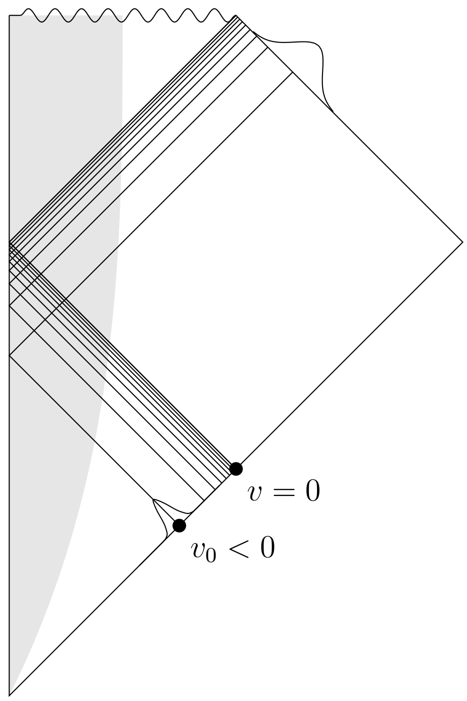
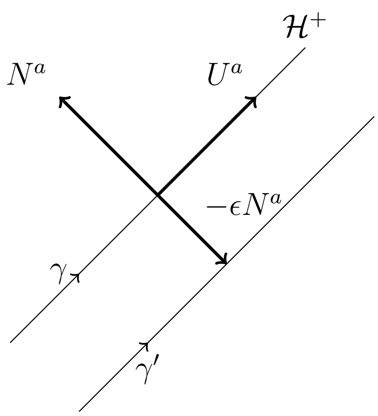


Figure 53. Surfaces of constant phase accumulate near event horizon and past extension of horizon generators.



Spherical symmetry implies that we can choose N^μ such that $N^\theta = N^\phi = 0$. Outside the collapsing matter we know that $\partial/\partial V$ is tangent to the affinely parameterized generators of $\mathcal{H}^+$, so we can choose $U^a = (\partial/\partial V)^a$ there. Since N^μ is null and not parallel to U^μ we must then have $N^V = 0$. From $U \cdot N = -1$ we obtain

$$N = C \frac{\partial}{\partial U} \tag{10.88}$$

for some positive constant C (since g_{UV} is constant on $\mathcal{H}^+$ outside the matter). Hence, outside the collapsing matter, the deviation vector $-\epsilon N^a$ connects γ to a null geodesic γ' with

$$U = -C\epsilon \tag{10.89}$$

Fom the definition of U we have

$$u = -\frac{1}{\kappa} \log(-U) \tag{10.90}$$

Hence, at late time, γ' is an outgoing null geodesic with

$$u = -\frac{1}{\kappa}\log(C\epsilon) \tag{10.91}$$

Let $F(u)$ denote the phase of the wavepacket p_i on $\mathcal{I}^+$. Then the phase everywhere along γ' must be

$$S = F\left(-\frac{1}{\kappa}\log(C\epsilon)\right) \tag{10.92}$$

At $\mathcal{I}^-$, γ, γ' are ingoing radial null geodesics. In (u, v) coordinates this implies that U^a is a multiple of $\partial/\partial u$. The metric near $\mathcal{I}^-$ has the form

$$ds^2 = -dudv + \frac{1}{4}(u-v)^2 d\Omega^2 \tag{10.93}$$

so spherical symmetry and the fact that N is null and not parallel to U implies

$$N = D^{-1}\frac{\partial}{\partial v} \qquad \text{at } \mathcal{I}^- \tag{10.94}$$

for some positive constant D, which implies that γ' intersects $\mathcal{I}^-$ at

$$v = -D^{-1}\epsilon \tag{10.95}$$

Combining with (10.92), we learn that the phase on $\mathcal{I}^-$ is, for small $v < 0$,

$$S = F\left(-\frac{1}{\kappa}\log(-CDv)\right) \tag{10.96}$$

Hence on $\mathcal{I}^-$ we have

$$p_i^{(2)} \approx \begin{cases} 0 & v > 0 \\ A(v)\exp\left[iF\left(-\frac{1}{\kappa}\log(-CDv)\right)\right] & \text{small } \; v < 0 \end{cases} \tag{10.97}$$

where the amplitude $A(v)$ is a smooth positive function. This shows that, on $\mathcal{I}^-$, most of our late time wavepacket is squeezed into a small region near $v = 0$ where the logarithm varies rapidly. To determine B_{ij} we now have to decompose this function into positive and negative frequency "in" modes on $\mathcal{I}^-$.

