–VERSION –July 24, 1994–

GR SHORT COURSE

William L. Burke

Contents

A2. Bibliography

Preface

In the beginning there was a two quarter course in General Relativity (GR) here at UC Santa Cruz. As time passed I kept expanding the mathematical preliminaries, adding examples from electrodynamics, mechanics, thermodynamics, and so on. After a while I realized that it was no longer a General Relativity course. I split the first quarter off into the course that became the book *Applied Differential Geometry.*

I tried to repackage the second quarter so that it would be possible to take it without having the mathematical preliminaries, provided one was willing to take the mathematics without derivations or examples. Few students try this, and I am not sure how successful that is.

It has been useful, nonetheless, to carve out the essential ideas of GR and to present them in a ten week package. These are my class notes for such a course. I have no intention for them to evolve into a book, but do intend for them to evolve into a better and better set of notes to supplement the class.

Some short, pregnant remarks in class require a careful and precise statement. The harried notetaker may scramble these or miss them all together. There are a lot of two liners in here that you can fruitfully expand into a term paper.

I am grateful for all the help that students in the class have given to me. Particular thanks to Ian Walker, Kathleen Wu, and David Reitzel.

I prefaced the 1992 version of these notes with the disclaimer that "These are my rough class notes, only lightly proofread. In the 1994 version I have filled a number of the logical and calculational potholes, and I have added many new figures.

Introduction

These notes present GR as a mature theory, much as electrodynamics is presented in Jackson, say. We are just going to leap in and do a bunch of interesting calculations.

Critical study

Of course GR is by no means as well tested as electrodynamics. We are going to neglect the critial analysis of how well GR has been verified, mainly because this class is constrained to just ten weeks duration. Look at Cliff Will, *Was Einstein Right?* for a discussion of the tests, so called, of Einstein's theory.

Philosophy

My ideas on the philosophical basis of gravitation come mainly from Hans Reichenbach, *The Philosophy of Space and Time*. Even though it is quite old and out dated, this is well worth reading. His discussion of the black hole is quite intelligent and completely wrong. A good way to appreciate the achievement of Eddington, Finkelstein, Kruskal, and Wheeler in straightening all this out is to see what the conventional wisdom was before their work. A later Reichenbachian treatment can be found in Lawrence Sklar, *Space, Time, and Spacetime*.

History

Here is a synopsis of the lineage of this course. I learned relativity from reading J. L. Synge, and from a course taught by Frank Estabrook after the untimely death of H. P. Robertson. Frank was a retreaded nuclear physicist who taught himself relativity also from Synge's books. Later I studied with Kip Thorne and wrote a thesis on gravitational radiation damping, but I have never considered myself in the John Wheeler school of relativity. Coming to UCSC with its good mathematics department, I proceeded to remedy my deficient math background, and this led to the present situation.

1. Gravity as a Classical Field

We will start from the idea of a classical field and its associated field energy.* To be positive definite, this field energy will be quadratic (or worse) in the fields. Were it not positive definite, then the stability of the vacuum (ground state) would be seriously in question. Classical field theories with a vector potential look like electromagnetism, and have like particles repelling each other. Theories with scalar or second rank tensor potentials have like particles attracting each other.

The sources of the fields will be the same geometrical objects as the potentials. In electrodynamics the source of the field is the 4-current. In a scalar theory it would be a scalar quantity; the trace of the stress energy tensor is the most likely quantity. In a tensor theory, the source of gravity would be the entire stress energy tensor. Since the stress tensor for a light signal is traceless, the observation that gravity bends light suggests that all (or at least most) of gravity is caused by a tensor field theory.

Now there is a paradox for theories with likes attracting. When two particles move together, the fields increase, ultimately doubling. The field energy is thus twice what the field energy was when the particles were separated. In electrodynamics, one has to do work to push the like particles together, and this can (and does) all balance. In an attractive theory, we get work out of the particles as they come together. How does this all work out?

The trouble is with the force law in the theory, which cannot be chosen arbitrarily, but must be consistent with the field equations and the expression for the energy in the field. It is not possible to choose this force law so that the force preserves the rest mass. That is, you cannot make the 4–force Lorentz perpendicular to the 4–velocity. Thus as you move the particles together they lose rest mass, and at a rate to make the energy balance work out. Now when particles lose mass, they change their sizes. So presto, despite your best intentions to be a classical physicist, you are stuck with a theory with measurements that do not follow Euclidean geometry. The two themes of this short course

* Our picture is that energy and momentum is sloshing around in the field, and coming out where there is a particular "twist" in the field lines indicating a source. Thus does the field exert forces on charges.

will be the failure of GR to behave as our Euclidean intuition expects, and the similar failure to act as our Newtonian intuition expects. The calculation of a consistent field energy and force law for a hypothetical scalar field is done in section 41 of *Applied Differential Geometry* (ADG).

[That section is in error, however, and you can in fact do that calculation easily with forms.]

This rescaling of masses will not mess up atomic theory only if it is universal. This shows up experimentally in the Eotvos experiment, and theoretically in the equivalence of active gravitational mass (what causes gravitational fields) and passive gravitational mass (what causes inertia). It is this universality that is the salient feature of gravitation. It results in its being an unshieldable force. Thus you cannot measure the gravitational force locally, because there are no unforced particles for comparison. This universality is what legitimizes the construction of a geometric theory.

[There is an interesting heuristic development of gravitation as a classical field in some dittoed but unpublished lecture notes of Feynman from a class in GR that he taught once in 1963.]

Geometric Gravity

Clifford in the late nineteenth century was the first to really try to explain gravity by geometry. This failed. After special relativity, with space and time married into spacetime, it became possible.

General Relativity is a theory which maps some part of the universe, perhaps all of it, into the geometric structure of a differentiable manifold with a Lorentz metric and other fields (electic, magnetic, etc.) defined on it. The metric in classical GR gives you a geometric structure that you use to model the physical behavior of clocks, measuring rods, and particles moving freely, under the influence of no forces except for gravity. Since the gravitational force is universal, we modify the notion of *free particle* to mean free except for possibly gravitational forces.

We cannot go into the fundamentals; we lack both the time and the mathematical skills. We should provide a detailed clock treatment based on quantum mechanics, and the matter should satisfy quantum field equations.

The interpretation of a situation in GR is fundamentally different from the usual interpretation of a classical field. The field itself alters the meaning of the solution. Until you know what the metric is, you do not know how to interpret the coordinates. This is very hard to make instinctive.

Field Equations

When starting Jackson, you are given the field equations. You might get some historical discussion. So too here. The field equations of GR relate the curvature tensor to the matter distribution. Nothing invariant can be formed from the first derivatives of the metric tensor. Coordinates, called Gaussian coordinates, can always be chosen so that the first derivatives vanish at any particular point. The curvature tensor is the only invariant that can be formed from the second derivatives. This is not obvious. The sources of the fields must be the matter, and the only invariant matter field is the stress energy tensor.

The stress energy tensor is a symmetric, two index tensor that generates conserved fluxes from symmetries. If we have a Killing vector (metric symmetry generator) k^μ, then we can form a vector from the stress energy tensor

$$j^\mu = T^{\mu\nu}\, k_\nu$$

and using the metric, turn this into a 1–form, and then using the star operator, turn this into a twisted 3–form. This twisted 3–form represents the flux of the conserved stuff associated with the symmetry. There is a balance law for this stuff; its exterior derivative is either zero, or the stuff disappears into known sinks.

There are ten symmetries of Lorentz spacetime, and so there are ten conserved quantities. In GR these are only locally conserved, because in general there are no Killing vectors. This local conservation is important to us because we cannot arbitrarily prescribe the motion of the sources of gravity. In the end, everything must be self consistent. For example, we cannot calculate the gravitational radiation caused by a particle that starts moving at some time, because that motion does not conserve linear momentum. The source that kicked the particle must be included. Because this conservation law is expressed in terms of the covariant derivative with respect to the metric, and the metric is what we are solving for, you can see the vicious, nonlinear circle that the field equations present to us.

We will proceed in this course to first study given symmetrical situations. We will introduce test particles into these solutions and pretend that they do not break the symmetry. This will allow us to develop a geometric, as opposed to a Newtonian, intuition. Then we will consider the general weak field situation, where the vicious circle does not arise. Out will go more Euclidean and Newtonian intuition. Then we go on to cosmology and gravitational collapse; replacing weak fields with symmetry. Without some falling from virtue, one can make little progress in GR.

You should also read the introduction to Chapter IX in ADG, and Section 59.

Stress–Energy Tensors

The geometric object called the stress–energy tensor, $T^{\mu\nu}$, plays as important a role in GR as the charge and current vector in electrodynamics. Here there are two ways to look at the stress–energy tensor, beyond its role as the source of the gravitational field. It is the array of local energy, momentum, and stress. These three concepts are related in that the flux of energy is momentum, and the flux of momentum is stress. The other way to look at the stress–energy tensor is that it is the generator that takes symmetries and produces conserved quantities. The stress–energy tensor is a two index symmetric tensor.

The basis for mechanics in flat spacetime are the ten conservation laws: for energy, momentum, angular momentum, and center–of–mass. These come from ten symmetries that we can describe by Killing vectors. These Killing vectors k^μ satisfy

$$\mathcal{L}_k \mathcal{G} = 0,$$

in components (ADG pg 143)

$$g_{\mu\nu,\sigma} k^\sigma + g_{\mu\sigma} k^\sigma{}_{,\nu} + g_{\nu\sigma} k^\sigma{}_{,\mu} = 0.$$

For a metric–derived covariant derivative we have

$$g_{\mu\nu;\sigma} = 0,$$

and this leads to

$$k_{\mu;\nu} + k_{\nu;\mu} = 0.$$

The expression with partial derivatives is more useful for actual calculation, while the expression with covariant derivatives is more useful in formal derivations. The flat spacetime Killing vectors are

$$\frac{\partial}{\partial t}, \frac{\partial}{\partial x}, \ldots, x\frac{\partial}{\partial y} - y\frac{\partial}{\partial x}, \ldots, x\frac{\partial}{\partial t} + t\frac{\partial}{\partial x}, \ldots.$$

The conservation laws constrain the stress–energy tensor

$$T^{\mu\nu}{}_{;\nu} = 0.$$

This contains total mechanical truth, more comprehensive than just $F = ma$. This correct treatment of mechanics is due to Cauchy (although not in this language, he lived before Killing, for example). We use the same constraint on the divergence of the stress–energy tensor in general relativity.

Example: Pressure–free matter, often called dust, is described by a 4–velocity λ^μ and a local, comoving density ρ. In the old days this was called proper density. The stress–energy tensor for dust is

$$T^{\mu\nu} = \rho\,\lambda^\mu \lambda^\nu.$$

Its covariant divergence gives us

$$T^{\mu\nu}{}_{;\nu} = \lambda^\mu (\rho\lambda^\nu)_{;\nu} + \rho\lambda^\mu{}_{;\nu}\lambda^\nu.$$

The derivative of the 4–velocity along the path is the 4–acceleration, and it is Lorentz orthogonal to the 4–velocity because the 4–velocity is normalized. Thus we can extract information from this by dotting a 4–velocity through it. This gives us

$$(\rho\lambda^\nu)_{;\nu} = 0,$$

which is the law of conservation of matter (as opposed to energy), and then we must also have the 4–acceleration vanishing. Pressure–free matter streams along geodesics.

Electrodynamics as a classical field

Let us develop the general properties of a spin–1 classical field, i.e. one described by a potential which is a vector, and having a field quantity

$$F_{\mu\nu} = A_{\nu,\mu} - A_{\mu,\nu}.$$

Every classical field must have a stress–energy tensor, and to keep the energy positive definite (necessary for stability) it should be quadratic. The only two quadratic terms possible are

$$F^{\mu\alpha} F_{\alpha}^{\cdot\nu} = F_{\alpha}^{\cdot\nu} F^{\mu\alpha} = F^{\nu\alpha} F_{\alpha}^{\cdot\mu}$$

and

$$F^{\alpha\beta} F_{\alpha\beta}\, g^{\mu\nu},$$

both symmetric.

What are their divergences?

$$(F^{\mu\alpha} F_{\alpha}^{\cdot\nu})_{;\nu} = F^{\mu\alpha} F_{\alpha\ ;\nu}^{\cdot\nu} + F^{\mu\alpha}{}_{;\nu} F_{\alpha}^{\cdot\nu}$$

Note the danger that results if you leave out the little dots. These placeholders are important when dealing with antisymmetric tensors. You don't know the sign of F_{β}^{α}. We have also

$$(F^{\alpha\beta} F_{\alpha\beta})_{;\nu} g^{\mu\nu} = 2F^{\alpha\beta}{}_{;\mu} F_{\alpha\beta}.$$

Because of the potential, we have

$$F_{\mu\nu;\sigma} + F_{\nu\sigma;\mu} + F_{\sigma\mu;\nu} = 0,$$

and we want to use this to kill off the garbage in the divergence of the stress–energy tensor. By garbage I mean those terms that are not local, i.e. that involve derivatives.

Looking over the terms with one free index contracted with an external F, we see that there will be two terms with the free index inside, and one with it outside the differentiation. This leads us to write one term

$$F_{\mu\alpha;\nu} F^{\alpha\nu} = \frac{1}{2} F_{\mu\alpha;\beta} F^{\alpha\beta} + \frac{1}{2} F_{\beta\mu;\alpha} F^{\alpha\beta}$$

and so we take $\frac{1}{4}$ of the other term to find

$$(F^{\mu\alpha} F_{\alpha}^{\cdot\nu} + \frac{1}{4} F^{\alpha\beta} F_{\alpha\beta}\, g^{\mu\nu})_{;\nu} = F^{\mu\alpha} F_{\alpha\ ;\nu}^{\cdot\nu}.$$

Our only hope for a balance law that is local is to have a field equation

$$F^{\alpha\nu}{}_{;\nu} = 4\pi j^{\alpha}.$$

This fixes the force law, and tells us the rate that momentum, for example, comes out of the electromagnetic field and onto its sources with a local density

$$F^{\mu\alpha} j_{\alpha}.$$

2. Orbit Problem

[This follows sections 60 to 63 in ADG.]

We start with a study of the properties of a given gravitational situation. Later we will see how the gravitational field is itself generated by the matter. This section is the equivalent of studying the motion of charged particles in electric and magnetic fields, with charges so small that they do not significantly modify the fields. The gravitational problem is considerably more complicated than the electrodynamic one. While a given electric field can be described as a function of space, the gravitational field contains within itself the definition of the space. Thus we must find simultaneously both the motion of the particles and the structure of the spacetime.

Geodesics

The motion of uncharged test particles in a given gravitational field is along the geodesics of the Riemannian metric of the spacetime. This is not an assumption, but rather an extremely delicate theorem. It is easy to prove for bodies whose mass goes to zero faster than their size, but it is also true in more extreme cases, so that even tiny black holes follow geodesics in the background spacetime.

A curve is a geodesic with special affine parameter u if the curve

$$u \mapsto \gamma(u),$$

satisfies

$$\nabla_{\dot\gamma}\,\dot\gamma = 0.$$

[See Appendix A1 on tensor notation if this is not familiar.]

The most effective way to generate the geodesic equations for a given spacetime is to use the Lagrangian

$$E = \int g_{ij}\,\dot X^i\,\dot X^j\,du,$$

to generate the Euler-Lagrange equations. This variational principle generates the curves with special affine parametrization. The variational principle with a square root in it (least time) would allow all possible parametrizations. These are more complicated equations, and have difficulty with light signals, whose tangent vectors cannot be normalized.

Symmetries and Conserved Quantities

A Killing vector k^μ satisfies

$$k_{i;j} + k_{j;i} = 0.$$

This is manifestly covariant, but computationally involves a lot of cancellation. More useful for computations is the form

$$g_{\mu\nu,\sigma} k^\sigma + g_{\sigma\nu} k^\sigma{}_{,\mu} + g_{\sigma\mu} k^\sigma{}_{,\nu}.$$

Here comma denotes partial derivative.

Such a Killing vector generates a conserved quantity L

$$L = k \cdot \dot{\gamma}.$$

Schwarzschild orbit problem

The simplest nontrivial gravitational field is the spherically symmetric solution outside of a spherial mass. You might have expected us to start with a plane gravitational field, which you would expect to be even simpler. Alas, there is no such thing as a plane gravitational field. It would be of infinite extent, involve infinite mass, and not be self consistent. The closest we can come to a plane solution will be to look at the outer regions of the spherical solution. The simplest part of the solution there can be recognized as just flat spacetime in accelerating coordinates; this is the Principle of Equivalence.

For the present we just take the following spacetime as a given:

$$\mathcal{G} = -(1 - \frac{2m}{r})\, dt^2 + dr^2/(1 - \frac{2m}{r}) + r^2(d\theta^2 + \sin^2\theta\, d\phi^2).$$

This representation is only valid for $r > 2m$.

This spacetime has three Killing vectors representing spherical symmetry. The orbits can be studied in the equatorial plane, $\theta = \pi/2$. There we have a Killing vector

$$\frac{\partial}{\partial \phi},$$

and if we use λ^μ for the tangent to the curve, normalized so that

$$\lambda \cdot \lambda = -1$$

for particles with rest mass and

$$\lambda \cdot \lambda = 0$$

for light signals, then we have that

$$L = \lambda_\phi$$

is a constant of the motion, which we will call angular momentum (really angular momentum per unit mass).

There is also a time symmetry, and so we have another conserved quantity

$$E = \lambda_t.$$

We have a third algebraic relation

$$\mathcal{G}(\lambda, \lambda) = -\kappa,$$

where $\kappa = 1$ for particles with rest mass and zero for light signals.

Because we have three algebraic relations for the three components of the velocity, we will be able to find most of the interesting features without even writing down the Euler-Lagrange equations. If we collect up the above three equations and discard a positive term involving the square of the radial velocity, we find that the orbit must satisfy an inequality

$$L^2 - r^2(E^2 - \kappa) - 2mr\kappa \leq \frac{2mL^2}{r}.$$

Light rays

For light rays the angular momentum and the energy are not independent, and only their ratio matters. The inequality is

$$L^2 - r^2 E^2 \leq \frac{2mL^2}{r},$$

or, better

$$\frac{r^3}{(r - 2m)} \geq (L/E)^2.$$

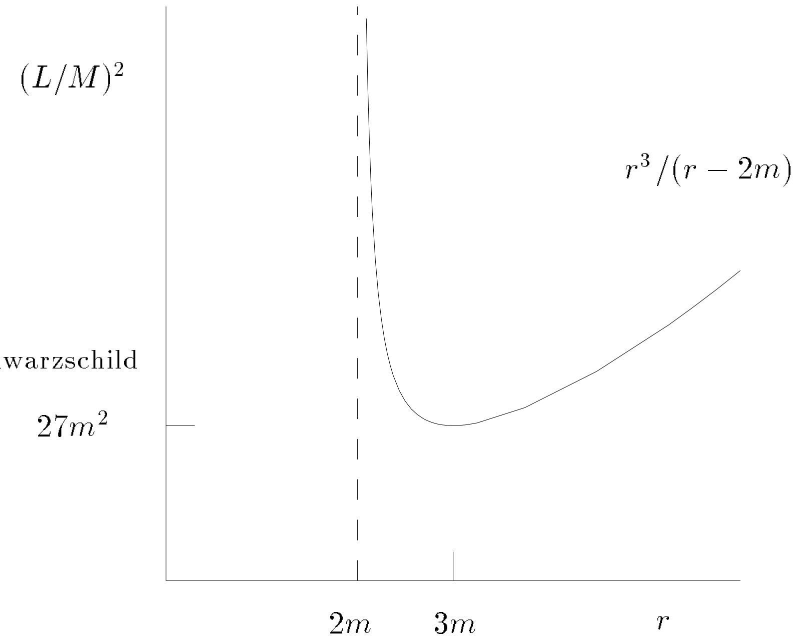


Figure 2-1. Light rays in the Schwarzschild metric.

From a sketch of this you can see that an incoming light ray turns when $\lambda^r = 0$, when the inequality is in fact an equality.

Light rays with $(L/E) < 27m^2$ will not have any turning point in the region that we can discuss and in fact, go down the "black hole".

For large values of r the spacetime is nearly Euclidean. There the tangent vector to the critical light ray, with

$$\frac{L}{E}^{2} = 27\,m^2$$

is given by

$$\lambda = \frac{\partial}{\partial t} + \frac{\sqrt{27}}{r^2}\frac{\partial}{\partial \phi} + \frac{\partial}{\partial r}.$$

The angle between this vector and the radial vector is given by

$$\theta = \sqrt{27}m/r,$$

and this says that the cross section for this spacetime to swallow up incoming light is $27\pi m^2$.

Particles with nonzero rest mass

We discuss this case by looking separately at the two functions of r

$$L^2 - r^2(E^2-1) - 2mr$$

and

$$2mL^2/r.$$

A turning point occurs when these curves intersect.

If the intersection is a double point, a point of tangency, then this is a circular orbit (inner and outer turning points coincident). For circular orbits we have

$$\begin{aligned} E &= \frac{(r-2m)}{\sqrt{r(r-3m)}}, \\ L &= a\sqrt{m/(a-3m)}. \end{aligned}$$

These circular orbits are stable for large r. When there is a triple contact, at $r = 6m$, the circular orbits go unstable. Looking at the light ray orbits, you can see that they have an unstable circular orbit at $r = 3m$.

3. Light Deflection

[This follows sections 64 and 65 in ADG. Go there for more information.]

One of the earliest confirmations of GR was the observation that light was deflected by a body. For a mass m and impact parameter $b \gg m$

$$\theta = 4m/b,$$

This was an important confirmation of GR because the Newtonian "back of the envelope" argument gives a deflection which is half that amount.

You cannot derive this deflection angle from the conservation laws that we gave earlier, you need the actual orbital shape in space. One way to derive this would be to start from the equation

$$\frac{d\phi}{dr} = \frac{\lambda^\phi}{\lambda^r}.$$

[The straightforward calculation can be found in Weinberg, and a clever treatment in Robertson and Noonan.]

Here I want to pursue an approach that is more involved, not any easier computationally, but which conveys some physical insight into the general properties of spin-2 classical field theories. It will shed some light on *why* the deflection is twice what you would expect.

Spacetime with two metrics

Instead of getting the correct equation and then solving it approximately, we will set up an approximate situation right from the start, considering the far out reaches of the Schwarzschild spacetime as a perturbation away from flat Minkowski spacetime.

We start with the general machinery for talking about a spacetime with two metrics defined on it. One is the true metric, and one will be a simpler metric that we will consider to be the unperturbed metric. For our problem here this simpler metric will be flat spacetime.

Write the curved spacetime metric as

$$*g_{\mu\nu} = g_{\mu\nu} + h_{\mu\nu},$$

where $*g_{\mu\nu}$ is the exact metric, and $g_{\mu\nu}$ is the background metric. Each of these metric tensors leads to a connection and a covariant

[This is not the Hodge star operator.]

derivative. Recall that the difference between two connections is a tensor. This tensor here plays the role of the gravitational field tensor. The geodesic equation for a light ray can be written

$$\sigma^i{}_{;j}\sigma^j + \Lambda^i{}_{jk}\,\sigma^j\sigma^k = 0,$$

where the semicolon denotes the covariant derivative in the background spacetime, and the field tensor is given approximately (linear approximation) by

$$\Lambda^i{}_{jk} = \frac{1}{2}\,g^{im}(h_{mj;k} + h_{mk;j} - h_{kj;m}).$$

This is the gravitational analog of the electromagnetic Lorentz force law

$$\sigma^i{}_{;j}\sigma^j + (e/m)F^i{}_j\,\sigma^j = 0.$$

Light deflection

The approximate form of the Schwarzscild spacetime metric is

$$h_{\mu\nu} = (2m/r)(dt^2 + dr^2).$$

In rectangular coordinates the unperturbed light ray is the curve

$$u \mapsto (t, x, y, z) = (u, u, b, 0).$$

Two components of the gravitational force contribute here. The $\Lambda^y{}_{tt}$ term, which is the analog of the Coulomb force, and couples with the t component of the particles 4-velocity, and the $\Lambda^y{}_{xx}$ term, which couples to the square of the x component of the particles velocity. This is like a magnetism-magnetism force, with no analog in ordinary electrodynamics. For a light ray, this double magnetic force is comparable with the Coulomb force, and this changes the answer from the Newtonian expectation. This double magnetism does not have the handed-symmetry that we associate with ordinary magnetism. It has the symmetry of the other velocity squared forces, thus it acts like pressure forces rather than like gyroscopic forces.

The path is an approximation of a simple path in rectangular coordinates, and the coordinate basis vectors there are orthonormal, so we choose to work in rectangular coordinates. The easiest way to transform $h_\mu\nu$ is to use

$$r^2 = x^2 + y^2 + z^2,$$
$$dr = \frac{x\,dx + y\,dy + z\,dz}{r}.$$

Thus

$$h_{\mu\nu} = (2m/r)\left(dt^2 + \frac{x^2\,dx^2}{r^2} + \frac{xy\,(dx\,dy + dy\,dx)}{r^2} + \dots\right),$$

and the nonzero components can be easily read off from the expansion of this.

The field components needed are

$$\begin{aligned}
\Lambda^y{}_{tt} &= my/r^3,\\
\Lambda^y{}_{tx} &= 0,\\
\Lambda^y{}_{xx} &= 2my/r^3 - 3mx^2\,y/r^5.
\end{aligned}$$

The velocity vector has unit components, approximately, so the deflection angle (for small angles) is

$$\theta = \int_{-\infty}^{\infty}\left(\frac{3mb}{r^3} - \frac{3mx^2 b}{r^5}\right)\,dx = \frac{4m}{b}.$$

It is important to realize here that the incoming and outgoing light signal vectors are being compared at large r where spacetime is nearly Euclidean.

Gravitational mirage

The gravitational deflection of light is well observed and an important astrophysical effect. It allows us to estimate the total mass of a distant object. This is sometimes called a "gravitational lens", although this is quite a misnomer. The optics of the situation do not resemble those of any lens you are familiar with. The most familiar situation is the mirage. We have very little experience with light deflectors with little or no symmetry.

One interesting result is that the $1/r$ deflection law lets us use two dimensional potential theory. From this we can conclude that light passing through a ring of material is not deflected at all, and that light passing outside will be deflected as if all the mass were in a point at the center.

"There is no way to be virtuous and still do things in the real world."

4. Weak Gravity

The preceeding section gave us the approximate gravitational field around a point mass. The gravitational force was radial, and given in polar coordinates by

$$(m/r^2)\left(-dt^2 + 2dr^2 - 2r^2\,d\theta^2 - 2r^2\sin^2\theta\,d\phi^2\right)\cdot(\lambda\otimes\lambda)\frac{\partial}{\partial r}$$

where λ^μ is the particle 4-velocity, and the centered dot is tensor-product evaluation. The term

$$-(m/r^2)\,dt^2$$

is the Newtonian force. The other terms are needed to describe fast moving particles. What do I mean by fast? Since this is a linear approximation, there are unknown corrections to the Newtonian term of order (m/r), and so if

$$v^2 \gg m/r$$

then these velocity dependent corrections will dominate the nonlinear corrections. This means that the velocities must be greater than the escape velocity. You can't use this method to calculate the magnetic analog of the gravitational force on a planet, for example.

Linear Gravity

We develop this using two metrics. Similar material can be found in MTW Chapter 18.

We seek solutions in the form

$$*g_{\mu\nu} \sim g_{\mu\nu} + \epsilon h_{\mu\nu} + \ldots$$

[I have explicitly included the small parameter here since there are many terms of different orders.]

the same as we did in the light deflection calculation. Here $g_{\mu\nu}$ is a known solution. The solution $*g_{\mu\nu}$ is a perturbation of that known solution. The known solution might be flat space, a Robertson–Walker universe, Schwarzschild spacetime, etc. The parameter ϵ will probably come from some of the terms in the stress energy tensor $T^{\mu\nu}$. For this section, we perturb flat space, although not necessarily in rectilinear coordinates.

The first step is to calculate the Riemann tensor. For the moment, go to coordinates where $\Gamma^\mu_{\nu\sigma} = 0$, and work only to first order in ϵ. We have no symmetries for the perturbed space, and there are really no tricks that will help us shortcut this calculation. We use the longhand tensor formula to calculate these.

$$*\Gamma^\mu_{\nu\sigma} \sim \frac{\epsilon}{2} g^{\mu\alpha}\left[h_{\alpha\nu;\sigma} + h_{\alpha\sigma;\nu} - h_{\nu\sigma;\alpha}\right]$$

$$*R^\mu_{\nu\sigma\tau} = *\Gamma^\mu_{\nu\tau,\sigma} - *\Gamma^\mu_{\nu\sigma,\tau} + O(\epsilon^2)$$

$$*R^\mu_{\nu\sigma\tau} \sim \frac{\epsilon}{2} g^{\mu\alpha}\left[h_{\alpha\nu;\tau\sigma} + h_{\alpha\tau;\nu\sigma} - h_{\nu\tau;\alpha\sigma} - h_{\alpha\nu;\sigma\tau} - h_{\alpha\sigma;\nu\tau} + h_{\nu\sigma;\alpha\tau}\right]$$

The first and fourth terms above cancel. The Ricci tensor

$$R_{\mu\nu} \equiv R^\beta_{\cdot\mu\beta\nu}$$

$$*R_{\mu\nu} \sim \frac{\epsilon}{2} g^{\alpha\beta}\left[-h_{\mu\nu;\alpha\beta} - h_{\alpha\beta;\mu\nu} + h_{\alpha\mu;\nu\beta} + h_{\alpha\nu;\mu\beta}\right]$$

$$*R \sim \epsilon g^{\nu\beta} g^{\gamma\delta}\left[-h_{\alpha\beta;\gamma\delta} + h_{\alpha\gamma;\beta\delta}\right]$$

The Einstein tensor associated with this is a big mess. It is simplified by introducing the new variable $\bar{h}_{\mu\nu}$:

$$\bar{h}_{\mu\nu} \equiv h_{\mu\nu} - \frac{1}{2} h_{\alpha\beta}\, g^{\alpha\beta}\, g_{\mu\nu}.$$

This is similar to the operation in which you remove the trace, but with a factor of a half rather than a quarter (four dimensions!) means that instead it reverses the trace. Because it reverses the trace, the same operation restores h from $\bar{h}$.

With this definition, the Einstein tensor simplifies to

$$*G_{\mu\nu} = \frac{\epsilon}{2}\left[-\bar{h}_{\mu\nu;\alpha\alpha} + \bar{h}_{\alpha\mu;\nu\alpha} + \bar{h}_{\alpha\nu;\mu\alpha} - g_{\mu\nu}\, \bar{h}_{\alpha\beta;\alpha\beta}\right].$$

Here I am using the shorthand which leaves out the obvious factors of $g^{\alpha\beta}$ needed to make the repeated indices at the same level make sense. Thus

$$\bar{h}_{\alpha\mu,\nu\alpha} = \bar{h}_{\alpha\mu,\nu\beta}\, g^{\alpha\beta}.$$

The Einstein Field Equations are then

$$*G_{\mu\nu} = 8\pi\ *T_{\mu\nu}.$$

This should look familiar to you. We are at the point in electrodynamics just before Lorentz gauge is introduced. Maxwell's Equations in terms of a potential are

$$A_{\mu;\alpha\alpha} - A_{\alpha;\alpha\mu} = 4\pi\, j_\mu .$$

Actually this is half of them, the other half are satisfied identically by the potential. Thus we are making gravity look like just another classical field.

The next step, then, is to choose a gauge condition. We can do this only if there is a gauge invariance. The above system of equations (these are ten coupled p.d.e.'s) has the invariance

$$h_{\mu\nu} \mapsto h_{\mu\nu} + V_{\mu;\nu} + V_{\nu;\mu} .$$

where V_μ is any vector field. This transformation takes solutions into other solutions. It reflects the coordinate invariance of the field equations. A more extensive treatment of this can be found in MTW, Box 18.2. Note that this invariance is on h, not $\hbar$

We can use the gauge invariance to impose the condition

$$\hbar_{\mu\alpha;\alpha} = 0.$$

This is usually called "Lorentz Gauge" in analogy with the situation in electrodynamics. This gives us a wave equation

$$\hbar_{\mu\nu;\alpha\alpha} = -16\pi\, T_{\mu\nu} .$$

If we take the above wave equation and take the divergence of both sides, and recall that we are in flat space (but possibly in curvilinear coordinates), so that covariant derivatives commute, then the Lorentz gauge condition gives us the consistency condition on the stress–energy tensor

$$T_{\mu\alpha;\alpha} = 0.$$

Raising indices with the unperturbed metric, remember, we can write this in the better form

$$T^{\mu\alpha}{}_{;\alpha} = 0,$$

with no funny conventions to possibly mislead.

Linearization Paradox

This leads to a consistency problem. It is common to all linearizations. The linear perturbations have no way to interact. Thus our gravitational test masses will not attract one another.

The resolution of this paradox requires a more sophisticated understanding of perturbation theory than you have been exposed to. The straightforward linearization present above may not be valid for long times, particularly, for times as long as $t \sim 1/epsilon$. This invalidity occurs because terms like ϵt will appear in the solution. These terms ruin the long time validity of the expansion.

The resolution of the paradox lies in the elimination of these terms, called secular terms. You may have come across these ideas in mechanics under the name *method of averaging*. The only way to secure long time validity for the linearization is to let the test masses interact, even though there is no formal need for it.

5. Gravitational Waves

We use the linear theory developed in the preceeding section to study homogeneous solutions, free gravitational fields without sources in the region of study. These are the simplest solutions to study, and yet many of the general problems, in particular coordinate covariance appear here.

The approximate calculation of the last section produced a class of spacetimes. Now we study those spacetimes using exactly the same techniques as we used in the orbit problem: symmetries, light signals, and particle orbits.

Let us start by getting a geometric grasp on the objects that we are going to talk about. The next two diagrams show two dimensional slices through two different spacetimes. One of these has a gravitational wave passing through it, perpendicular to the paper and the other is just flat spacetime in wavy coordinates. You might glance at the figures to see if you can spot the difference. The most significant feature of the two spacetimes is their spatial symmetry.

In these figures the vertical lines would be world lines of geodesics. You can see this from Huyghen's construction, which was covered in some detail in my book *Spacetime Geometry Cosmology*. You can think of the vertical lines, then, as the worldlines of a family of dust motes, and it is natural to ask what it would be like for people on the dust motes to be doing astronomy.

To do astronomy from these dust motes, one wants to trace out the trajectories of light signals. These lines must be everywhere tangent to the light cones of the metric figures. Since these spacetimes are symmetric, the light signal worldlines are also invariant under spatial translations (horizontal). To measure the doppler shift between observers on different dust motes, it is sufficient to look at the vertical intervals between two of these horizontally translated world lines. Why vertical? Because that is the world line of the clock used to measure the doppler shift.

You can see that there is a doppler shift that can be measured in the first spacetime. There is not one in the second spacetime. The first

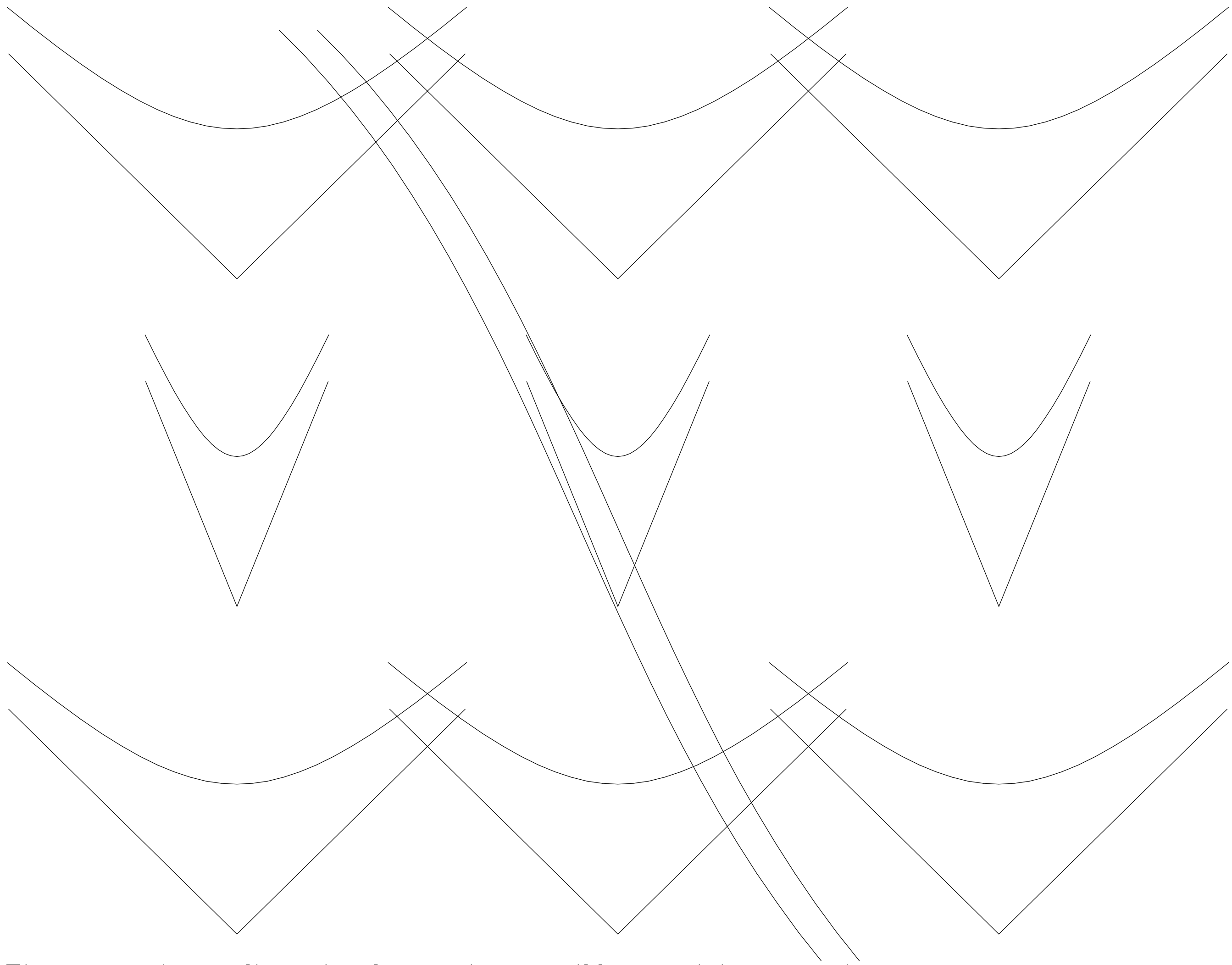


Figure 5-1. A two dimensional spacetime possibly containing a gravitational wave.

spacetime is a real gravitational wave; the second is just flat spacetime in coordinates that have been squashed vertically in the middle of the diagram.

The doppler shift observable in the first spacetime is an example of the non–Newtonian behavior of general relativity. Here we have two observers which never accelerate (they would feel it if they did). For a while they are not moving apart (no doppler shift). Then, without any acceleration, there is suddenly a doppler shift. This cannot happen in Newtonian physics. This gives us an operational means for distinguishing this spacetime from flat Minkowski spacetime.

Plane Waves

Here is an analytic treatment of the four dimensional version of the

Figure 5-2. Another candidate spacetime for a gravitational wave.

above geometric situation. We are going to find the plane–wave homogeneous solutions to the linearized equations. The exact metric will be flat spacetime plus the perturbation

$$*g_{\mu\nu} = g_{\mu\nu} + h_{\mu\nu},$$

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2} h\, g_{\mu\nu},$$

writing h for the trace, with the wave equation

$$\Box \bar{h}_{\mu\nu} = 0$$

[No sources.]

and the gauge condition

$$\bar{h}_{\mu\alpha;\beta}\, g^{\alpha\beta} = 0.$$

We use the shorthand for the gauge condition

$$\bar{h}_{\mu\alpha;\alpha} = 0.$$

These solutions have a gauge invariance

$$h_{\mu\nu} \sim h_{\mu\nu} + V_{\mu;\nu} + V_{\nu;\mu},$$

for any vector V. This is an equivalence relation on the solutions.

All directions are equivalent; we will take our waves to be going in the z–direction. The wave equation has solutions

$$\bar{h}_{\mu\nu} = e_{\mu\nu}\, F(z-t).$$

Here $e_{\mu\nu}$ is a constant tensor, and F is an arbitrary function of one variable giving the wave amplitude profile. So far we have ten degrees of freedom in the polarization. Now we have to reduce these to those that are consistent with the gauge condition, and distinct under gauge transformations. Both of these are miserably hard questions in the full nonlinear theory. It is good to meet them on simple ground first.

First we go through the cases for the gauge condition. Note that both of these are linear relations.

Suppose $\bar{h}_{tt} \neq 0$. The gauge condition gives us two non–trivial equations

$$-\bar{h}_{tt,t} + \bar{h}_{tz,z} = 0$$

$$-\bar{h}_{zt,t} + \bar{h}_{zz,z} = 0$$

The only solution for arbitrary F is the polarization tensor with non–zero zz, zt and tt components:

$$e_{(1)} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix}$$

If $\bar{h}_{tt} = 0$, then the bottom block all has to vanish.

The off diagonal two–by–two blocks give us two more candidates

$$e_{(2)} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \end{pmatrix}, e_{(3)} = \begin{pmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}$$

Finally, the fully transverse polarizations all satisfy the gauge condition.

$$e_{(4)} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, e_{(5)} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, e_{(6)} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Now we use the equivalence relation to discard duplicate solutions from this list of six. Suppose we pick the vector V in the gauge transformation so that we have

$$V_x = G(z - t),$$
$$G'(u) = F(u).$$

all other components zero. This leads to an $h_{\mu\nu}$ field

$$h_{\mu\nu} = \begin{pmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 \end{pmatrix}$$

This is traceless, and so it is the same expression for $\bar{h}$. Thus two of our polarizations, $e_{(2)}$ and $e_{(3)}$ are gauge equivalent to zero. The last two possibilities for V lead to

$$e_{(1)} \sim e_{(4)} + e_{(5)}$$

$$e_{(1)} \sim -e_{(4)} - e_{(5)},$$

that is,

$$e_{(1)} \sim 0$$

$$e_{(4)} + e_{(5)} \sim 0.$$

The only two polarizations for gravitational waves are thus

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \qquad \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

These two are called the transverse, traceless representation. One could expect that there are just two independent polarizations; the spin of a massless particle (graviton) can only be aligned with or against the motion.