So far we have been working with normalizable wavepackets built by superposing modes of definite frequency. But now we will assume that p_i contains only the single positive frequency $\omega_i > 0$ so $F(u) = -\omega_i u$. This means that p_i is neither normalizable nor localized at late time (as assumed above) but it makes the rest of the calculation easier. The result is the same as a more rigorous calculation using wavepackets. We will also use ω to label the modes i.e. we will write p_ω instead of p_i (there will be additional labels (l, m) but we will suppress these). For this function p_ω we have on $\mathcal{I}^-$:

$$p_\omega^{(2)} \approx \begin{cases} 0 & v > 0 \\ A_\omega(v) \exp\left[i\frac{\omega}{\kappa}\log(-CDv)\right] & \text{small } \; v < 0 \end{cases} \tag{10.98}$$

Similarly we will use a basis of "in" modes f_σ such that f_σ has frequency $\sigma > 0$, i.e., $f_\sigma = (2\pi N_\sigma)^{-1} e^{-i\sigma v}$ on $\mathcal{I}^-$ where N_σ is a normalization constant. Writing $p_\omega^{(2)}$ in terms of $\{f_\sigma, \bar{f}_\sigma\}$ is therefore just a Fourier transform w.r.t. v on $\mathcal{I}^-$. Since $p_\omega^{(2)}$ is squeezed into a small range of v near $v = 0$ (or would be if it were a wavepacket), its Fourier transform will involve mainly high frequency modes, i.e. large σ. For such modes, the Fourier transform is dominated by the region where $p_\omega^{(2)}$ oscillates most rapidly, i.e., near $v = 0$. So we can use the above expression and approximate the amplitude $A_i(v)$ as a constant. The Fourier transform is therefore

$$\tilde{p}_\omega^{(2)}(\sigma) = A_\omega \int_{-\infty}^{0} dv\, e^{i\sigma v} \exp\left[i\frac{\omega}{\kappa}\log(-CDv)\right] \tag{10.99}$$

with inverse

$$\begin{aligned} p_\omega^{(2)}(v) &= \int_{-\infty}^{\infty} \frac{d\sigma}{2\pi} e^{-i\sigma v} \tilde{p}_\omega^{(2)}(\sigma) \\ &= \int_0^\infty d\sigma N_\sigma \tilde{p}_\omega^{(2)}(\sigma) f_\sigma(v) + \int_0^\infty d\sigma \bar{N}_\sigma \tilde{p}_\omega^{(2)}(-\sigma) \bar{f}_\sigma(v) \end{aligned} \tag{10.100}$$

the first term picks out the positive frequency components and second term the negative frequency components. Hence in (10.80) we have

$$A_{\omega\sigma}^{(2)} = N_\sigma \tilde{p}_\omega^{(2)}(\sigma) \qquad B_{\omega\sigma} = \bar{N}_\sigma \tilde{p}_\omega^{(2)}(-\sigma) \qquad \omega, \sigma > 0 \tag{10.101}$$

The integral in (10.99) is not convergent but this is an artefact of working with non-normalizable states. It would converge if we used wavepackets so we will manipulate it as if it converged. We will want to extend the integrand into the complex v-plane so we define the logarithm with a branch cut in the lower half plane:

$$\log z = \log|z| + i \arg z \qquad \arg z \in (-\pi/2, 3\pi/2) \tag{10.102}$$

which makes the integrand in (10.99) analytic in the lower half plane. If $\sigma > 0$ then the integrand in $\tilde{p}_\omega^{(2)}(-\sigma)$ decays as $v \to \infty$ in the lower half v-plane. Consider the semi-circular contour shown in Fig. 54. The integral around this contour vanishes by Cauchy's theorem. The integral around the curved part of the semi-circle vanishes as $R \to \infty$ (at least it would if we were working with wavepackets, by Jordan's lemma).

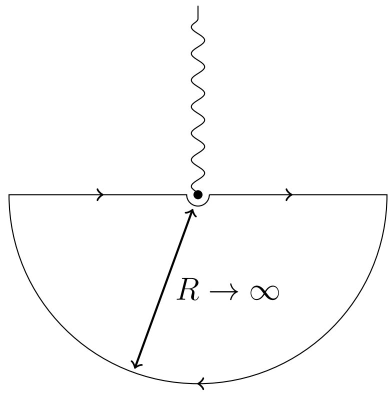


Figure 54. Choice of contour in complex v-plane.