6. THE Linearly Polarized Gravitational Wave

The previous section showed that there are only two distinct polarizations for a gravitational wave. If you rotate coordinates by 45°, you will transform these into each other. Furthermore, any linear combination of them is just a rotation by some angle. Thus there is a unique linearly polarized wave. The spacetime is

$$\mathcal{G} = -dt^2 + dz^2 + [1 + F(z - t)]\, dx^2 + [1 - F(z - t)]\, dy^2.$$

where F is an arbitrary function giving the dimensionless amplitude of the wave. We now proceed to study this like any other spacetime. The effects of gravity show up in the measurement of lengths as well as in forces. This is how gravitation differs from other classical fields.

Symmetries

The above spacetime has three Killing vectors

$$\frac{\partial}{\partial x}, \quad \frac{\partial}{\partial y}, \quad \frac{\partial}{\partial z} + \frac{\partial}{\partial t}.$$

Only at this point are we justified in calling it a *plane* wave. We have just discovered that it has the symmetry of a plane: two Killing vectors that commute. From these symmetries we find three conserved quantities:

$$P = \lambda_x,$$

$$Q = \lambda_y,$$

and

$$U = \lambda_z + \lambda_t.$$

Together with the normalization condition, this gives us four equations for the four components of the 4–velocity. We can thus argue in a manner similar to the black–hole orbit calculation.

A Cloud of Dust

Suppose that we have a test particle at rest at some initial time. Then the invariants are

$$P = 0, \quad Q = 0, \quad U = 1.$$

The components of the 4–velocity must satisfy

$$\lambda_x = \lambda_y = 0, \quad \lambda_z + \lambda_t = -1,$$

and

$$(\lambda_z)^2 - (\lambda_t)^2 = -1.$$

From this it follows that for all time

$$\lambda = \frac{\partial}{\partial t}.$$

In these coordinates, a particle at rest stays at rest.

Does this tell us that a gravitational wave does not exert any forces on particles? NO. It only tells you what these coordinates mean. In this representation, the coordinates are the Lagrangian coordinates for a cloud of free particles, a dust cloud.

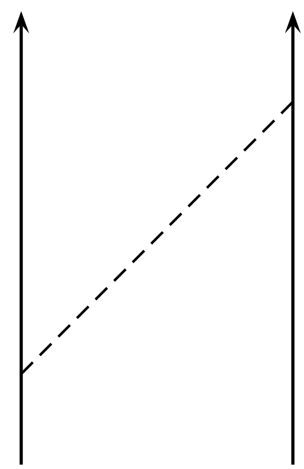
Astronomy from Dust Motes

Suppose there are observers on two of these dust motes with world lines

$$I : s \mapsto (t, x, y, z) = (s, 0, 0, 0),$$

$$II : s \mapsto (t, x, y, z) = (s, L, 0, 0),$$

where L is some constant length. Let I send a photon to II. These photons are subject to the same conservation laws. Their 4–velocities are normalized not to -1 but to 0.



For this photon σ_x will be constant. This is related to the wave number k observed for the photon

$$k = \sqrt{g^{xx}}\sigma_x.$$

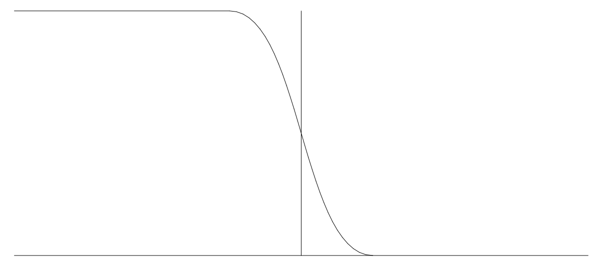
This "half–way raised index" is called the physical component of σ in MTW. It really only makes sense with a diagonal metric. Because F is small, we have

$$k = (1 - \frac{F}{2})\sigma_x.$$

where the amplitude F is evaluated at the point where the photon is being measured. Thus the ratio

$$\frac{k_{EMIT}}{k_{RECEIVE}} \sim 1 + \frac{1}{2}[F(-t_{II}) - F(-t_I)].$$

Figure 6-1. Pulse shape for our wave.



Example: Suppose $F(u)$ looks like Figure 6-2, with a constant value of A before $u = 0$, and a smooth drop to zero for times after. The fall time of the amplitude will be assumed to be small compared with the spacing L between the dust motes.

Let I continually send out photons of constant wavelength. This observer need be no more than an atom. What does II measure? The geometry is sketched in Figure 6-1. The wave edge hits I at time t_I and hits II at the later time $t_{II} = t_I + L$.

Well, before t_I there is no frequency shift because the wave edge hasn't arrived. During the time that the pulse edge has hit I but not II, the observer sees a wave–number ratio

$$\frac{k_{EMIT}}{k_{RECEIVE}} \sim 1 + \frac{A}{2}.$$

After the wave has passed over both dust motes, the frequency ratio is again unity.

A good astronomer would interpret the frequency shift as a velocity

$$v \sim \frac{k_E - k_R}{k} \sim \frac{A}{2},$$

with positive velocity being recession. Since the shift is observed for a time L, this velocity would be interpreted as a change in the distance of an amount

$$\Delta L \sim \frac{AL}{2}.$$

Observer I would not consider it amiss to send a message to II saying something like "I say you move". But the situation is completely symmetric. EACH FREELY FALLING OBSERVER SEES THE OTHER MOVE. This observation is inconsistent with the idea of a Galilean reference frame. It shows one way in which GR violates our usual intuition. Note that we cannot answer the question as to whether this frequency shift is caused by actual motion. There is no intrinsic frame to measure motion against. There could be an arbitrary number of these observers in a line transverse to the wave. Each would feel at rest, and observe the others move. So you might be tempted to say that none move, but that space is created in between all of them. While comforting, that too, goes beyond what is actually there. What is there is just $g_{\mu\nu}$ and its consequences.

Polarization

One can look at dust motes in other directions. There is no effect if you look in the z–direction. From our conservation law

$$\sigma_z + \sigma_t = \text{const.}$$

and because the 4–velocity is a null vector, the metric is diagonal, and the zz and tt components of the metric are unaffected by the wave

$$\sigma_z = \sigma_t.$$

Thus σ_z must be constant.

In the reverse x–direction one sees the same effect. In both y–directions one would see the opposite effect. This double sign change is the behavior of quadrupole symmetry. This is why one says that the graviton is a spin–2 particle. Contrast this with the spin–1 behavior of an electromagnetic wave.

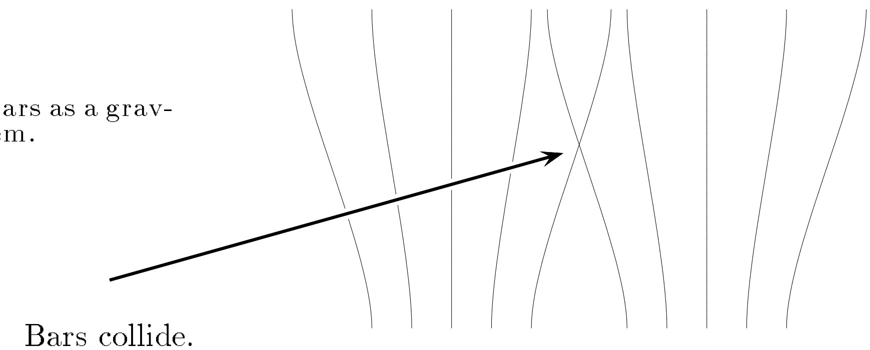
Elastic Bodies

Suppose there is an elastic bar floating out in our dust cloud. What happens as the wave goes by?

If the bar is soft, then inertial forces dominate. The bar has natural frequencies below the wave frequencies. This bar behaves just like a dust cloud.

If the bar is stiff, that is, its natural frequencies are all higher than the frequencies found in the wave, then it keeps its proper length constant. For a stiff bar inertial forces can be ignored. The bar will move relative to the local dust cloud. Two such bars close to one another could collide as the wave goes by, since their centers will follow the local dust cloud. An observer riding on the end of a stiff bar would measure an acceleration as the wave hit. His worldline would be curved relative to a geodesic worldline.

Figure 6-2. Two stiff bars as a gravitational wave hits them.



If the bar is soft, then inertial forces dominate. The bar has natural frequencies below the wave frequencies. This behaves just like a dust cloud.

For the intermediate cases we can write a low–velocity Lagrangian for the bar as follows. This would be useful if you were designing an antenna to measure gravitational waves. If we have two masses connected by a spring, the potential energy will be in the spring

$$\text{P.E.} = \frac{k}{2}(l - l_0)^2,$$

where l is the proper length of the spring, not the coordinate length. Because there are no inertial forces in these dust–mote coordinates, the kinetic term in the Lagrangian is just

$$\text{K.E.} = \frac{m}{2}v^2.$$

Note that the system remains at constant z as the wave passes by. Verify this by considering the total momentum and other invariants of the system.

Expanding and using the smallness of the wave amplitude F, we find

$$\mathcal{L} = m\dot{q}^2 - \frac{k}{2}(2q + \frac{l_0 F}{2})^2.$$

The generalized coordinate q is such that the dust mote coordinates of the particles are

$$x, y = \pm(q + \frac{l_0}{2}).$$

What about a very long bar? After the wave hits, the bar finds itself strained, and thus starts to transfer momentum. This momentum collects at the ends or anywhere else that there are changes in the elastic properties. The ends start to move, and an elastic wave propagates inward from the two ends of the bar. These elastic discontinuities have converted a gravitational wave into an elastic wave.

Strain–Free Coordinates

Dust mote coordinates have several good features: simple spatial dependence, well–behaved at large distances, manifestly planar symmetry, and a ready physical interpretation. On the other hand, laboratory sized objects (aluminum bars, the moon, and so on) are elastic bodies, and these care more about strains than accelerations. To simplify the study of such objects we can go go a different representation of the plane wave. This is also useful for developing a covariant intuition. It is the same plane wave. That different effects appear to be there shows you how little you can trust the usual decomposition of effects into inertial and elastic. All that separates them is a coordinate transformation.

The goal of the following calculation is to push the metric perturbation from the $x-y$ block of the metric to the $z-t$ block. One finds the correct coordinate transformation by inspired and lucky fiddling, preferably with a large blackboard. The new coordinates will be denoted by overbars, with

$$\begin{aligned} x &= \bar{x}(1-\frac{1}{2}F(\bar{z}-\bar{t})) \\ y &= \bar{y}(1+\frac{1}{2}F(\bar{z}-\bar{t})) \\ z &= \bar{z}-\frac{1}{4}(\bar{y}^2-\bar{x}^2)F'(\bar{z}-\bar{t}) \\ t &= \bar{t}-\frac{1}{4}(\bar{y}^2-\bar{x}^2)F'(\bar{z}-\bar{t}). \end{aligned}$$

Thus we have the differentials

$$\begin{aligned} dx &= d\bar{x}(1-\frac{F}{2})-\frac{\bar{x}F'}{2}(d\bar{z}-d\bar{t}) \\ dy &= d\bar{y}(1-\frac{F}{2})+\frac{\bar{y}F'}{2}(d\bar{z}-d\bar{t}) \\ dz &= d\bar{z}-(\bar{y}\,d\bar{y}-\bar{x}\,d\bar{x})\frac{F'}{2}-(\bar{y}^2-\bar{x}^2)(d\bar{z}-d\bar{t})\frac{F''}{4} \\ dt &= d\bar{t}-(\bar{y}\,d\bar{y}-\bar{x}\,d\bar{x})\frac{F'}{2}-(\bar{y}^2-\bar{x}^2)(d\bar{z}-d\bar{t})\frac{F''}{4} \end{aligned}$$

Now square these up and plug them into the expression for $\mathcal{G}$ given at the beginning of this section. It was in anticipation of this substitution that we solved for the x's in terms of the $\bar{x}$'s. We cancel a mess of terms, and use the smallness of F, as in

$$F(\bar{z}-\bar{t}) \sim F(z-t),$$

justified by a Taylor's Series Expansion. We also assume that the derivatives of F are no larger than F itself. We find

$$\mathcal{G} = -d\bar{t}^2 + d\bar{z}^2 + d\bar{y}^2 + d\bar{x}^2 - (\bar{y}^2 - \bar{x}^2)(d\bar{z} - d\bar{t})^2 \frac{F''}{2}.$$

[Multiplication here is by the tensor product.]

This is clearly not valid for large x or y.

The equations of motion follow from the Lagrangian condition for geodesics

$$\delta \int ds = 0,$$

$$\delta \int (1 - \dot{x}^2 - \dot{y}^2 - \dot{z}^2 + (y^2 - x^2)\frac{F''}{2})^{\frac{1}{2}}\, dt = 0.$$

Assume small velocities to expand the radical ($\dot{t} = 1$, $\dot{x} \ll 1$, ...

$$\delta \int (1 - \frac{\dot{x}^2}{2} - \frac{\dot{y}^2}{2} - \frac{\dot{z}^2}{2} + (y^2 - x^2)\frac{F''}{4})\, dt = 0.$$

The Euler–Lagrange equations are

$$\ddot{x} = x\frac{F''}{2}$$
$$\ddot{y} = -y\frac{F''}{2}$$
$$\ddot{z} = 0$$

where we have also specialized to the case of a detector small compared with a wavelength. This allowed us to discard terms like $(y^2 - x^2)F'''$. Over a region with size smaller than a wavelength around the origin, there are no strains at all.

The diagram was drawn in Mathematica by

```
PlotGradientField[x^2-y^2, {x, -1, 1},{y, -1, 1}]
```

These forces are often called "pseudoforces". Like centrifugal force, they can be changed or even eliminated by working in a different coordinate system. The worldlines of free particles are always geodesics of the true metric, and there are never any true gravitational

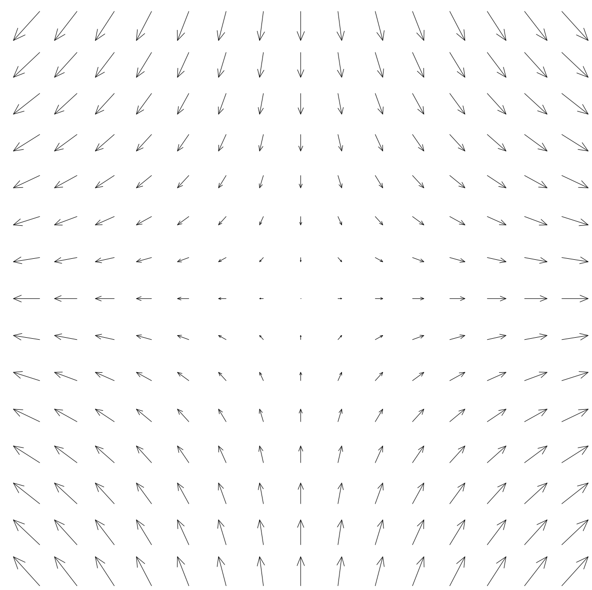


Figure 6-3. Gravitational-wave force field.

forces. This leads ultimately to the paradox that there really isn't any way to define gravitational energy.

Reconcile the Two Representations

Look at a free particle (dust mote) a distance L away from the origin, in the x–direction, say. We have

$$\ddot{x} = \frac{F''L}{2},$$
$$\ddot{y} = 0,$$
$$\ddot{z} = 0.$$

The solution is

$$x = L + \frac{L}{2}F(z - t).$$

The particle moves relative to these coordinates. Since these are strain free coordinates, this is an actual fluctuation in proper length. This is the length change that was inferred from the Doppler measurements.

The figure in Section 5 was drawn in nearly these strain–free coordinates.

General Polarization

Rotate coordinates by 45° to new coordinates

$$u = \frac{x+y}{\sqrt{2}},$$
$$v = \frac{x-y}{\sqrt{2}}.$$

The equations of motion in the new coordinates are

$$\ddot{u} = v\frac{F''}{2},$$
$$\ddot{v} = u\frac{F''}{2}.$$

If we just relabel the u and v back to x and y, we will have the correct equations for a rotated wave. The correct equations of motion for waves of both polarizations will clearly be

$$\ddot{x} = \frac{x}{2}F''_+ + \frac{y}{2}F''_\times,$$
$$\ddot{y} = \frac{x}{2}F''_\times - \frac{y}{2}F''_+,$$
$$\ddot{z} = 0.$$

The above force field can be derived from a potential, and be represented by field lines just like electrostatics. See MTW Box 37.2 and Figure 37.2.

Gravitational wave antennae

The physical effects of gravitational waves have only been seen indirectly. The energy lost in the waves has made noticable changes in the orbits of binary stars. To observe the waves in the laboratory is being attempted, but so far eludes us. Still, it is important to appreciate the measurement ideas since this is where the theory contacts reality.

All measurements of the gravitational field are differential measurements. This is because all particles have the same ratio of inertial to gravitational mass. The differential gravitational field is responsible for the tides, and is represented by the Riemann tensor. In the following figures I sketch the pattern of tidal forces and the differential forces found in gravitational waves.

For laboratory size apparatus, including even the entire planet, the detecting system will be much smaller than the wavelength of the

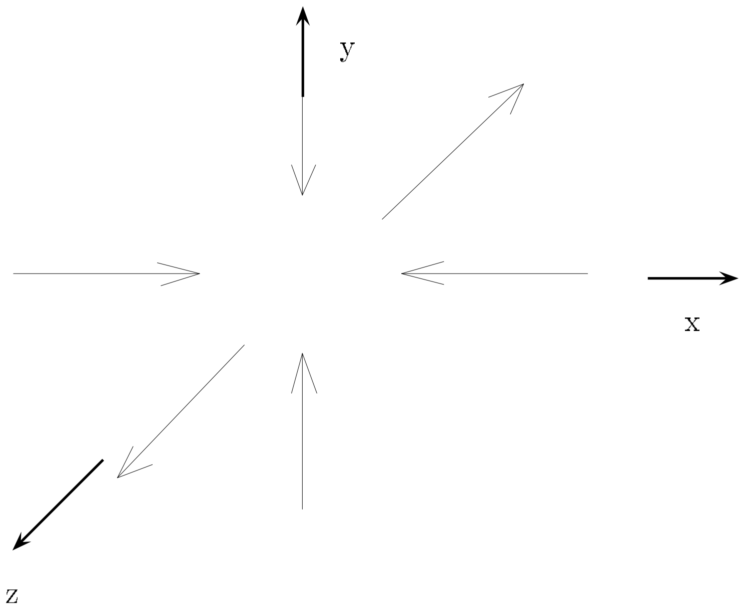


Figure 6-4. The tidal force field.

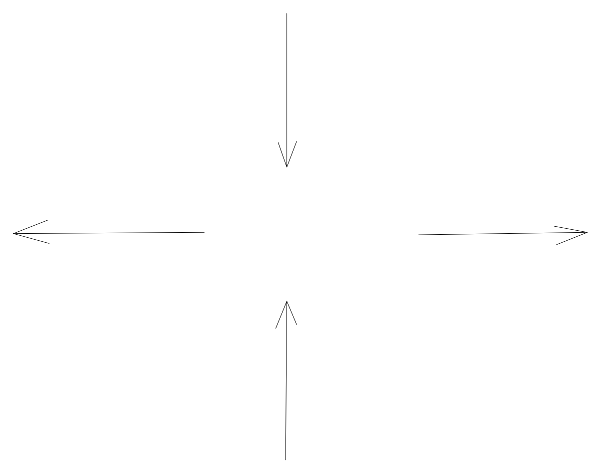


Figure 6-5. The differential forces in a gravitational wave.

radiation. For such systems the force is just a quasi-static force field. Thus for a circularly polarized gravitational wave, the force field pictured in figure 6-2 rotates at twice the frequency of the wave, in the transverse plane to the waves motion.

7. Sources of Gravitational Waves

To connect these waves to their sources, we need to solve the inhomogeneous wave equations that were given in Chapter 4. It is easiest to think about things in the space–time decomposition of the $h_{\mu\nu}$ field into scalar, vector, and tensor parts. The source of the scalar part is the local energy density. This energy includes rest–mass energy. The source of the vector part is the local momentum density, and the source of the tensor part is the local stress density.

$$\begin{aligned}
\nabla^2 \mathbb{H} - \partial_t^2 \mathbb{H} &= -16\pi \mathbb{S}, \\
\nabla^2 \mathbb{E} - \partial_t^2 \mathbb{E} &= -16\pi \mathbb{P}, \\
\nabla^2 \psi - \partial_t^2 \psi &= -16\pi \rho, \\
\nabla \cdot \mathbb{H} - \partial_t \mathbb{E} &= 0, \\
\nabla \cdot \mathbb{E} - \partial_t \psi &= 0.
\end{aligned}$$

Because energy is conserved, the local energy density can only change by flow; it cannot just disappear. Thus there cannot be a monopole wave solution. Because of the symmetry of a monopole, there is no direction for the energy to flow. In addition, momentum is conserved, and there are no solutions with monopole spatial symmetry for the vector perturbation either.