Hence we have, for $\sigma > 0$

$$\begin{aligned}\tilde{p}^{(2)}_\omega(-\sigma) &= -A_\omega \int_0^\infty dv\, e^{-i\sigma v} \exp\left[i\frac{\omega}{\kappa}\log(-CDv)\right] \\ &= -A_\omega \int_0^\infty dv\, e^{-i\sigma v} \exp\left[i\frac{\omega}{\kappa}\left(\log(CDv)+i\pi\right)\right] \\ &= -A_\omega e^{-\omega\pi/\kappa} \int_{-\infty}^0 dv\, e^{i\sigma v} \exp\left[i\frac{\omega}{\kappa}\log(-CDv)\right] \\ &= -e^{-\omega\pi/\kappa}\tilde{p}^{(2)}_\omega(\sigma) \end{aligned} \tag{10.103}$$

therefore

$$|B_{\omega\sigma}| = e^{-\omega\pi/\kappa}|A^{(2)}_{\omega\sigma}| \tag{10.104}$$

We now return to using wavepackets, for which the corresponding result is

$$|B_{ij}| = e^{-\omega_i\pi/\kappa}|A^{(2)}_{ij}| \tag{10.105}$$

Now the normalization of $p^{(2)}$ gives (upon substituting in the decomposition of $p^{(2)}$ in terms of $f, \bar{f}$)

$$\begin{aligned}T_i^2 = (p_i^{(2)}, p_i^{(2)}) &= \sum_j \left(|A^{(2)}_{ij}|^2 - |B_{ij}|^2\right) = \left(e^{2\omega_i\pi/\kappa}-1\right)\sum_j |B_{ij}|^2 \\ &= \left(e^{2\omega_i\pi/\kappa}-1\right)(BB^\dagger)_{ii} \end{aligned} \tag{10.106}$$

hence the expected number of late time "out" particles of type i is

$$\langle 0|b_i^\dagger b_i|0\rangle = \frac{\Gamma_i}{(e^{2\omega_i\pi/\kappa}-1)} \tag{10.107}$$

where $\Gamma_i \equiv T_i^2$. As explained above, Γ_i is the "absorption cross-section" for the mode f_i (the "in" mode with the same profile as the "out mode" p_i), i.e., the fraction of this mode that is absorbed by the black hole. This result is exactly the spectrum of a blackbody at the *Hawking temperature*

$$T_H = \frac{\kappa}{2\pi} \tag{10.108}$$

This result shows that particle production is not just a transient effect during gravitational collapse: surprisingly, there is a steady flux of particles at late time.

The above argument can be generalized to other types of free field e.g. a massive scalar field, an electromagnetic field or a fermion field. In all cases, the result is the same: a blackbody spectrum at the Hawking temperature. One can also generalize to allow for non-spherically symmetric collapse, and collapse to a rotating or charged black hole. In the latter cases, one finds that the temperature is still given by (10.108) and the black hole preferentially emits particles with the same sign angular momentum or charge as itself, just like a rotating or charged blackbody.

For an astrophysical black hole, the Hawking temperature is tiny: for Schwarzschild we have

$$T_H = 6 \times 10^{-8} \frac{M_\odot}{M} \mathrm{K} \tag{10.109}$$

this is well below the temperature of the cosmic microwave background radiation (2.7K) so astrophysical black holes absorb much more radiation from the CMB than they emit in Hawking radiation, Tiny black holes, with $M \ll M_\odot$, could have a non-negligible temperature. But there is no convincing evidence for the existence of such small black holes.

Notice that T_H decreases with M. So Schwarzschild black holes have negative heat capacity.