The solution with the simplest angular dependence is a quadrupole wave. The three parts achieve this overall symmetry in three different ways. If you know about angular momentum, you will recognize the three ways to get $J = 2$: spin 0 and $L = 2$, spin 1 and L=1, and spin 2 with $L = 0$. The energy density leads to a wave with quadrupole spatial behavior. The momentum density leads to a wave with dipole spatial behavior. The stress density leads to a wave with monopole spatial behavior. See Figure 1.

Rather than develop a fancy series and integral representation of the general solution, I am going to give a detailed treatment of the simplest quadrupole solution, the line quadrupole. This is simpler because it has axial symmetry. It has almost all of the typical characteristics of a gravitational wave source, however. Also, we will be able to use it as a building block for the general quadrupole source. Before I derive this solution, let me present the answer so you will appreciate where we are heading.

[If you want a derivation from general principles, look at Landau and Lifshitz, Classical Theory of Fields.]

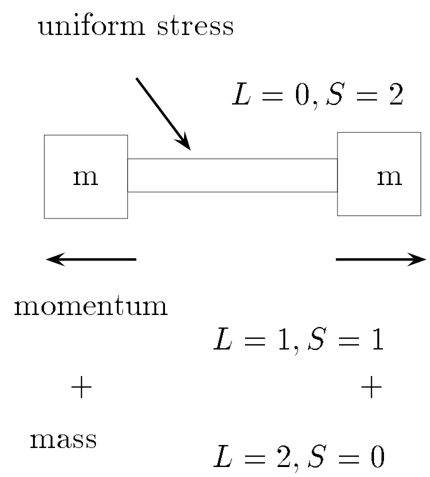
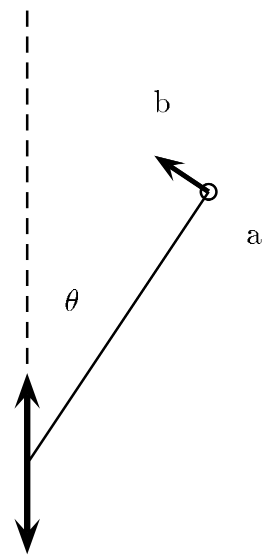


Figure 7-1. Symmetries of the quadrupole source.

The far field metric perturbation due to an axisymmetric quadrupole source is

$$\mathbb{H} = \frac{1}{2}\sin^2\theta \frac{Q''_{zz}(t-r)}{r}\mathbb{E}$$



The metric perturbation is only in the transverse direction. I am using a double–barred notation for dyadics (tensor product of two spatial vectors). The $\mathbb{E}$ dyadic is the polarization

$$\mathbb{E} = \hat{a}\otimes\hat{a} - \hat{b}\otimes\hat{b}$$

where the vectors a and b are orthonormal vectors in the transverse plane, with a in the plane containing the axis of symmetry. Here r is the distance from the source, and Q_{zz} is the quadrupole moment, defined by

$$Q_{ab} = \int (3x_a x_b - r^2\delta_{ab})\rho\, dV.$$

[Be careful, there are many different normalizations for quadrupole moments.]

Complete Line Quadrupole Solution

Recall the general solution of the ordinary one dimensional wave equation

$$f(x,t) = W(t-x) + V(t+x).$$

Here W and V are functions of a single variable evaluated at $(t-x)$ and $(t+x)$. They correspond to right–going and left–going waves respectively. There is a corresponding solution for spherically symmetric

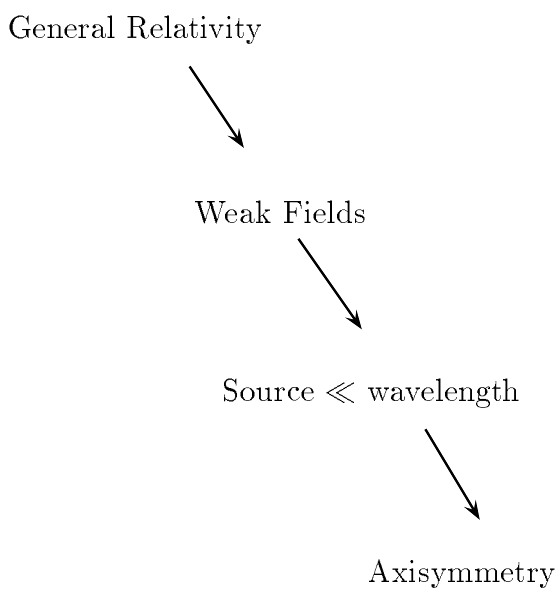


Figure 7-2. The slippery slope of approximations.

waves:

$$f(r,t) = \frac{W(t-r)}{r} + \frac{V(t+r)}{r}.$$

Now the two functions represent outgoing and incoming waves. For sources of radiation we are only interested in the outgoing waves. There are solutions similar to this for all of the multipoles. Combining them we can find a complete quadrupole solution for the linearized gravity equations

[Note the obvious spinology here.]

$$\psi = \frac{2z^2 - x^2 - y^2}{r^2}\left(\frac{W''(t-r)}{r} + 3\frac{W'(t-r)}{r^2} + 3\frac{W(t-r)}{r^3}\right)$$

$$\mathbb{E} = \frac{2z\hat{z} - x\hat{x} - y\hat{y}}{r}\left(\frac{W''(t-r)}{r} + \frac{W'(t-r)}{r^2}\right)$$

$$\mathbb{H} = \frac{2\hat{z}\hat{z} - \hat{x}\hat{x} - \hat{y}\hat{y}}{r}\left(W''(t-r)\right)$$

with outgoing waves. This is valid for all radii outside of the sources, [A complete derivation of this can be found in an appendix to my paper: J. Math. Phys. 12, 401-418 (1971).]

not just in the wave zone. The function W is related to the changing quadrupole moment of the source.

We are thinking here about a source that is small compared with the wavelength of the radiation. The next step is to match the near–zone field of the wave to the near–zone field of the quadrupole. For small r we have

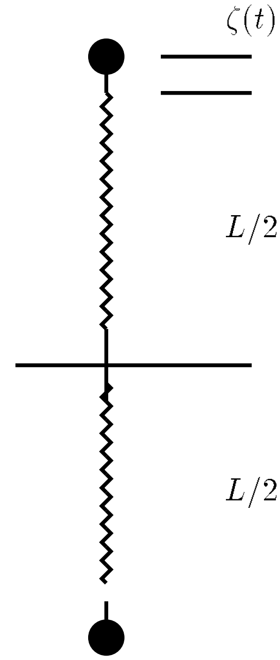
$$\psi \to (\frac{2z^2 - x^2 - y^2}{r^2})\, 3\frac{W(t)}{r^3}$$

and solving the simplest Newtonian gravity approximation

$$\nabla^2\psi = -16\pi\rho$$

relates this to the quadrupole moment.

Finally, we take the wave–zone fields, the $1/r$ fields, and make a gauge transformation to remove the longitudinal part of the wave. This eliminates the ψ and E parts of the perturbation.



Idealized Line Quadrupole

As an example of the use of this expression, consider two masses connected with a spring, *a la* freshman physics. Further, suppose that they execute only small oscillations about an equilibrium separation. The positions of the two masses along the z–axis will be given by a small function $\zeta(t)$

[Thus we layer two more approximations on top of everything: point masses and small oscillations.]

$$z = \pm(\frac{L}{2} + \zeta(t)).$$

Let us try to solve for the motion with initial conditions

$$\dot{\zeta} = 0,$$
$$\zeta = a.$$

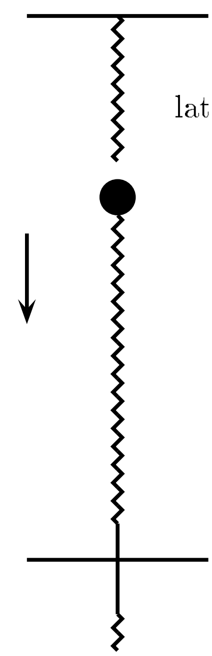
The mechanical motion of the system will be little affected by the radiation. This back reaction of the waves on the mechanical system will be the subject of a later chapter. Thus we have the solution

$$\zeta(t) = a\cos(\sqrt{\frac{k}{M}}t)$$

[We are even ignoring regular damping here.]

We now compute the quadrupole moment.

$$\begin{aligned} Q_{zz} &= \int (3z^2 - r^2)\rho \, dV \\ &= 2M(2(\frac{L}{2} + \zeta)^2) \\ &= ML^2 + 4ML\zeta(t) + \mathcal{O}(\zeta^2) \end{aligned}$$



The ML^2 part of the quadrupole moment is constant in time and so does not lead to any radiation. Thus it looks for all the world like the changing quadrupole moment is $4ML\zeta(t)$.

Why is $4ML\zeta(t)$ not the quadrupole moment that we want to stick into our expression? While it is the quadrupole moment of the two masses, and we are ignoring the mass of the spring as is usual in Freshman physics, we are making a mistake that we cannot ignore. To work with General Relativity, we must make sure that we have conserved energy and momentum. How have we gone wrong? Everything is fine after the system starts. The problem is with the time interval containing the initial conditions. The assumption is that before that time the system was not oscillating, but held at $\zeta = a$. Unlikely as it sounds, the latch that holds the system must be considered.

We will put in an idealized latch and compute its effects. We assume that it has a small mass μ and a spring constant κ. We cannot make the spring constant too small or it will not be strong enough to hold the system. We use a variable η to represent the latch degree of freedom. Of course we need a latch on each side. Balancing the forces gives us

$$\zeta k = \eta\kappa$$

in equilibrium. When the latch is released, the upper mass moves downward, starting its oscillation, and the latch springs upward, doing its own oscillations. The latch motion will be (ignoring damping for the moment)

$$\eta(t) = \frac{k}{\kappa} a \cos\sqrt{\frac{\kappa}{\mu}} t.$$

This has a quadrupole moment

$$Q_{\text{LATCH}} = -4\mu L\eta(t).$$

The factor of L appears here because the latch must be located at the position of the mass it is holding. No action at a distance allowed. The total quadrupole moment, which is what must appear in the radiation formula, is

$$Q_{\text{TOTAL}} = 4MLa\cos\sqrt{\frac{k}{M}}t - 4\mu La\cos\sqrt{\frac{\kappa}{\mu}}t.$$

Now you can see what will happen. For small μ, there is a vanishingly small contribution to the quadrupole moment. However, the frequency of the latch goes up as the mass of the latch decreases. What counts is the second derivative of the quadrupole moment

$$Q''_{\text{TOTAL}} = 4Lka\left(\cos\sqrt{\frac{k}{M}}t - \cos\sqrt{\frac{\kappa}{\mu}}t\right).$$

The radiation amplitude goes like the stress, ak, and does not depend on the frequency. Since the stress in the latch must equal the stress in the oscillator in order for the latch to hold it open, we have equal radiation amplitude from the latch and from the rest of the oscillator. This despite the disparity in masses. A well designed latch will be heavily damped, and so the radiation from the latch will only cancel the radiation from oscillator until they get out of phase, or the latch amplitude decays.

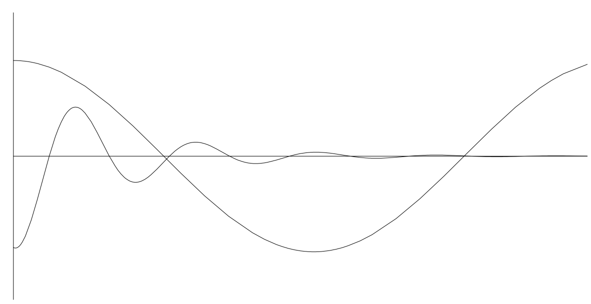


Figure 7-3. Radiation from the quadrupole and the latch.

General Quadrupole

Any quadrupole can be written as the sum of line quadrupoles.

> **Example:** If we write the z–axis line quadrupole in axes rotated by 45° in the z–x plane:
>
> $$\hat{z} \to (\hat{z} + \hat{x})/\sqrt{2}$$
> $$\hat{x} \to (\hat{z} - \hat{x})/\sqrt{2}.$$

then we have

$$4\hat{z}\hat{z} - 2\hat{x}\hat{x} - 2\hat{y}\hat{y} \rightarrow 3\hat{z}\hat{x} + 3\hat{x}\hat{z} - (2\hat{y}\hat{y} - \hat{z}\hat{z} - \hat{x}\hat{x}).$$

This allows us to write the off diagonal terms in the quadrupole moment tensor in terms of line quadrupoles inclined at 45°.

Example: A rotating pair of masses held together by a rod with stress in it is equivalent to four line quadrupoles in the plane of the orbit with a 90° phase shift between them, plus a static quadrupole perpendicular to the plane of the orbit.

You can use this decomposition to see that pole on you will get circularly polarized gravitational waves from the orbiting pair, while in the plane you will get linear polarization. The polar amplitude is larger since the amplitudes of two of them add, and then the intensities of the two out of phase add, so the final amplitude is larger than the in plane amplitude by a factor of $2\sqrt{2}$.

Transverse Traceless Representation

Here is a useful representation of the general quadrupole solution. Start with the quadrupole moment of a line quadrupole along the z– axis

$$Q(t) = Q_{zz}(t)(\hat{z}\hat{z} - \frac{1}{2}\hat{x}\hat{x} - \frac{1}{2}\hat{y}\hat{y}).$$

This Q is traceless; this follows from the definition of the quadrupole moments. If we project this transverse to the direction of motion in the far field, we replace (signs unimportant)

[The trace of a matrix forms a representation of spin zero, not spin two.]

$$\hat{z} \rightarrow \sin\theta\,\hat{a}$$
$$\hat{x} \rightarrow \hat{b}$$
$$\hat{y} \rightarrow \cos\theta\,\hat{a}.$$

If we apply this projection operation to the line quadrupole polarization tensor above we find

$$(\hat{z}\hat{z} - \frac{1}{2}\hat{x}\hat{x} - \frac{1}{2}\hat{y}\hat{y}) \rightarrow (\frac{3}{2}\sin^2\,\theta - \frac{1}{2})\hat{a}\hat{a} - \frac{1}{2}\hat{b}\hat{b}.$$

Now remove the trace which this acquired by the projection operation

$$\text{trace} = \frac{3}{2}\sin^2\,\theta - 1,$$

to find the transverse, traceless projection

$$\frac{3}{4}\sin^2\,\theta\,(\hat{a}\hat{a}-\hat{b}\hat{b}).$$

We will denote the result of this operation on Q by Q^{TT}. The operation depends on the direction of observation (propagation) but not on the direction of the coordinates. Thus we have

$$\mathbb{H}=\frac{2}{3r}\frac{d^2}{dt^2}(Q^{TT}).$$

All of the angular dependence has been absorbed in the transverse traceless projection operation.

Since the transverse traceless projection is a linear operator, the above expression applies to any quadrupole moment, not just the line quadrupole we started with. The operator is given by

$$Q^{TT}_{ab}=(\delta_{ac}-n_a n_c)Q_{cd}(\delta_{db}-n_d n_b)$$

where n is a vector in the direction of propagation. Seen in this form it is manifestly linear. Therefore the above expression applies to any quadrupole moment distribution whatsoever.

8. The Newtonian Approximation

The weak–field theory presented in the last few chapters was valid for weak fields with the space and time derivatives treated equally. This was appropriate for gravitational waves, which move at the speed of light. In a self–gravitating situation, however, the motions take place at a speed less than the speed of light. If we call the typical speed ϵ, then this will be caused by gravitational potentials of size ϵ^2. Time derivatives will be smaller than spatial derivatives by a factor of ϵ.

The source of the perturbation will have an expansion with the potential ψ going like ϵ^2, and hence generated by matter with density scaling like this. The momentum will scale like ϵ^3, since momentum is mass moving with this slow velocity. What about the stresses? If the motions are only influenced by self–gravity, then there are no stresses at all. If there were other stresses, small enough to keep the motions on the correct timescale, then they would be of size ϵ^4. The missing stress in the self–gravitating case is provided by the non–linearity: terms quadratic in ψ are of size ϵ^4, exactly what is needed. Thus we need to extend our theory to include terms quadratic in ψ in the sources of the dyH field.

[That a theory valid for motions as fast as light is not valid for slower speeds should seem paradoxical. If not you are either very sophisticated or haven't thought very hard about this.]

Weak–Field Slow–Motion Expansion

To calculate the gravitational stress follow the program: take a spacetime metric with the expansion

$$\begin{aligned} g_{tt} &\sim -1 + \tfrac{1}{2}\epsilon^2\psi + \ldots \\ g_{ta} &\sim -\epsilon^3\mathbb{E}_a \\ g_{ab} &\sim \eta_{ab}(1 + \tfrac{1}{2}\epsilon^2\psi) + \epsilon^4 H_{ab}. \end{aligned}$$

[The ψ term in the spatial part reflects the trick of using $\bar{h}$ instead of h.]

Now one calculates the Einstein tensor for this metric. The efficient way to do this is to either use a computer algebra package, or to use the method of moving frames. We will cover the method of moving frames when we do the Robertson–Walker spacetimes needed in cosmology, so for now you can just take the Einstein tensor as a gift.

$$
\begin{aligned}
G_{tt} &\sim -\tfrac{1}{2}\epsilon^2\nabla^2\psi, \\
G_{ta} &\sim -\tfrac{1}{2}\epsilon^3\left[\nabla^2 E_a - (\nabla\cdot\mathbb{E} + \frac{\partial\psi}{\partial t})_{,a}\right], \\
G_{ab} &\sim \epsilon^4\Bigg[-\tfrac{1}{2}\nabla^2 H_{ab} \\
&\quad + \tfrac{1}{8}\psi_{,a}\psi_{,b} + \tfrac{1}{4}\psi\psi_{,ab} - g_{ab}(\tfrac{3}{16}\nabla\psi\cdot\nabla\psi + \tfrac{1}{4}\psi\nabla^2\psi) \\
&\quad + \tfrac{1}{2}(H_{ac,c} - V_{a,t})_{,b} + \tfrac{1}{2}(H_{bc,c} - V_{b,t})_{,a} \\
&\quad - \tfrac{1}{2}g_{ab}\Big(\nabla\cdot(\nabla\cdot\mathbb{H} + \partial_t\mathbb{E}) + \partial_t(\nabla\cdot\mathbb{E} + \partial_t\psi)\Big)\Bigg].
\end{aligned}
$$

[The first equation is just the Poisson equation. The nonlinear stuff becomes the gravitational stress tensor, and the last bracket is flattened by the gauge condition.]

Again we want to impose the Lorentz gauge condition

$$
\begin{aligned}
\nabla\cdot\mathbb{E} + \frac{\partial\psi}{\partial t} &= 0, \\
\nabla\cdot\mathbb{H} + \frac{\partial\mathbb{E}}{\partial t} &= 0.
\end{aligned}
$$

With this gauge condition we have Poisson equations for the perturbations unlike linearized theory, where we had wave equations:

$$
\begin{aligned}
\nabla^2\psi &= -16\pi\rho, \\
\nabla^2\mathbb{E} &= -16\pi\mathbb{P}, \\
\nabla^2\mathbb{H} &= -16\pi(\mathbf{S} + \mathbf{S}_G).
\end{aligned}
$$

Here $\mathbb{P}$ is the momentum density, and $\mathbf{S}$ are the non–gravitational stresses, and $\mathbf{S_G}$ are the gravitational stresses. These come from non–linearity and are given by

$$
\mathbf{S}_G = -\frac{\pi}{8}\left[\frac{\psi_{,a}\psi_{,b}}{8} + \frac{\psi\psi_{,ab}}{4} - g_{ab}(\tfrac{3}{16}\nabla\psi\cdot\nabla\psi + \tfrac{1}{4}\psi\nabla^2\psi)\right]
$$

[The calculation of the fraction $\frac{3}{16}$ here involves over twenty hours of hard calculation.]

From this we recover Einstein's relation

$$
\nabla\cdot\mathbf{P} + \frac{\partial\rho}{\partial t} = 0.
$$

That is really the equivalent of $E = mc^2$ and a bit more. We also get the Newtonian force law, with the above expression for the gravitational stress

$$
\nabla\cdot\mathbf{S}_G = -\tfrac{1}{4}\rho\nabla\psi.
$$

This is $F = ma$, and a bit more. From the Poisson equation above we see that ψ is $\frac{1}{4}$ of the usual Newtonian potential.

Non–Uniformities in these expansions

This need to keep the sizes of derivatives under control means that the remarks I made in the discussion about detecting gravitational waves using elastic bodies need to be more carefully discussed. The approximations need to be rethought whereever there are discontinuities in the matter density. At such points the spatial derivatives are much larger than the temporal derivatives. Unfortunately a rediscussion of this is very complex, and would involve a careful discussion of the junction conditions for General Relativity. These are non–trivial and unfortunately non–obvious.

9. Gravitational Radiation Damping

A system with a time–dependent quadrupole moment will radiate gravitational waves. These waves then interact with systems in the wave–zone which could lead to their detection. These forward couplings are well separated from the back couplings by the extreme weakness of gravitation. The antenna will scatter waves, and an array of scatterers might have an index of refraction. The radiation emission will act back on the source, causing damping.

Radiation damping is a common phenomenon in the real world. The motion of a racing sailboat is severely affected by the energy lost in the waves. Ships designed to sail at a single speed, like oil tankers, are designed to minimize this radiation drag. At the present time the only evidence we have for the physical reality of gravitational radiation is the damping effect of the radiation on astronomical systems. This is circumstantial evidence at best, we have to be sure that we have properly accounted for all other damping mechanisms.