10.8 Black hole thermodynamics

Hawking's discovery implies that a stationary black hole is a thermodynamic object with temperature T_H. Hence the zeroth law of black hole mechanics can be regarded as the zeroth law of thermodynamics applied to a black hole (the temperature is constant throughout a body in thermal equilibrium). The first law of black hole mechanics can now be written

$$dE = T_H dS_{BH} + \Omega_H dJ \tag{10.110}$$

where

$$S_{BH} = \frac{A}{4} \tag{10.111}$$

This is identical in form to the first law of thermodynamics provided we interpret S_{BH} as the entropy of the black hole: this is referred to as the *Bekenstein-Hawking entropy*. Reinstating units we have (k_B is Boltzmann's constant)

$$S_{BH} = \frac{c^3 k_B A}{4G\hbar} \tag{10.112}$$

The second law of black hole mechanics now states that S_{BH} is non-decreasing classically. But S_{BH} *does* decrease quantum mechanically by Hawking radiation: the black hole loses energy by emitting radiation and therefore gets smaller. However, this radiation itself has entropy and the *total* entropy $S_{\text{radiation}} + S_{BH}$ does *not* decrease. This is a special case of the *generalized second law* (due to Bekenstein) which states that the total entropy

$$S = S_{\text{matter}} + S_{BH} \tag{10.113}$$

is non-decreasing in any physical process. Evidence in favour of this law comes from the failure of various thought experiments aimed at violating it.

The result that black holes have entropy has several consequences. First, plugging in numbers reveals that the entropy of a Schwarzschild black hole with $M = M_\odot$ is $S_{BH} \sim 10^{77}$. This is many orders of magnitude greater than the entropy of the matter in the Sun: $S_\odot \sim 10^{58}$. Hence the entropy of the Universe would be *much* greater if all of the mass were in the form of black holes. So our Universe is in a very special (i.e. low entropy) state. This observation is due to Penrose.

Second, Hawking's result treats the gravitational field classically. But statistical physics tells us that entropy measures how many quantum microstates correspond to the same macroscopic configuration. So a black hole must have $N \sim \exp(A/4)$ quantum microstates. What are these? To answer this requires a quantum theory of gravity. A statistical physics derivation of $S_{BH} = A/4$ is a major goal of quantum gravity research. String theory has been successful in doing this for certain "supersymmetric" black holes. Such black holes are necessarily extreme ($\kappa = 0$) and include the extreme Reissner-Nordstrom solution.

10.9 Black hole evaporation

The energy of the Hawking radiation must come from the black hole itself. Hawking's calculation neglects the effect of the radiation on the spacetime geometry. An accurate calculation of this backreaction would involve quantum gravity. However, one can estimate the rate of mass loss by using Stefan's law for the rate of energy loss by a blackbody:

$$\frac{dE}{dt} \approx -\alpha A T^4 \tag{10.114}$$

where α is a constant and we approximate Γ_i by treating the black hole as a perfectly absorbing sphere of area A (roughly the black hole horizon area) in Minkowski spacetime. Plugging in $E = M$ with $A \propto M^2$ and $T \propto 1/M$ gives $dM/dt \propto -1/M^2$. Hence the black hole evaporates away completely in a time

$$\tau \sim M^3 \sim 10^{71} \left(\frac{M}{M_\odot} \right)^3 \text{sec} \tag{10.115}$$

This is a very crude calculation but it is expected to be a reasonable approximation at least until the size of the black hole becomes comparable to the Planck mass (1 in our units) when quantum gravity effects are expected to become important.

This process of *black hole evaporation* leads to the *information paradox*. Consider gravitational collapse of matter to form a black hole which then evaporates away completely, leaving thermal radiation. It should be possible to arrange that the collapsing matter is in a definite quantum state, i.e., a pure state rather than a density matrix. However, the final state is a mixed state, i.e., only describable in terms of a density matrix. Evolution from a pure state to a mixed state is impossible according to the usual unitary time evolution in quantum mechanics.

Another way of saying this is: information about the initial state appears to be permanently lost in black hole formation and evaporation. This is in contrast with, say, burning an encyclopaedia. In that case one could reproduce (in principle) the information in the encyclopaedia if one collected all of the radiation and ashes and studied them very carefully. Not so with Hawking radiation, which appears to be *exactly* thermal and hence contains no information about the initial state apart from its mass, angular momentum and charge.

Hawking interpreted this apparent paradox as indicating that quantum mechanics would need modifying in a full quantum theory of gravity. Most other physicists take a more conservative view that information is not really lost and that there are subtle correlations in the Hawking radiation which take a long time to appear but could, in principle, be used to reconstruct information about the initial state. However, this idea has run into trouble recently: if one assumes this, as well as several other cherished beliefs about black hole physics (e.g. nothing special happens at the event horizon, QFT in curved spacetime is a good description of the physics until the black hole reaches the Planck scale) then one runs into a contradiction (Almheiri, Marolf, Polchinski & Sully 2012).