In ordinary physics one calculates damping by finding the energy that appears at infinity. This is tricky in General Relativity because there is no global energy conservation. In addition, one may want more information about the damping than just the total energy. The radiation may cause energy redistribution in a system with several degrees of freedom. The way to deal with this is to find the damping forces directly. Either way leads to the expression for the radiated energy

$$L_{\mathrm{GW}} = \frac{1}{45}\left\langle \dddot{Q}_{ab}\, \dddot{Q}_{ab} \right\rangle.$$

[Note the third time derivative in the expression.]

Here L is the gravitational energy loss rate due to quadrupole gravitational radiation. Energy loss rate is dimensionless. $L = 1$ corresponds to 3.63×10^{59} ergs/sec. The luminosity of the sun is 1.075×10^{-26}. The quadrupole moment Q is related to the familiar moment of inertia tensor by

$$Q_{ab} = 3(I_{ab} - \frac{1}{3}\delta_{ab}\,\mathrm{tr}\, I).$$

Finally, the angle brackets denote a time average over many cycles.

The size of this result is found from the scaling

$$\dddot{Q}_{ab} \sim \frac{(\text{mass of moving part})\,(\text{size of system})^2}{(\text{time scale})^3}.$$

If the system has motions on the scale of its size, for example a binary star, but not our system with small oscillations, then

$$\begin{aligned}\dddot{Q}_{ab} &\sim \frac{(\text{mass})\,(\text{typical velocity})^2}{(\text{timescale})},\\ &\sim \frac{(\text{kinetic energy})}{(\text{time})},\\ &\sim (\text{internal power}).\end{aligned}$$

Thus we see that the radiated power goes as the square of the internal power. This power is dimensionless and much smaller than one, so the radiated power is small.

> **Example:** The gravitational wave luminosity for a teacher waving his hands can be estimated
>
> $$\begin{aligned} m &\approx kg = 2\times 10^{-36}\,\text{sec}\\ T &\approx 1\,\text{sec}\\ v &\approx 1\,\text{m/sec} = 3\times 10^{-9}\\ L_{\text{internal}} &\approx 2\times 10^{-54}\\ L_{\text{GW}} &\approx 4\times 10^{-108} = 1.5\times 10^{-48}\,\text{erg/sec}.\end{aligned}$$

Acoustics Example

To see the general run of such a calculation, let me work a scalar example. Consider the emission of sound waves from a vibrating sphere. We assume small amplitudes, and work in a linearized theory. Acoustics is a code word that means irrotational in fluid mechanics. Thus there exists a velocity potential ϕ such that

$$\mathbf{v} = \nabla\phi,$$

and the pressure differences are given by

$$p - p_0 = \rho_0 \frac{\partial\phi}{\partial t}.$$

This velocity potential satisfies the wave equation

$$\nabla^2\phi - \frac{\partial^2\phi}{\partial t^2} = 0.$$

We have taken units where the sound speed is unity.

The boundary condition at the surface of the vibrating sphere, taken to have a small radius ϵ, will be given by

$$\frac{\partial \phi}{\partial r} = \frac{d\zeta}{dt} \qquad \text{at } r = \epsilon.$$

where $\zeta(t)$ describes the radial oscillation of the sphere. The dynamics of the sphere for small oscillations is given by the linear equation

$$\frac{d^2\zeta}{dt^2} + \zeta = \kappa \frac{\partial \phi}{\partial t}.$$

Here κ is a coupling constant which determines how stiff the sphere is relative to the impedance of the air. We have picked a timescale using the unloaded frequency of the shell. We are going to consider only the case of a heavy shell, with κ very small, even compared with ϵ.

Now in this problem there are two length scales, and this is a warning that the problem will be singular. One length is the wavelength of the radiation, and the other is the size of the shell. We work here the case where the body is much smaller than a wavelength, $\epsilon \ll 1$. The coordinates that we have been using to describe the problem have time and space derivatives of the same size, and are appropriate for a description of the sound waves. We need a different coordinate system to describe the sound field in the immediate neighborhood of the sphere. Call this near–zone coordinate R

$$R \equiv r/\epsilon.$$

For an inner expansion, valid near the body, we take

$$\phi \sim \epsilon A(R,t) + \epsilon^2 B(R,t) + \cdots$$

and the terms will satisfy

$$\nabla^2 A = 0,$$
$$\nabla^2 B = 0.$$

with boundary conditions at $R = 1$

$$\frac{\partial A}{\partial R} = \frac{d\zeta}{dt},$$
$$\frac{\partial B}{\partial R} = 0.$$

We do not yet have the outer boundary condition for these. They have solutions

$$A = -\frac{1}{R}\frac{d\zeta}{dt} + \alpha(t),$$
$$B = \beta(t).$$

With arbitrary functions of time α, β.

An expansion valid in the wave zone will have the form

$$\phi \sim \epsilon^2 a(r,t) + \cdots,$$

and the general solution of this which represents only an outgoing wave will be

$$a(r,t) = W(t-r)/r,$$

with W another arbitrary function.

[Matching is a technical term for the precise way of fitting these two asymptotic expansions together using the methods of singular perturbations.]

To match these two solutions together, we ensure that the small–r behavior of this outer solution matches the large–r behavior of the near–zone solution. We look at intermediate values of r, smaller than a wavelength but still larger than the sphere. In these nondimensional units, $r \sim \sqrt{\epsilon}$ would do. For intermediate r we have

$$\epsilon^2 \frac{W(t-r)}{r} = \epsilon \frac{W(t-\epsilon r)}{R} \to \epsilon \frac{W(t)}{R} - \epsilon^2 W'(t).$$

For these to match we must have

$$W(t) = -\frac{d\zeta}{dt},$$
$$\alpha(t) = 0,$$
$$\beta(t) = \frac{d^2\zeta}{dt^2}.$$

Notice how the near–zone motion $\zeta(t)$ determines the function $W(t)$, which then determines the homogeneous solution $\beta(t)$ carrying the direction of time in from the out-going waves boundary condition.

The near–zone field is thus

$$\phi \sim -\epsilon \frac{1}{R}\frac{d\zeta}{dt} + \epsilon^2 \frac{d^2\zeta}{dt^2},$$

and this leads to a force on the shell, and an equation of motion

$$(1+\epsilon\kappa)\frac{d^2\zeta}{dt^2} + \zeta - \epsilon^2\kappa\frac{d^3\zeta}{dt^3} = 0.$$

The radiation shifts the frequency by adding a term to the inertial of the shell. This accounts for the air that is moved back and forth by the shell as it oscillates. In fluid mechanics it is called entrained mass. There is also a term which is in quadrature with this which can extract energy from the oscillator. One calculates the radiation resistance by including this term as as small correction. That this small parameter multiplies the highest derivative in the equation leads also to spurious solutions of the approximate equations. In electrodynamics these spurious solutions are called runaway solutions, and have generated much confusion among those unfamiliar with singular perturbation techniques.

A similar calculation in General Relativity leads to a similar resistive force

$$-(8\pi\sqrt{10}/75)\rho r \mathbf{Y}_{21M}\frac{d^5}{dt^5}Q_{2M}$$

Binary Star System

A crude estimate of the radiation from a binary star system can be made from Keppler's Law

$$a^3\omega^2 = m$$

which gives a velocity

$$v = a\omega$$

and a kinetic energy

$$ma^2\omega^2 = m^{\frac{5}{3}}\omega^{\frac{2}{3}}$$

The internal energy flow will be the flux of this kinetic energy, so the internal luminosity is

$$L_{\mathrm{internal}} = (m\omega)^{\frac{5}{3}}$$

and the gravitational wave luminosity

$$L_{\mathrm{GW}} = (m\omega)^{\frac{10}{3}}.$$

A more careful calculation gives the energy loss from two masses in a circular orbit via quadrupole gravitational waves to be

$$L_{\mathrm{GW}} = (1.2L_{\odot} \text{ or } 4.7\times 10^{42} \text{ erg/sec})\frac{m_1^2\, m_2^2\, \omega^{\frac{10}{3}}}{(m_1+m_2)^{\frac{2}{3}}},$$

where the masses should be in solar masses and the rotation frequency in hz. The radiation will be at frequency 2ω.

Non–circular orbits tends to radiate more, radiate at higher harmonics of the orbital frequency, and tend to become circular with time. For circular orbits, the orbits will spiral into one another with

$$a/a_0 = (1 - t/\tau)^{\frac{1}{4}},$$

with timescale

$$\tau = \tfrac{5}{256} \frac{a_0^4}{m_1 m_2 (m_1 + m_2)}.$$

Ordinary stars undergoing this spiral process would suffer tidal distortions which would absorb energy and drastically affect this rate. Such processes must be fully understood before such tests can be used to validate General Relativity.

Wave Energy

While we have expressions for the energy lost, it would be nice to an expression for the energy in the wave. In general such an expression does not exist. In the limit that the waves are high–frequency ripples on a slowly changing background spacetime, the tensor

[This is another singular perturbation problem.]

$$T^{\mathrm{GW}}_{\mu\nu} = \tfrac{1}{32\pi} \left\langle h^{\mathrm{TT}}_{\alpha\beta,\mu} h^{\mathrm{TT}}_{\alpha\beta,\nu} \right\rangle$$

can be used to represent the energy of the wave. Again the angle brackets denote averages over several waves, and the TT stands for transverse and traceless.

This tensor gives the correct energy loss rates and also the effects of the gravitational waves on the background curvature of the universe as a whole. It cannot be used to localize the wave energy, and cannot be used to predict the non–linear evolution of the wave.

Falling Charges

In the General Relativity paradigm, an object sitting on my desk is being accelerated. If is electrically charged, does it then radiate electrodynamic waves? This is an ancient and honorable paradox.

We might first inquire what the wavelength of the radiation involved will be. The only timescale is given by g, and so the radiation will have a wavelength of around a light year. Already we are in trouble with our naive thinking, because the wave zone is much larger than the size of the earth, and we certainly can't use the Principle of Equivalence here, which presupposes a uniform gravitational field.

A careful calculation carried out with the matching ideas of this section shows that the correct thinking is not that "accelerated charges radiate", but the changing field strengths propagate as waves. A charge placed midway between two orbiting masses will be unmoved, but the fields will be affected by the masses and there will be electrodynamic radiation from this system.

10. Cosmology and General Relativity

While General Relativity is a minor correction to most of everyday physical situations, it is absolutely necessary for an understanding of cosmology. While one can treat an infinite flat universe as a convenient local approximation, an argument similar to that used in Olber's paradox shows that Newtonian theory cannot treat an infinite universe. The nonlinearities in General Relativity allow the infinite universe to be described without any divergences.

When do we need General Relativity? In our units where $G = 1$, GR will be important when we have

$$M/R \sim 1, \qquad \text{that is} \quad R \sim 1/\sqrt{\rho}.$$

The universe contains at least the galaxies that we see, with a mass of around $10^{11}\, M_{\odot}$. The spacing between galaxies is roughly a megaparsec, so the smoothed out density of just the galaxies is $\rho_{\mathrm{GAL}} \approx 5 \times 10^{-37}\mathrm{sec.}^{-2}$. This says that GR will be important for length scales of 2×10^{18} sec. Since the ages of rocks, nuclei, and the planets is around 3×10^{17} sec, any study of the universe must involve GR in its structure, dynamics, and interpretation.

Symmetric Spaces

Not looking for trouble, we start with the simplest non–trivial idealization. We try to find a model of the universe which has no preferred direction, is isotropic, and has no preferred location, is homogeneous. In the old days, one added to this list the requirement that there be no preferred time, that the universe be stationary. Now we think that the redshift of galaxies can only be explained by an evolving universe, and we drop the requirement that all times be equivalent.

The above used the subjunctive, a necessary mode in cosmology. Very little is known for certain. One must always keep an open mind. General Relativity has been only mildly confirmed, and could go out tomorrow. For all we know, next decade we will again be believing in a tired light explanation for the redshift, and a stationary universe might again be preferred. It is points for General Relativity as a theory that it cannot explain a static universe, and that we don't observe one either.

Example: In 1917 Einstein proposed that the universe be modeled by the spacetime

$$\mathcal{G} = -dt^2 + R^2 \left[d\chi^2 + \sin^2\chi \, (d\theta^2 + \sin^2\theta \, d\phi^2) \right]$$

with R a constant. This is a fully symmetric model of the universe.

Example: The spatial metric in the above example is the 3–sphere. It is one of the three distinct homogeneous and isotropic 3–spaces with a positive definite metric. Its topology is carefully discussed in SGC.

The matter in the universe must not single out a preferred direction, and so in these symmetric models must have a 4–velocity given by (on average) $\frac{\partial}{\partial t}$.

The spacetime in the above example has an unreasonable $T^{\mu\nu}$ if you expect it to satisfy Einstein's equations. If you try to repair this by messing with the equations, then you end up with an unstable model. We will discuss this later. This is the basis for the statement that General Relativity is only compatible with an expanding universe.

It is easy to visualize the symmetric two spaces. Look at spaces formed by pasting together quadrilaterals, with different numbers of them meeting at every vertex. If four meet at every vertex, then we generate an infinite checkerboard. If three meet at every vertex, we generate the closed space which is the surface of the cube. If five or more meet, then we generate an infinite space that resembles the familiar Escher drawings.

An interesting question that you can pose is whether the knight's tour sketched above closes or not. Take two steps forward, then turn left. Repeat indefinitely.

The symmetries of this tiling are best represented not by finding translations and rotations, but by taking the reflections in the sides of the fundamental scalene triangle. One of these fundamental triangles is sketched inside each of the squares.

Another picture of this tiling with five squares meeting at every point is interleaved at the end of this section. It shows how five of these squares can be put together into an X-shape, and then the tiling of space with these.

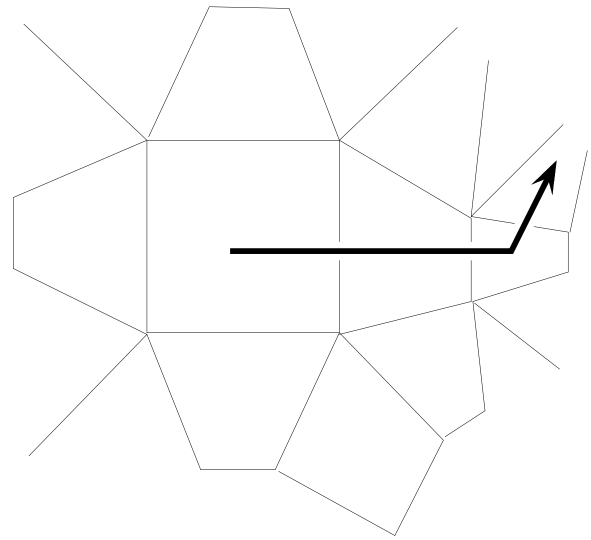


Figure 10-1. A sketch of five squares meeting at every vertex, and a possible knight's move.

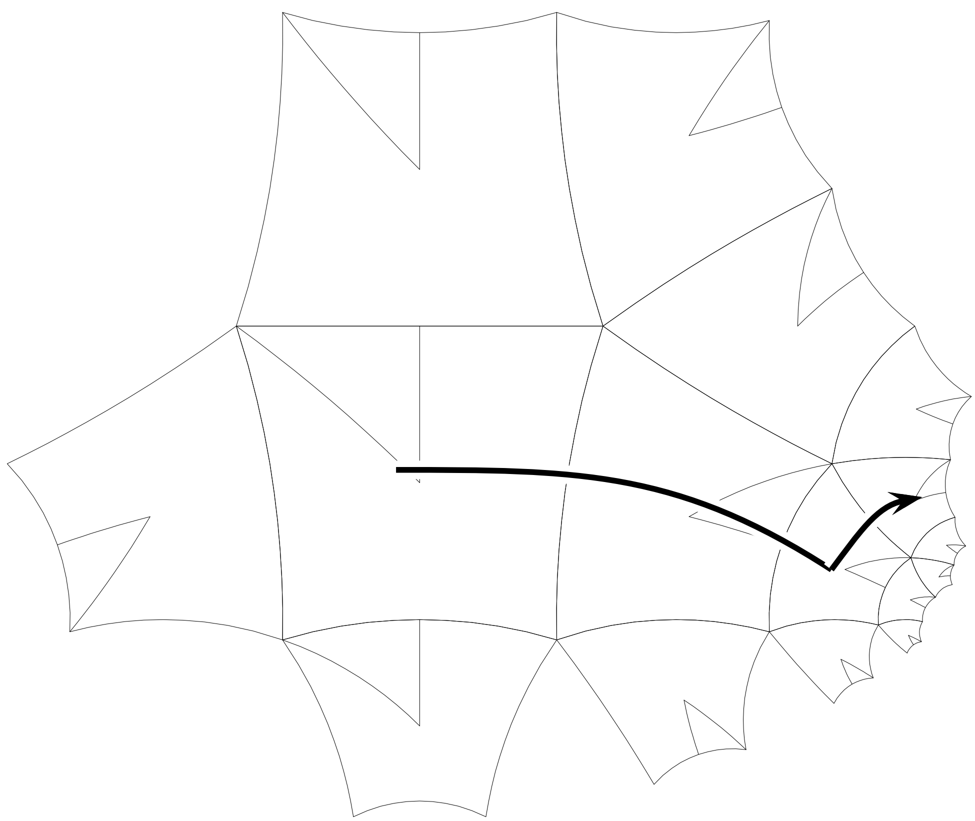


Figure 10-2. The Poincare disk representation of the preceeding sketch.

Robertson–Walker Spacetimes

A spacetime model for cosmology is a spacetime plus matter worldlines. In some cosmologies the matter is an insignificant addition, and does not show up in the $T^{\mu\nu}$. The spacetimes that satisfy our symmetry requirements can be made up by using a symmetric 3–space for the spatial part, and adding that to a dynamic time variable.

One can make a complete catalog of the symmetric 3–spaces:

$$\mathcal{G} = d\chi^2 + \sin^2\chi\, d\Omega^2 \qquad 0 \le \chi \le \pi,$$
$$\mathcal{G} = dr^2 + r^2\, d\Omega^2 \qquad 0 \le r \le \pi,$$
$$\mathcal{G} = d\chi^2 + \sinh^2\chi\, d\Omega^2 \qquad 0 \le \chi \le \infty.$$

[Here $d\Omega^2$ is the metric on the surface of the 2–sphere.]

These are called the 3–sphere, Euclidean 3–space, and the 3–pseudo- sphere. Here χ is dimensionless, while r has dimensions of length. We will write these generically as

$$\mathcal{G} = d\chi^2 + S^2(\chi)\, d\Omega^2.$$

A Robertson–Walker spacetime takes one of these symmetic 3–spaces and bolts on a time axis:

$$\mathcal{G} = -dt^2 + R^2(t)(d\chi^2 + S^2(\chi)\, d\Omega^2).$$

These are useful in any geometric theories of the universe, not just the models used in General Relativity.

Curvature Computation

We describe the Robertson–Walker spacetime with the orthonormal frame

$$\omega^t = dt,$$
$$\omega^\chi = R\, d\chi,$$
$$\omega^\theta = RS\, d\theta,$$
$$\omega^\phi = RS\sin\theta\, d\phi.$$

[Note that the indices here are labels for different one-forms, not the components of a single one-form.]

This list of four orthonormal 1-forms contains all of the metric information.

The connection forms satisfy

$$d\omega_i + \omega_{ij} \wedge \omega^j = 0,$$
$$\omega_{ij} + \omega_{ji} = 0.$$

and we want to solve this by inspection. The indices on the forms are just labels, and when we lower an index, nothing happens except that the sign changes for the t–index. The connection forms are

[This was done with more detail in the last chapter of ADG.]

$$\begin{aligned}
\omega_{t\chi} &= -R'\,d\chi, \\
\omega_{t\theta} &= -R'S\,d\theta, \\
\omega_{t\phi} &= -R'S\sin\theta\,d\phi, \\
\omega_{\chi\theta} &= -S'\,d\theta, \\
\omega_{\chi\phi} &= -S'\sin\theta\,d\phi, \\
\omega_{\theta\phi} &= -\cos\theta\,d\phi.
\end{aligned}$$

The symmetry of the spacetime is not apparent when these are expressed in terms of a coordinate basis. We can write these

$$\begin{aligned}
\omega_{t\chi} &= -(R'/R)\,\omega^{\chi}, \\
\omega_{t\theta} &= -(R'/R)\,\omega^{\theta}, \\
\omega_{t\phi} &= -(R'/R)\,\omega^{\phi}, \\
\omega_{\chi\theta} &= -(S'/RS)\,\omega^{\theta}, \\
\omega_{\chi\phi} &= -(S'/RS)\,\omega^{\phi}, \\
\omega_{\theta\phi} &= -(\cos\theta/RS\ \sin\theta)\,\omega^{\phi}.
\end{aligned}$$

Now for the curvature forms. These satisfy

$$\Omega_{ij} = d\omega_{ij} + \omega_{il}\wedge\omega^{l}{}_{j}.$$

It is sufficient to calculate $\Omega_{t\chi}$ and $\Omega_{\chi\theta}$, a time-space one and a space-space one. The rest follow from symmetry. The symmetry must be apparent in the Ω_{ij} because they are measureable, while the ω_{ij} are not.

$$\begin{aligned}
\Omega_{t\chi} &= d\omega_{t\chi} + \omega_{t\theta}\wedge\omega^{\theta}{}_{\chi} + \omega_{t\phi}\wedge\omega^{\phi}{}_{\chi}, \\
&= -R''\,dt\wedge d\chi \\
&= -(R''/R)\,\omega^{t}\wedge\omega^{\chi}.
\end{aligned}$$

$$\begin{aligned}
\Omega_{\chi\theta} &= d\omega_{\chi\theta} + \omega_{\chi t}\wedge\omega^{t}{}_{\theta} + \omega_{\chi\phi}\wedge\omega^{\phi}{}_{\theta}, \\
&= -S''\,d\chi\wedge d\theta + (R'\,d\chi)\wedge R'S\,d\theta \\
&= (-S'' + SR'R')\,d\chi\wedge d\theta
\end{aligned}$$

We can handle the three separate symmetric spaces if we define the index κ to be plus one for the 3–sphere, minus one for the 3–pseudosphere, and zero for Euclidean space. Then we have

$$S'' = -\kappa S$$

and

$$\Omega_{\chi\theta} = \frac{(R')^2 + \kappa}{R^2} \quad \omega^\chi \wedge \omega^\theta.$$

The Riemann tensor in orthonormal frame components is given by

$$\Omega_{ij} = \hat{R}_{ij|kl|} \quad \omega^k \wedge \omega^l.$$

[The vertical bars stand for the ordered summation. This elimanates duplicate terms and extraneous factors.]

This is really the equivalent of a generating function for the Riemann tensor components; i.e. it is an expansion whose terms given the components. Here it is a finite expansion. This is useful in situations where many of the components are zero.

The final shortcut to the computation of the Riemann tensor is an identity that is totally unmotivated. By some miracle we have the Einstein tensor

$$G^\mu_\nu = -\delta^{\mu\gamma\delta}_{\nu\alpha\beta} R^{|\alpha\beta|}{}_{|\gamma\delta|}.$$

This is useful because the right–hand side is so easy to compute. This is easy to prove, but wicked to motivate. The 6–index tensor is called a Kronecker–delta. It has a value of one if the top and bottom indices are an even permutation of one another, minus one if an odd permutation, and zero otherwise.

[See section 27 in ADG.]

The Einstein tensor is another physical object, and must have the same symmetry as the curvature tensor. We compute the components in the orthonormal frame, denoted here by the hat over the kernal.

$$\begin{aligned}\hat{G}^t_t &= -\delta^{t\gamma\delta}_{t\alpha\beta} \hat{R}^{|\alpha\beta|}{}_{|\gamma\delta|} \\ &= -\hat{R}^{\chi\theta}{}_{\chi\theta} - \hat{R}^{\chi\phi}{}_{\chi\phi} - \hat{R}^{\theta\phi}{}_{\theta\phi}.\end{aligned}$$

To find the needed Riemann tensor components, look at

$$\Omega_{\chi\theta} = \frac{(R')^2 + \kappa}{R^2} \omega^\chi \wedge \omega^\theta.$$

this implies that

$$\hat{R}_{\chi\theta\chi\theta} = \frac{(R')^2 + \kappa}{R^2},$$

and the same for the others. Thus

$$\hat{G}^t_t = -3(\frac{(R')^2 + \kappa}{R^2}).$$

A similar computation gives us

$$\hat{G}^\chi_\chi = -\left(2R''/R + (R'/R)^2 + \kappa/R^2\right)$$

All other components of the Einstein tensor are zero or related to these by symmetry.

11. Friedmann Models

The Friedmann models are generated by taking a Robertson–Walker spacetime, imposing the Einstein field equations, and using an equation of state appropriate to either dust or radiation. The last section showed that the Robertson–Walker spacetime, given by the moving frame

$$\begin{aligned} \omega^t &= dt, \\ \omega^\chi &= R\, d\chi, \\ \omega^\theta &= RS\, d\theta, \\ \omega^\phi &= RS \sin\theta\, d\phi. \end{aligned}$$

has an Einstein tensor given by

$$-3\frac{R'R' + \kappa}{R^2}\,\omega^t \otimes \omega^t$$
$$-[2\frac{R''}{R} + (\frac{R'}{R})^2 + \frac{\kappa}{R^2}]\,(\omega^\chi \otimes \omega^\chi + \omega^\theta \otimes \omega^\theta + \omega^\phi \otimes \omega^\phi).$$

[Note the advantages of the explicit notation here: we need no remark about all other components zero, plus the expansion in an orthonormal frame is explicitly indicated.]

We now invoke the Einstein equations. $T^{\mu\nu}$ will have to have the same symmetry. Therefore we must be studying what is called a perfect fluid, and one which is at rest in the orthonormal frame. We write

$$T = \rho\,\omega^t \otimes \omega^t + p\,(\omega^\chi \otimes \omega^\chi + \omega^\theta \otimes \omega^\theta + \omega^\phi \otimes \omega^\phi).$$

Here ρ is what is often called the proper energy density, and p is the local pressure, i.e. the local momentum flux density. We need to add to this a constitutive equation relating the density and the pressure. Typical equations of state are

dust	$p = 0,$
radiation	$p = \frac{1}{3}\rho,$
stiff matter	$p = \rho.$
"dressed" vacuum	$p = -\rho = -\Lambda.$

[Stiff matter is not known to exist. It has a sound speed equal to the speed of light.]

The Einstein equations written for the Robertson–Walker spacetime are usually called the Friedmann equations. They are

$$2\frac{R''}{R} + (\frac{R'}{R})^2 + \frac{\kappa}{R^2} + 8\pi p = 0,$$
$$(\frac{R'}{R})^2 + \frac{\kappa}{R^2} - 8\pi\rho/3 = 0.$$

Critical Dust Universe

To see the structure of the system, let us look at nearly the simplest case, the critical dust universe, with

$$p = 0, \qquad \kappa = 0.$$

Now one cannot freely specify ρ, that would give us too many equations. Look at the amount of matter in a comoving volume ρR^3

[Here κ is the curvature indicator.]

$$\begin{aligned} \tfrac{8\pi}{3}\frac{d}{dt}(\rho R^3) &= \frac{d}{dt}(RR'R'), \\ &= (R')^3 + 2RR'R'', \\ &= 0. \end{aligned}$$

This should have been expected. The equation $T^{\mu\nu}{}_{;\nu} = 0$ implies that matter must be conserved. We showed this earlier.

Since we have a conserved quantity, we should give it a name. Define

$$\varpi \equiv \tfrac{8\pi}{3}\rho R^3.$$

Reducing the order of our system from two to one. The remaining equation is

$$(\frac{R'}{R})^2 = \varpi/R^3$$

easily solved

$$R(t) = (\tfrac{9\varpi}{4})^{\frac{1}{3}}\, t^{\frac{2}{3}}.$$

This spacetime is called the Einstein–deSitter spacetime.

$$\mathcal{G} = -dt^2 + (\tfrac{9\varpi}{4})^{\frac{2}{3}} t^{\frac{4}{3}}\big(d\chi^2 + \chi^2\, d\Omega^2\big).$$

In this spacetime dust motes follow lines χ, θ, ϕ all constant. We have t measuring proper time along their world lines. Do we need to check that these curves are geodesics? No. That is guaranteed by the self–consistency of the field equations, which guarantees that $T^{\mu\nu}{}_{;\nu} = 0$.

Because this universe has a singularity with infinite density a finite amount of time in its past, it has a big bang in it.

We can play with this spacetime fruitfully. The space part looks like a form of spherical polars, so try the coordinate transformation

$$\begin{aligned} Z &= \chi \cos\theta, \\ X &= \chi \sin\theta \cos\phi, \\ Y &= \chi \sin\theta \sin\phi. \end{aligned}$$

Then the metric is

$$\mathcal{G} = -dt^2 + \left(\tfrac{9\varpi}{4}\right)^{\frac{2}{3}} t^{\frac{4}{3}} \big(dX^2 + dY^2 + dZ^2\big).$$

Now introduce a new time variable T such that

$$\begin{aligned} dt &= t^{\frac{2}{3}}\, dT \left(\tfrac{9\varpi}{4}\right)^{\frac{1}{3}} \\ 3t^{\frac{1}{3}} &= \left(\tfrac{9\varpi}{4}\right)^{\frac{1}{3}} T, \end{aligned}$$

then

$$\mathcal{G} = \left(\tfrac{\varpi}{4}\right)^2 T^4 \big(-dT^2 + dX^2 + dY^2 + dZ^2\big).$$

The metric differs from the Minkowski metric by a simple factor. Such a metric is said to be *conformal* to the Minkowski metric, or conformally flat.

Conformal Maps

A spacetime map

$$C : (M, \mathcal{G}) \mapsto (M, \mathcal{G}')$$

is called a conformal map if we have a function Ω such that

$$\mathcal{G} = \Omega^2\, \mathcal{G}.$$

If $\mathcal{G}'$ is flat, then $(M, \mathcal{G})$ is called *conformally flat.*

A spacetime with electromagnetic field should satisfy Maxwell's equations as well. In index notation these can be written

$$\begin{aligned} \left[\sqrt{-g}\, F^{\mu\nu}\right]_{,\nu} &= 0, \\ F_{\mu\nu,\sigma} + F_{\nu\sigma,mu} + F_{\sigma\mu,\nu} &= 0. \end{aligned}$$

[Only partial derivatives! That these can be written without using the connection reflects the simplicity possible when these are written using differential forms.]

There is an important identity useful in the index notation needed here

$$\Gamma^{\alpha}_{\mu\alpha} = \frac{(\sqrt{-g})_{,\mu}}{\sqrt{-g}}$$

where $\sqrt{-g}$ uses g, the determinant of the covariant components of the metric tensor. The useful extension of our conformal map is

$$C(M, \mathcal{G}, F^{\mu\nu}) = (M, \Omega^2\mathcal{G}, \frac{F^{\mu\nu}}{\Omega^4}).$$

[C doesn't preserve solutions of the Einstein equations or we would be knee deep in exact solutions.]

which preserves solutions of Maxwell's Equations.

Since the Einstein–deSitter spacetime is conformally flat we can use all of our flat–space electrodynamics. In particular, in flat spacetime light propagates without changing its wavelength. This implies that in the conformally flat coordinates (X, Y, Z) the wavelength doesn't change. In physical coordinates this says that the wavelength goes like $R(t)$. This spacetime has a systematic redshift.

Closed Matter Universes

We have disposed of the $\kappa = 0$ case, which is pathological because of this extra symmetry. The more typical behavior is given by the open and closed universes. When we have $\kappa = 1$, the Friedmann equations reduce to

$$(R')^2 = \frac{\varpi}{R} - 1 = \frac{\varpi - R}{R}.$$

so that here the constant ϖ is the maximum size of the universe. If we introduce a dimensionless time η from

$$\frac{dt}{d\eta} = R(t)$$

then we can find a parametric solution of the Friedmann equations. This alone would justify introducing this dimensionless time, which I call arctime. But note that it is also physically natural, measuring time in terms of the angular distance around the universe travelled by a photon since the beginning.

$$R(\eta) = \tfrac{\varpi}{2}(1 - \cos\eta),$$
$$t(\eta) = \tfrac{\varpi}{2}(\eta - \sin\eta).$$

Looking at the parametric equations we see that all of these universes are similar. They differ only in their size, measured by the parameter ϖ.

Example: To verify our interpretation of arctime, look at the worldline of a radial light signal. Since it must be null, we must have

$$
\begin{aligned}
-dt^2 + R^2\, d\chi^2 &= 0, \\
\frac{d\chi}{dt} &= \frac{1}{R} \\
\frac{d\chi}{d\eta} &= 1.
\end{aligned}
$$

While it is theoretically useful to work things out in terms of the size of the universe ϖ and the current arctime η, neither of these is easily observed. The observable parameters are the current value of the local expansion rate, H, the Hubble parameter

$$H \equiv \frac{R'}{R}.$$

Its relation to the theoretical quantities is given by

$$\varpi H = \frac{2\sin\eta}{(1-\cos\eta)^2}$$

where I have written this so that both sides of the equation are manifestly dimensionless. Since the Hubble parameter has dimensions, it must involve ϖ.

The local acceleration of the expansion is also observable in principle:

$$q \equiv -\frac{R''R}{R'R'}.$$

It is dimensionless and so cannot involve ϖ

$$q = \frac{1-\cos\eta}{\sin^2\eta}.$$

A final observable is the local matter density. This involves dimensions, and thus ϖ. A related dimensionless variable uses the Hubble constant to remove the dimensions from the density

$$\Omega \equiv \frac{8\pi\rho}{3H^2} = \frac{\varpi}{R^3H^2}.$$

and in terms of arctime

$$\Omega = \frac{2(1-\cos\eta)}{\sin^2\eta}.$$

Any one of the dimensionless variables η, q, Ω can be used to tell time in the expanding universe.

In addition to being observable, the quantitites H and q are useful because expessions involving them are often identical in the open and closed universes, while expressions involving η usually require $\sin\eta \mapsto \sinh\eta$ shifts.

12. The Redshift

Our first little bit of applied geometry will be to calculate the redshift in these models. One method would be to exploit the conformal symmetry, but this only works for the $\kappa = 0$ model. A more general approach is to use the arc–time parameter.

Recall that we took the metric in the form

$$\mathcal{G} = -dt^2 + R^2(t)(d\chi^2 + S^2(\chi)\, d\Omega^2).$$

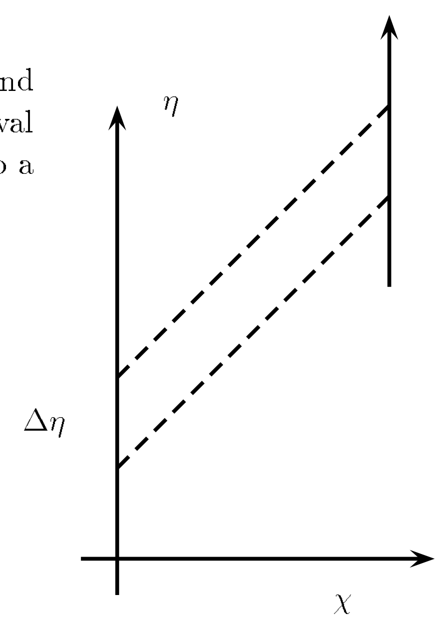
and defined a new time parameter η by

$$\frac{dt}{d\eta} = R(t).$$

This transforms the metric into the form

[Really it transforms only the chart representing the metric.]

$$\mathcal{G} = R^2(t)(-d\eta^2 + d\chi^2 + S^2(\chi)\, d\Omega^2).$$



In these coordinates light signals which stay at constant θ and ϕ, call these radial, travel along 45° lines. The coordinate interval $\Delta\eta$ between two of these worldlines is constant. This translate into a proper–time interval

$$\frac{\Delta t}{R} = \text{constant}.$$

Thus we have the wavelength going like the radius:

$$\lambda \propto R$$

In terms of the usual redshift parameter z we have

$$1 + z = \frac{R_{NOW}}{R_{THEN}}.$$

Example: A QSO photon sent out from $z = 3$ was sent from a universe which was one quarter of the present size. The cosmic microwave background radiation temperature was then 12°. If we had some sort of remote thermometer applicable to material in thermal equilibrium with that radiation, then we would be able to verify that the redshift observed in the QSO was indeed the same as the redshift that we have calculated here.

Example: Consider the Hubble constant, defined earlier by

$$H_0 \equiv \frac{R'}{R},$$

[In cosmology the subscript 0 usually refers to the present time, not the origin of time.]

For small redshift we have

$$z \sim \frac{\Delta\lambda}{\lambda} = \frac{R'\,\Delta t}{R} = H_0\,\Delta t$$

This gives us another interpretation of the Hubble parameter: it is the dimensioned constant that converts the dimensionless measure of distance z to proper distance. In the same way, we can think of R as the dimensioned parameter that converts the dimensionless distance parameter χ to proper distance. Thus

$$(z/\chi) = H_0 R_0.$$

Conformal Symmetry

These expanding universes are always similar. As they expand, only the scale of length changes. We don't expect an ordinary Killing vector in the time direction, but there is an analog which describes conformal symmetry as opposed to isometry. A conformal Killing vector satisfies

$$k_{\mu;\nu} + k_{\nu;\mu} = \phi\, g_{\mu\nu}$$

where ϕ is any scalar function.

For a particle moving with 4–velocity λ^μ we defined the conformal quantity

$$W \equiv k^\alpha\,\lambda_\alpha.$$

Then it is easy to calculate that

$$\frac{dW}{ds} = k_\alpha a^\alpha + \tfrac{1}{2}\phi(\lambda_\alpha\,\lambda^\alpha).$$

Here a^μ is the proper acceleration, and s is proper time.

Do the Friedmann universes have a conformal Killing vector? To calculate this we convert the equation into non–covariant form

$$g_{\mu\nu,\alpha}\, k^{\alpha} + g_{\mu\alpha}\, k^{\alpha}{}_{,\nu} + g_{\nu\alpha}\, k^{\alpha}{}_{,\mu} = \phi\, g_{\mu\nu}.$$

Let us try the vector $k = \frac{\partial}{\partial\eta}$. Because the components are constant, it is easy to plug this into the above equation. We need only consider

$$g_{\mu\nu,\alpha}\, k^{\alpha} = \phi\, g_{\mu\nu}.$$

For any component of $g_{\mu\nu}$ we have

$$\frac{\partial}{\partial\eta} = 2RR'\frac{dt}{d\eta}(\frac{g_{\mu\nu}}{R^2})$$

so the answer is yes, this is a conformal Killing vector with

$$\phi = \frac{2R'}{R}\frac{dt}{d\eta} = 2R'.$$

[The prime here is derivative with respect to t.]

Using this conformal Killing vector, we compute the photon redshift seen by an observer at rest in the universe. For a photon, $\lambda\cdot\lambda = 0$ and so W is constant. An observer with 4–velocity u^{μ} will see a frequency

$$2\pi\nu = u\cdot\lambda$$

and u is proportional to k

$$u = \frac{\partial}{\partial t} = \frac{k}{R}$$

so that

$$2\pi\nu = \frac{k\cdot\lambda}{R}$$

again reproducing our result, with

$$R\nu = \text{constant}.$$

Radial Geodesics

The Lagrangian for geodesics is just the metric "divided" by ds^2:

$$\mathcal{L} = R^2\left[(-\frac{d\eta}{ds})^2 + (\frac{d\chi}{ds})^2\right].$$

The Euler–Lagrange equations are

$$\frac{d}{ds}(-2\dot{\eta}R^2) = 2R\dot{R}(-\dot{\eta}^2 + \dot{\chi}^2)$$
$$\frac{d}{ds}(-2\dot{\chi}R^2) = 0$$

For these radial geodesics, the χ–momentum is conserved

$$\lambda_\chi = \text{constant}.$$

If in addition it is a massless particle, then

$$\lambda_\chi = \lambda_\eta.$$

For particles with mass we have the orthonormal frame

$$R\,d\eta \qquad \text{and} \qquad \frac{\partial}{R\partial\eta}$$
$$R\,d\chi \qquad \text{and} \qquad \frac{\partial}{R\partial\chi}$$

So we write the velocity in terms of the usual speed and γ factors:

$$\lambda = \gamma\frac{\partial}{R\partial\eta} + v\gamma\frac{\partial}{R\partial\chi}$$

and we have the component

$$\lambda^\chi = v\,\gamma/R,$$

and lowering the index gives

$$Rv\gamma = \text{constant}.$$

This shows how the proper motion of a galaxy, say, with respect to the cosmic rest frame redshifts.

13. Dust Universes

For practice we do some sample calculations in dust, that is, $p = 0$, universes. In all of these, $T^{\mu\nu}{}_{;\nu} = 0$ leads to conservation of mass

$$\tfrac{8}{3}\pi\rho R^3 = \text{constant} = \varpi.$$

[Verify this by differentiation and using the Friedmann equations.]

Thus it is sufficient to solve only the first–order equation

$$(\frac{R'}{R})^2 + \frac{\kappa}{R^2} - \frac{\varpi}{R^3} = 0,$$

Again we introduce arc–time

$$dt = R\,d\eta$$

and this converts the equation to

$$\frac{dR}{d\eta} = \sqrt{R(\varpi - \kappa R)}.$$

This can be integrated

$$R(\eta) = \varpi \sin^2(\eta/2) = \tfrac{1}{2}\varpi(1 - \cos\eta)$$
$$R(\eta) = \varpi \sinh^2(\eta/2) = \tfrac{1}{2}\varpi(\cosh\eta - 1)$$

We can then solve for proper time in terms of arc–time

$$\frac{dt}{d\eta} = R = \varpi \sin^2(\eta/2)$$

so

$$t(\eta) = \tfrac{1}{2}\varpi(\eta - \sin\eta)$$
$$t(\eta) = \tfrac{1}{2}\varpi(\sinh\eta - \eta)$$

All of the models are similar; the parameter ϖ only determines the scale.

In terms of these parameters we have the relations

Parameter	Closed	Critical	Open
R	$\frac{1}{2}\varpi(1-\cos\eta)$	$\frac{1}{4}\varpi\eta^2$	$\frac{1}{2}\varpi(\cosh\eta-1)$
t	$\frac{1}{2}\varpi(\eta-\sin\eta)$	$\frac{1}{12}\varpi\eta^3$	$\frac{1}{2}\varpi(\sinh\eta-\eta)$
$H\varpi$	$\frac{2\sin\eta}{(1-\cos\eta)^2}$	$\frac{8}{\eta^2}$	$\frac{2\sinh\eta}{(\cosh\eta-1)^2}$
$\frac{1}{3}\pi\rho\varpi^2$	$\frac{1}{(1-\cos\eta)^3}$	$\frac{8}{\eta^6}$	$\frac{1}{(\cosh\eta-1)^3}$
Ω	$\frac{2}{(1+\cos\eta)}$	1	$\frac{2}{(1+\cosh\eta)}$

The early stages of all of these universes resemble the critical case. You can see this by an expansion for small η.

Example: How far away are objects at $z = 2$ in an $\Omega = 2$, $H_0 = 50$ dust universe? Here 50 means 50 km/sec/mpc. 1/(50 km/sec/mpc) is 6.18×10^{17} sec. Since $\Omega > 1$ we need a closed model for the universe. From the table we have

$$1 + \cos\eta = 1, \qquad \eta = \pi/2$$

[These numbers have been chosen just to make the math easy, not because they are realistic.]

For an object at $z = 2$, it emitted it's light at an arc–time η given by

$$\frac{1-\cos\eta_0}{1-\cos\eta} = 1 + z = 3$$

and thus the light was emitted at a time $\eta = 0.84$ radians. This places the object at an arc–distance of 0.73 radians.

Example: Continuing the above example. If two such objects were separated on the sky by a small angle ϵ, what is their separation d in light–seconds? This is a little exercise in spherical trig. From the law for spherical triangles

$$\cos a = \cos b\, \cos c + \sin b\, \sin c\, \cos A,$$

we find the law for long skinny spherical triangles

$$a = A \sin b,$$

and so we have the arc–distance δ between the objects given by

$$\delta = \epsilon \sin \chi$$

This distance δ is converted into proper distance using the scale factor $R(\eta)$, not $R(\eta_0)$.

$$d = R(\eta_0)\epsilon \sin \chi = \tfrac{1}{3}R(\eta)\epsilon \sin \chi.$$

Finally we convert the parameter R into the somewhat better known here given parameter H using

$$HR = \frac{\sin \eta}{(1 - \cos \eta)} = 1$$

So finally

$$d = \frac{\sqrt{5}}{9}\epsilon\,(6.18 \times 10^{17}\text{sec}).$$

For an angular separation of $\epsilon = 5 \times 10^{-6}$ radians, the unit used in astronomy, the arc–second, we have a distance of 7×10^{11} light seconds, about 7 kpc in astronomical units.

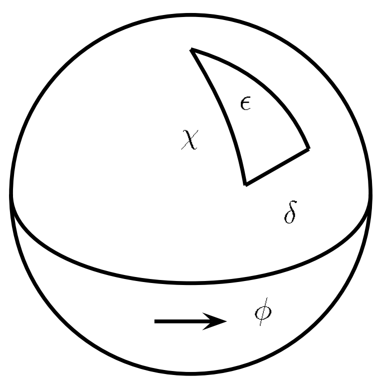


Figure 13-1. The two photon paths in a $\chi - \phi(\theta = \pi/2)$ diagram.

Example: The above example was done using what one would call synthetic spherical geometry. You could also take a super–careful approach and use analytical spherical trig. You could write the worldlines in our coordinate system. The most sensible location for us is at the pole, $\chi = 0$. This would give us a worldline given by

$$\chi = 0.$$

The angles θ and ϕ for this worldline are undetermined, since it is a singularity of the coordinate system.

Worldlines for the two objects can be taken to be

$$\begin{aligned} &\chi = \alpha, \quad \theta = \pi/2, \quad \phi = 0, \\ &\chi = \alpha, \quad \theta = \pi/2, \quad \phi = \epsilon, \\ &\alpha = \cos^{-1} \tfrac{2}{3} \end{aligned}$$

To see that this is correct, you would have to find the worldlines of the light signals, and see what angle they make with the observer. Unfortunately, the coordinate system is singular right at the observer's location. So this approach is really a lot of trouble and inferior to the direct synthetic approach.

Causality in the Early Universe

It is surprising that the early universe expands with a speed greater than the speed of light, $R' > 1$. Locally this is not observable, and entails no violence to Special Relativity. Spacetime is everywhere locally Minkowski, except at the singularity itself. Still it is disturbing. Look at a slice in the radial direction through any of these Friedmann universes.

In the figure the light signals follow dotted lines. If we pick a time, then there is a minimum separation beyond which the points have had time to exchange influences. Let us calculate this for the case of the cosmic microwave radiation. We take it to be formed at a redshift $z = 1000$, based solely on temperature arguments.

[This closed universe would require an enormous amount of missing mass, and it doesn't really live long enough. It is just for fun.]

If we stay with our closed universe then we have the arc–time of decoupling

$$\frac{1}{(1 - \cos \eta)} = 1001$$

[Unified theory anyone?]

and this gives an arc–time of about 2.5°. The spherical trig is now easy. The spherical triangle is again a long skinny triangle, and so the angle on the sky is also 5°.

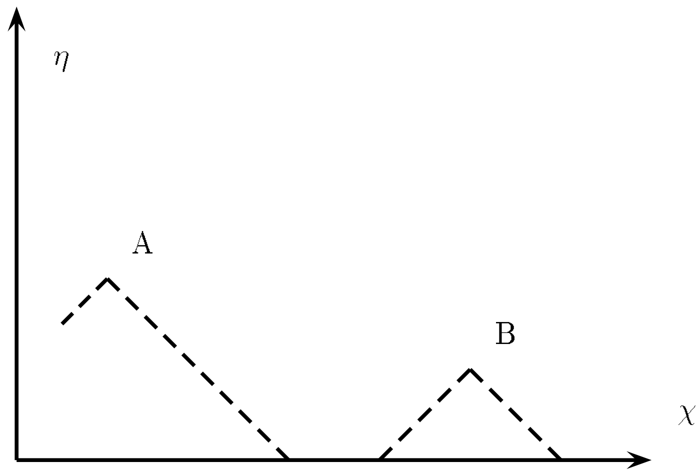


Figure 13-2. Points in the universe that have never had causal contact.

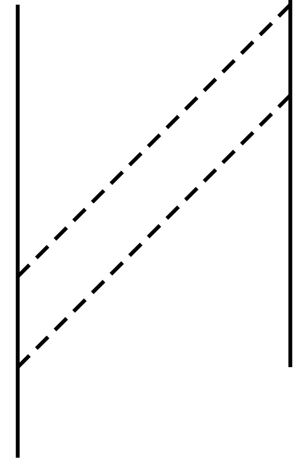
14. Inverse Square Law

Given a cosmological model, one must then decide what observable phenomena are explained by the model. Things like $R(t)$ are not directly observable. In this section I look at the basic relation between the luminosity of a source and the energy flux received. Were sources of known luminosity available, then we would be able to fix our epoch in a given cosmological model. In the idiom: to measure q_0.

Source Properties

We describe the source of radiation by its spectral energy density. In old books on relativity they only discussed the total energy flux from a source, in accord with measurement technology of the time. This spectral energy density we call L_ν, and it has units of energy per second per hertz.

There are a number of reasons, not all compelling, to prefer spectral energy density to energy per unit wavelength interval. In any event, it is easy to modify these results.



Geometric Description

We are going to describe this situation using analytic geometry in a Robertson–Walker cosmological model. Since the model is homogeneous we can put the source anywhere. The easiest calculation results if we put it at the North pole, $\chi = 0$.

We are going to follow the fate of the radiation that is emitted by the source in a time interval Δt and in a band of frequencies $\Delta\nu$. Both of these bands are to be small in the calculus sense. This corresponds to a number n of photons

$$n = \frac{L_\nu \, \Delta t \, \Delta\nu}{h\nu}.$$

We make the assumption that the number of photons is conserved. This is the adiabatic invariant for the electromagnetic problem, as well as many mechanical problems. We are neglecting absorbtion. The frequency of the photons decreases according to

$$\frac{\nu}{\nu_0} = (1+z) = \frac{R_0}{R}.$$

At their reception, the photons which cross a standard area in a standard time within a specified spectral band are counted. This is the flux density F_ν, with units of energy per sec per area per hertz. Let us collect n_0 photons over a time Δt_0, over a spectral band $\Delta\nu_0$ over an area A_0. This corresponds to a flux

$$F_{\nu_0} = \frac{n_0 h\nu_0}{\Delta t_0 \Delta\nu_0 A_0}.$$

Let us think about collecting all of the photons over a full 2–sphere. Then we have $n = n_0$. Also, we have

$$\frac{\Delta\nu}{\nu} = \frac{\Delta\nu_0}{\nu_0}$$

because the redshift is linear, and

$$\frac{\Delta t_0}{\Delta t} = (1+z).$$

This is the redshift of the integration time.

For the area of the collection sphere, we have

$$A_0 = 4\pi R_0^2 \, S^2(\chi)$$

where χ is the arc–distance to the source.

Putting all of this together gives us

$$F_{\nu_0} = \frac{L_\nu \Delta t \, \Delta\nu \, h\nu_0}{h\nu \, \Delta\nu_0 \, \Delta t_0 \, A_0}$$

and this simplifies to

$$L_\nu = (1+z) 4\pi R_0^2 \, S^2(\chi) \, F_{\nu_0}.$$

Now χ, remember, comes from

$$(1+z) = \frac{R(\eta_0)}{R(\eta_0 - \chi}$$

where we need to know the epoch η_0 of reception.

Hubble Diagram

This is a generic name for the observed flux, corrected to constant emitted luminosity, plotted against redshift. Let me present my own unorthodox way of plotting this data.

The data should be presented in such a way as to make the effects of the uncertainties as clear as possible. The intrinsic luminosity L_ν of the sources is not well known; we may be able to correct individual sources to a "standard candle", but the absolute level will be poorly known, as is R_0. All of this leads to a multiplicative uncertainty in the observed flux. To display this, we plot the log of the flux. This converts the uncertainty to a ridgid translation, something the eye does quite naturally.

A prominent part of the Hubble diagram is just the inverse square law. Many workers have published diagrams of $\log z$ against $\log d$ to see this as a linear effect. My approach to this is that believing in the inverse square law of optics for small distances, to remove it from the data and look at the residual. You could call this the inverse square law anomaly.

At the other end of the distance scale, there is a redshift singularity at $z = 0$ that can also be removed. Thus we are led to the anomaly parameter defined by

$$\phi_\nu \equiv \tfrac{1}{2} \log_{10} \frac{z^2 F_{\nu_0}}{(1+z)},$$

with

$$\frac{\nu}{\nu_0} = (1+z).$$

This ϕ_ν is an observational parameter sensitive to the geometrical effects of the universe that appear in the departure from the inverse square law of optics.

The function ϕ_ν can be computed from observational data. How is its theoretical shape predicted? For any Robertson–Walker model we have

$$F_{\nu_0} = \frac{L_\nu}{(1+z)4\pi R_0^2 S^2(\chi)}$$

and so

$$\phi_\nu = \tfrac{1}{2}\log\left[\frac{z^2 L_\nu}{(1+z)^2 4\pi R_0^2 S^2(\chi)}\right]$$

that is

$$\phi_\nu = \log\left(\frac{z}{(1+z)S(\chi)}\right) \quad + \quad \text{constant.}$$

The constant contains the multiplicative uncertainty referred to earlier.

As $z \to \infty$ we have $z/(1+z) \to 1$ and $\chi \to \eta_0$ so ϕ_ν goes to some constant. Because of this the curves for large z contain little information, and so we plot ϕ against the variable $z/(1+z)$, which plots the entire universe between zero and one.

Example: In a $q_0 = 1$ dust universe, we have

$$\eta_0 = \pi/2$$

and this leads to

$$\frac{z}{(1+z)} = \sin\chi.$$

Thus the plot of ϕ is a horizontal line for this model.

15. Lemaître Universes

Einstein thought that the idea of a closed, no–boundary universe was so appealling that he was willing to modify his field equations to accomodate it. Of course at that time the universe was thought to be time–translation invariant, and so he tried to make a static model of the universe. Had he troubled to check the stability of his proposed model, he would have discovered that it was unstable, and would have avoided what has achieved notoriety as his "most famous mistake".

In fact Einstein had two chances to "predict" the expanding universe ahead of time. The first came when there were no static solutions to his field equations. The second came when he failed to notice that the solution with the cosmological constant was unstable. Either of these predictions would have increased his fame even further, and increased even further the damage the he and Dirac did to the style of theoretical physics.

Einstein Static Universe

Let us take the Robertson–Walker metric with $\kappa = 1$

$$\mathcal{G} = -dt^2 + R^2(d\chi^2 + \sin^2\chi\, d\Omega^2),$$

and now R is a constant. Solving for $T^{\mu\nu}$ we find

$$8\pi p = -\frac{1}{R^2},$$
$$\frac{8\pi\rho}{3} = \frac{1}{R^2}.$$

Apparently to persist in this state requires a negative pressure. This disturbed him, and he arbitrarily put the change into the field equations instead, in what came to be called the cosmological constant term. Today we view these terms as a form of matter with a Lorentz–invariant equation of state, and consider a "dressed" form of the vacuum as a candidate for such matter.

[In any event, once considered, hard to disconsider.]

The Friedmann equations now read, in either case,

$$2\frac{R''}{R} + (\frac{R'}{R})^2 + \frac{\kappa}{R^2} + 8\pi p - \Lambda + 8\pi p = 0,$$
$$(\frac{R'}{R}^2) + \frac{\kappa}{R^2} - \frac{8\pi\rho}{3} - \frac{\Lambda}{3} = 0.$$

Consistency demands that Λ be constant. Now, and as intended, there is now a dust solution

$$R = \text{constant},$$
$$\frac{8\pi\rho}{3} = \frac{1}{R^2} + \frac{\Lambda}{3} = \frac{4}{3R^2},$$
$$p = 0.$$

and this leads to the cosmological model called the Einstein static universe. [Might be mistakes in these!]

Stability of the Einstein Static Universe

To show the instability of this solution, it suffices to show any growing perturbation. To show stability is much harder, then you have to show that every perturbation is bounded. Using divine inspiration, we consider the perturbation

$$\mathcal{G}' = -dt^2 + (R + \delta(t))^2 (d\chi^2 + \sin^2\chi\, d\Omega^2)$$

The Friedmann equations linearize to

$$\delta'' - \frac{\delta}{R^2} = 0,$$

and as advertised, this has a growing exponential solution, which implies instability. We haven't considered modes with angular assymmetry, these could be even more unstable.

Lemaître Universes

What about non–static solutions of the equations with cosmological constant? In the ordinary Friedmann equations we had one dimensioned constant and one dimensionless constant. The dimensionless constant determined the "time". We could use either q_0, or Ω_0, or η_0 to represent the evolution in scale–free form. While η was convenient in terms of light propagation, it suffers from the singular behavior as you change types in the universe. For many purposes the parameter Ω is the most convenient measure of "time". In these Lemaître universes we pick up one more dimensionless parameter. There is still only one scale to the spacetime. Now the state of the universe is two–dimensional in scale–free coordinates.

A useful second parameter relates to the energy density of the energy described by the cosmological constant

$$\lambda \equiv \frac{\Lambda}{3H^2}.$$

[Note that this is the Hubble parameter, not the Hubble constant. It is a function of time.]

The model evolves in a two–dimensional state space. The relative sizes of Ω and Λ allow you to judge the importance of the cosmological constant effects.

The evolution of the universe in this state space is given by the vector field

$$(\Omega - 2\lambda - 1)\frac{\kappa}{RH}\Omega\frac{\partial}{\partial\Omega} + (2 + \Omega - 2\lambda)\frac{\kappa}{RH}\lambda\frac{\partial}{\partial\lambda}.$$

This gives the rate of evolution with respect to arc–time. If you only want to trace out the curves without regard to the time, then you can drop the common factor of κ/RH. Evolution in the state space is shown in the following figure.

[See Ehlers and Rindler, Ann. NY Acad. Sci. volume unknown, for the extension of this to include radiation.]

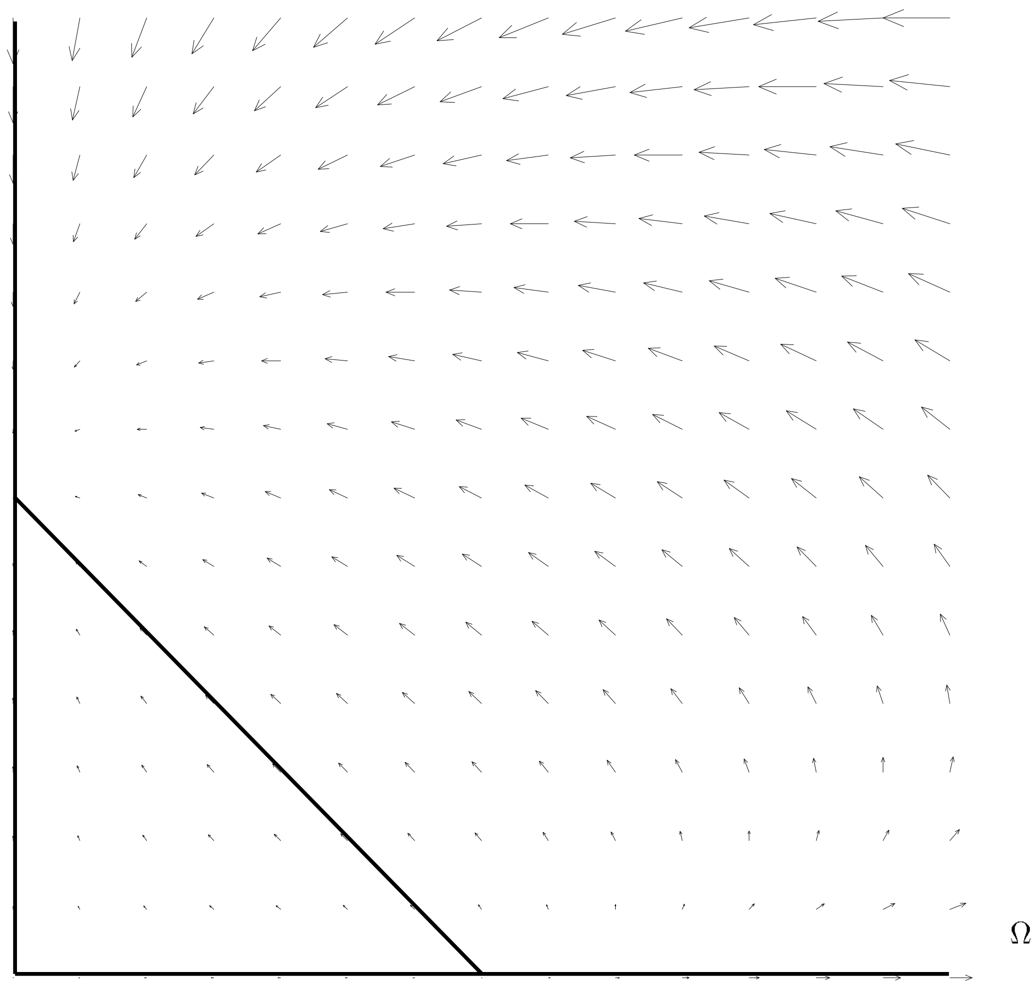


Figure 15-1. The Lemaître universe evolution.

The parameter RH is given by

$$(R')^2 = (RH)^2 = \frac{\kappa}{\lambda + \Omega - 1},$$

and you will have to be careful about the square root in a computer algorithm. The other Friedmann equation gives us

$$2RR'' = \kappa \frac{2\lambda - \Omega}{\lambda + \Omega - 1}$$

These two equations let you compute all other quantities of interest. The deceleration parameter is given by

$$q = -\frac{RR''}{R'R'} = \frac{\Omega}{2} - \lambda.$$

All solutions crossing the line $\Omega = 2\lambda$ have an inflection point.

The $\kappa = 0$ models lie along the line

$$\Omega + \lambda = 1.$$

Thus closed models are outside of the triangle and open models are inside. Both have inflection points.

Where are the zeroes of this vector field. points. These are the points $(1, 0)$ and $(0, 1)$. The latter point is a sink of the flow, with the universe expanding forever with finite speed and vanishing density. The former point is a parabolic Einstein–de Sitter model which is a neutral point of the flow. The Einstein static universe is nowhere to be seen. Why? It is off at infinity. Static means zero H, and we have normalized all of our variables with H^2 in the denominator.

The Dynamics

A useful model also needs the age parameter HT and the redshift. To find these we must integrate differential equations. If we define the dimensionless radius a by

$$a \equiv R/R_0$$

then we have

$$\frac{da}{d\eta} = \frac{R}{R_0}\frac{dR}{dt} = aRH,$$

and we have an expression for RH on terms of our state variables above. The redshift is given by

$$1 + z = 1/a$$

since we have defined $a_0 = 1$. The above equation must be integrated from the present time backward.

Similarly, we can find the time since the big bang, if there was one, by integrating backward

$$\frac{d}{d\eta}(H_0 T) = RH_0 = y(H_0 R_0)$$

and then setting the constant of integration so that the initial time is zero. You could cover state space with a plot of the dimensionless variable HT in this way.

Radiation Universes

You can also add radiation into the model. This is not complicated as long as the radiation does not couple with the matter. Unlike the matter, the energy density of this radiation goes like

$$\rho_\nu R^4 = \text{constant}$$

This means that even a small amount of radiation today, if it came from the distant past, had a more significant contribution to the dynamics of the universe then. Said forwards, even a significant amount of radiation in the past would be redshifted into insignificance today. We can describe the energy density of the radiation by a similar parameter

$$\omega \equiv \frac{8\pi\rho_\nu}{3H^2}$$

Our state space is now three dimensional, and the dynamics is given, up to a normalization, by the vector field

$$(-1+\Omega+2\omega-2\lambda)\Omega\frac{\partial}{\partial\Omega}+(-2+\Omega+2\omega-2\lambda)\omega\frac{\partial}{\partial\omega}+(2+\Omega+2\omega-2\lambda)\lambda\frac{\partial}{\partial\lambda}.$$

[See Ehlers and Rindler.]

16. Gravitational Collapse

As an example of this physical situation, we work out the collapse of a cloud of dust. We will do this by patching together a uniform dust interior and an empty outside solution.

The interior solution will be a piece of one of our closed, dust–filled cosmological models

$$\mathcal{G} = -dt^2 + R^2(t)(d\chi^2 + \sin^2\chi\, d\Omega^2).$$

We will cut a piece out of this, of radius χ_0. The cut will be a 3–surface, the 2–sphere crossed with a time axis.

The exterior solution will be the Schwarzschild solution that we studied earlier for the orbit problem.

$$\mathcal{G} = -(1 - \frac{2m}{r})dt^2 + \frac{dr^2}{(1 - \frac{2m}{r})} + r^2\, d\Omega^2.$$

This is not in coordinates that have any simple relation to those used in the cosmological model. Those were comoving coordinates; here we have instead $\frac{\partial}{\partial t}$ being the generator of time shift symmetry.

To patch pieces of these two solutions together we require that the acceleration of the surface points be the same in the two spacetimes, and that the curvature of the surface be the same as seen from the two sides. Here the surface will be a geodesic, and the curvature will require that r as a measure of the area of the spherical surfaces be continuous.

Interior Solution

We take a piece of our cosmological solution, given in parametric form by

$$R(\eta) = \frac{\varpi}{2}(1 - \cos\eta),$$

and the actual radius of curvature of the edge at an arc–distance of χ will be

$$r(\eta) = \frac{\varpi}{2}\sin\chi_0(1 - \cos\eta).$$

Along the line where we want to join the two solutions the proper time is given by

$$t(\eta) = \frac{\varpi}{2}(\eta - \sin\eta).$$

We are only considering matter with no pressure, so this surface will be a geodesic, even when we remove the surrounding material.

Exterior Solution

Can we find a geodesic in the Schwarzschild geometry with these properties. One way would be to just plug in the above and check, except we would not know what time parameter to use. If we just proceed systematically, the radial geodesics are constrained by two equations:

$$\lambda_t = \text{constant},$$
$$\lambda \cdot \lambda = -1.$$

If we call the radial and temporal components of λ $\dot{r}$ and $\dot{T}$, then we must have

$$(1 - 2m/r)\dot{T} = E, \quad \text{constant},$$
$$-(1 - 2m/r)\dot{T}^2 + \frac{\dot{r}^2}{(1 - 2m/r)} = -1$$

and these can be combined to give an equation that looks just like the Friedmann equation

$$(\frac{\dot{r}}{r})^2 + \frac{(1 - E^2)}{r^2} - \frac{2m}{r^3} = 0.$$

These will have the same solution provided we choose the constants so that

$$r = R\sin\chi_0,$$

that is,

$$\frac{1 - E^2}{\sin^2\chi_0} = 1,$$
$$\sin\chi_0 = \sqrt{1 - E^2}.$$

and

$$\frac{2m}{\sin^3\chi_0} = \varpi,$$
$$m = \frac{\varpi}{2}\sin^3\chi_0.$$

This provides us with a solution for the collapse of a dust cloud. Unfortunately for our logical development, the dust cloud collapses right through the radius $r = 2m$, beyond which we have not yet, in this class, found the correct solution. We defer this to a later section. Everything seems to be ok outside of $r = 2m$, and we don't seem to be able to avoid the difficulty of passing through $r = 2m$. A galaxy of stars behaves pretty much like dust. This cloud collapsing under its self–gravity, we would need to remove all of its angular momentum for this to happen, would pass through $r = 2m$ before the stars started to overlap.

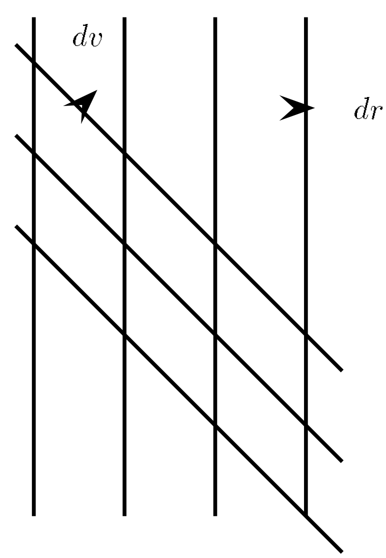
Must everything collapse? Of course not, we are surrounded by examples of objects which hold themselves up with pressure gradients.

Eddington–Finkelstein Coordinates

The first step toward straightening out the final details of gravitational collapse was the coordinate system developed by Eddington and independently by David Finkelstein. Roughly speaking, the idea is to straighten out at least one set of null geodesics so they travel along 45° lines. When we study the full solution of this problem, Kruskal coordinates, we will see that what he did was to straighten out both sets of null geodesics, incoming and outgoing, at the same time.

The idea is to introduce an *advanced time* parameter, like $t + r$ in flat spacetime. If we look at the equation for radial null geodesics

$$\frac{r\,dr}{r - 2m} = -dt,$$



this is easy to integrate

$$(r - 2m) + 2m \log(r - 2m) = -t + v,$$

and this integration constant v is our advanced time:

$$v = t + r - 2m + 2m \log(r - 2m).$$

This leads to a metric

$$\mathcal{G} = -\frac{r - 2m}{r} dv^2 + 2\,dv\,dr + r^2\,d\Omega^2.$$

This is best represented in oblique coordinates.

The metric is now well behaved at $r = 2m$, showing that it was indeed just a coordinate singularity. This now leads to a physically reasonable picture of the gravitational collapse of our dust cloud.

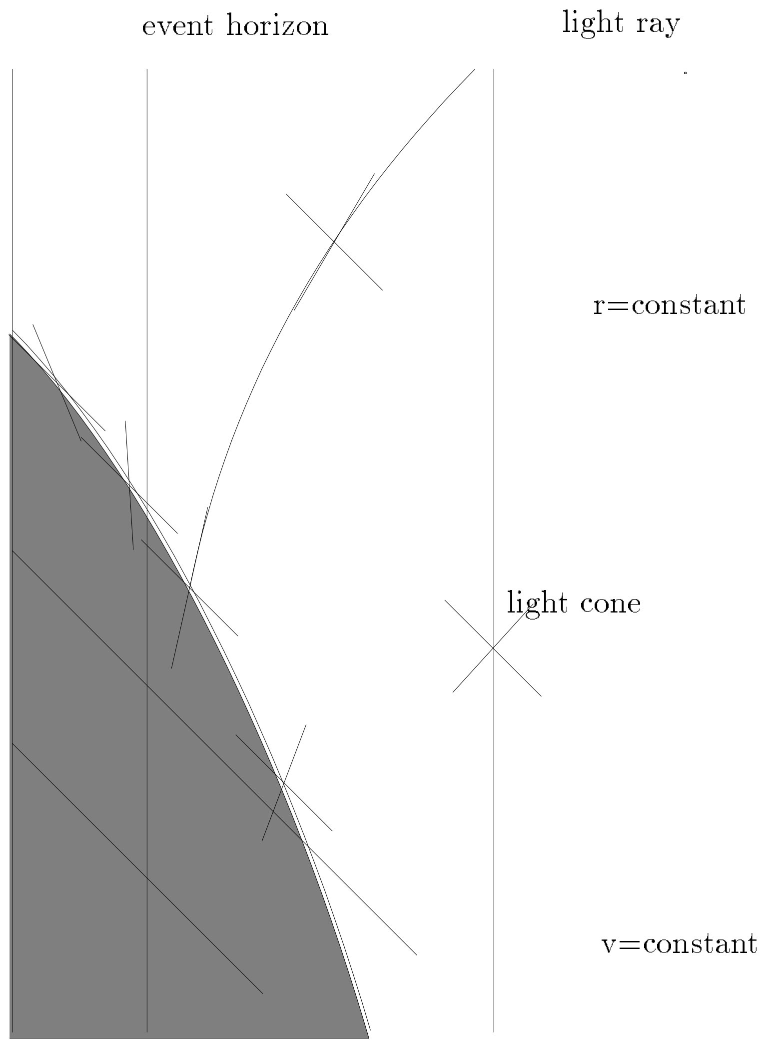


Figure 16-1. Eddington–Finkelstein coordinates.

17. Relativistic Stellar Structure

We start by discussing Birkhoff's theorem (1923). Any spherically symmetric, empty Einstein spacetime has one other translational Killing vector. The spacetime is either

$$\mathcal{G} = -(1 - \frac{2m}{r})dv^2 + \frac{du^2}{(1 - \frac{2m}{r})} + r^2\, d\Omega^2,$$

with

$$r > 2m, \qquad r(u, v) = u,$$

or the spacetime

$$\mathcal{G} = -\frac{dv^2}{(\frac{2m}{r} - 1)} + (\frac{2m}{r} - 1)du^2 + r^2\, d\Omega^2,$$

with

$$r < 2m, \qquad r(u,v) = v.$$

[There is a discussion of this in W.B. Bonnor's article in Recent Developments in General Relativity, pg 167.]

What this means, roughly, is that spherical symmetry is enough to ensure that the spacetime is static, without separately requiring it. Another way to appreciate the content of the theorem is to consider any spherical object. If you go inside it and do anything you wish that does not disturb the spherical symmetry, then the external gravitational field cannot change. This is consistent with our earlier result that there is no monopole gravitational radiation. Yet another meaning for the theorem is than inside of a spherically symmetric distribution of mass there are no gravitational effects, just as in Newtonian theory, despite the nonlinearities.

Proof of Birkhoff's Theorem

We assume that the metric can be written in the form

$$\mathcal{G} = -e^{\gamma}\, dv^2 + e^{\alpha}\, du^2 + r^2\, d\Omega^2.$$

Here as before the three functions α, γ, and r are all functions of u and v. There are three types of regions, not mutually exclusive. We might have r be a suitable substitute for the spacelike coordinate u, or the timelike coordinate v, or finally, r might be constant, and not useful for a coordinate at all. This third case does not arise here, but does arise when one considers the electrically charged solution. A great deal of confusion is avoided by the systematic choice here that keeps u a spacelike coordinate no matter what.

It is routine, but tedious, to calculate the Einstein tensor for the above metric. A regular calculation can be found in Synge, *Relativity, the General Theory*, pg 270ff. This is a useful general calculation, but be careful, his Riemann tensor has the opposite sign to ours.

Translating Synge's expressions, we have

$$G^u{}_u = e^{-\alpha}(\frac{r_u^2}{r^2} + \frac{r_u\gamma_u}{r}) - \frac{1}{r^2} - e^{-\gamma}(2\frac{r_{vv}}{r} + \frac{r_v^2}{r^2} - \frac{r_v\gamma_v}{r}),$$

$$G^\theta{}_\theta = G^\phi{}_\phi =$$

$$e^{-\alpha}\big(\frac{r_{uu}}{r} + \tfrac{1}{2}\gamma_{uu} + \tfrac{1}{4}\gamma_u^2 + \frac{r_u}{2r}(\gamma_u - \alpha_u) - \tfrac{1}{4}\alpha_u\gamma_u\big),$$

$$-e^{-\gamma}\big(\frac{r_{vv}}{r} + \tfrac{1}{2}\alpha_{vv} + \tfrac{1}{4}\alpha_v^2 + \frac{r_v}{2r}(\alpha_v - \gamma_v) - \tfrac{1}{4}\alpha_v\gamma_v\big),$$

$$G^v{}_v = e^{-\alpha}(\frac{2r_{uu}}{r} + \frac{r_u^2}{r^2} - \frac{r_u\alpha_u}{r}) - \frac{1}{r^2} - e^{-\gamma}(\frac{r_v^2}{r^2} + \frac{r_v\alpha_v}{r}),$$

$$G^u{}_v = e^{-\alpha}(-2\frac{r_{uv}}{r} + \frac{\alpha_v r_u}{r} + \frac{\gamma_u r_v}{r}).$$

In these equations the subscripts stand for partial derivatives, without the usual commas:

$$\alpha_u \equiv \frac{\partial\alpha}{\partial u}.$$

We explore the regions successively. In the first case we pick

$$r(u, v) = u.$$

The $G^u{}_v$ equation gives us

$$\frac{\partial\alpha}{\partial v} = 0. \qquad (17.1)$$

From $G^u{}_u - G^v{}_v$ we get

$$\frac{\partial}{\partial u}(\alpha + \gamma) = 0. \qquad (17.2)$$

From $G^v{}_v$ we find

$$e^{-\alpha}\frac{\partial\alpha}{\partial u} = \frac{e^{-\alpha} - 1}{r},$$

and this can be integrated to

$$\partial_u\big(r(e^{-\alpha} - 1)\big) = 0.$$

Thus we have

$$e^{\alpha} = \frac{1}{1 - \frac{2m}{r}}. \qquad (17.3)$$

From equation (17.1) we know that the constant of integration does not depend on v either.

From equation (17.2)

$$e^{\gamma} = C(v)\big(1 - \frac{2m}{r})\big).$$

The metric in this region can thus be written

$$\mathcal{G} = -(1 - \frac{2m}{r})(B(v)\,dv)^2 + \frac{du^2}{(1 - \frac{2m}{r})} + r^2\,d\Omega^2,$$

with $r(u,v) = u$ and the condition $u > 2m$, from our demand that u and v always be the spacelike and timelike coordinates. We can pick a new time variable satisfying

$$dw = B\,dv,$$

and this puts the solution into the familiar Schwarzschild form. We see that without demanding it we have asymptotic flatness and no time dependence. Clearly we have the Killing vector

$$\frac{\partial}{\partial v},$$

and this Killing vector is timelike.

The constant has been picked to be m, suggestive that it is the mass of this solution. To verify that, one should go back to the orbit calculation and look at Kepler's Law for large radii.

The next region proceeds in just the same way, leading to another spacetime with a metric

$$\mathcal{G} = -\frac{dv^2}{(\frac{2m}{r} - 1)} + \big(\frac{2m}{r} - 1\big)\,du^2 + r^2\,d\Omega^2,$$

with $r(u,v) = v$ and restricted to $v < 2m$. Now we have a spacelike Killing vector.

The third region would have $r =$ constant, and this is inconsistent with the $G^u{}_u$ equation, so there are only two solutions here. Now we have the Killing vector

$$\frac{\partial}{\partial u},$$

and this Killing vector is spacelike. It is not at all clear what this solution means, or what its relation to the first solution is. It is a correct solution to the equations, however, not a meaningless brute force extension of the first solution by passing to "imaginary" time.

Relativistic Stellar Structure

Suppose we start with a spherically symmetric metric, do not require it to be static, and allow it to contain matter with pressure. It will be useful to avoid the t–ish and r–ish labels for coordinates, and write it in the form

$$\mathcal{G} = -e^{\gamma}dv^2 + e^{\alpha}du^2 + r^2\,d\Omega^2.$$

Here $\gamma, \alpha,$ and r are all possibly functions of u and v.

Example: As a first check we verify that these equations can handle flat space. That has the functions

$$\begin{aligned} \alpha &= 0, \\ \gamma &= 0, \\ r(u,v) &= u. \end{aligned}$$

Only two equations are non–trivial. We have

$$\frac{r_u}{r} = \frac{1}{r},$$

and

$$G^u{}_u = \frac{1}{r^2} - \frac{1}{r^2} = 0,$$

and so on.

Schwarzschild Interior Solution

We want this to be static, and for this it is sufficient to study the $r = u$ case. The extra freedom, the result of bitter experience, is not needed here.

The Einstein tensor is

$$\begin{aligned} G^u{}_u &= e^{-\alpha}\left(\frac{1}{r^2} + \frac{\gamma_u}{r}\right) - \frac{1}{r^2}, \\ G^\theta{}_\theta &= e^{-\alpha}\left(\frac{\gamma_{uu}}{2} + \tfrac{1}{4}\gamma_u^2 + \frac{\gamma_u - \alpha_u}{2r} - \tfrac{1}{4}\alpha_u\gamma_u\right), \\ G^v{}_v &= e^{-\alpha}\left(\frac{1}{r^2} - \frac{\alpha_u}{r}\right) - \frac{1}{r^2}, \\ G^u{}_v &= 0. \end{aligned}$$

and using Einstein's equations for a perfect fluid we have

$$\begin{aligned} e^{-\alpha}\left(\frac{1}{r^2} + \frac{\gamma_u}{r}\right) - \frac{1}{r^2} &= 8\pi p, \\ e^{-\alpha}\left(\frac{\gamma_{uu}}{2} + \tfrac{1}{4}\gamma_u^2 + \frac{\gamma_u - \alpha_u}{2r} - \tfrac{1}{4}\alpha_u\gamma_u\right) &= 8\pi p, \\ e^{-\alpha}\left(\frac{1}{r^2} - \frac{\alpha_u}{r}\right) - \frac{1}{r^2} &= -8\pi\rho. \end{aligned}$$

Example: Now we should pause and see if we can recover the empty Schwarzschild solution, to check on the equations and practice with the structure of the equations.

The first equation leads to the separable equation for α

$$\frac{dr}{r} = \frac{d\alpha}{1 - e^{\alpha}}.$$

This is easy to integrate if we change variables (back again) to

$$w = e^{\alpha}.$$

And we find

$$w = \frac{1}{1 - \frac{1}{cr}}.$$

This will agree with what we know provided we take the constant of integration to be $1/2m$.

[Be careful here, you need to treat $w>1$ and $w<1$ separately.]

Finally, if we subtract the first two equations we find

$$\frac{d}{dr}(\alpha + \gamma) = 0.$$

Since they should both vanish at infinity, we find

$$\gamma = -\alpha,$$

and we have recovered the Schwarzschild solution.

Back to our search for an interior solution. With a little manipulation we can put the first equation into the form

$$\frac{d}{dr}\big(\frac{r}{2}(1 - e^{-\alpha})\big) = 4\pi r^2 \rho.$$

In classical stellar structure one defines a useful function which is the "mass interior" function:

$$m(r) \equiv \int_0^r 4\pi \rho r^2 \, dr.$$

This lets us solve for α

$$e^{-\alpha} = 1 - \frac{2m(r)}{r}.$$

We still have two unknowns, p and γ. The second equation after some manipulation leads to the equation

$$\tfrac{1}{2}\frac{d\gamma}{dr} = \frac{4\pi r^3 p + m}{r(r - 2m)}.$$

We still need the constaint that is always floating around the Einstein equations as an identity. This is going to provide the support equation, that classically relates the gravitational potential gradient to the pressure gradient. I had to fiddle around with several approaches before I found this one. Start with the second Einstein equation:

$$e^{-\alpha}(1 + r\gamma') - 1 = 8\pi p r^2.$$

Since there are only r derivatives here, I use a prime for simplicity. Now differentiate this equation with respect to r, and use the third Einstein equation to eliminate the γ'' term.

$$8\pi r^2 \frac{dp}{dr} = -\frac{r - 2m}{2}\gamma'(\gamma' + \alpha').$$

Now we know γ' from the above work. Moosh this all together to find finally the Oppenheimer–Volkoff equation

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

which, together with

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

and an equation of state

$$p = p(\rho),$$

forms a basis for computing relativistic stellar models.

Schwarzschild Interior Solution

An easily obtained solution, not realistic, is matter which is in equilibrium with constant density, independent of the pressure. This is incompressible matter, which of course disagrees with special relativity. Still, it is a valid solution to the Einstein equations without massive falling from virtue.

Since the density is constant, we can find the function $m(r)$

$$m(r) = \tfrac{4}{3}\pi r^3 \rho \qquad r < a,$$
$$m(r) = \tfrac{4}{3}\pi a^3 \rho \qquad r > a,$$

where $r = a$ is the surface of the star. The Oppenheimer–Volkoff equation is separable

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

[There is a continuation of this discussion in MTW, pg 610-611.]

18. Kruskal Spacetime

Can we fit together the two solutions found in the preceeding section? Then the radial coordinate would cover the entire range for zero to infinity. We want to find a covering spacetime so that in one region it looks like the first solution, and in another region, it looks like the second. Arguing against any such possibility, and this blinded people for decades, is the loss of time translation symmetry. The Killing vector in the second solution is spacelike. How is this to be reconciled?

We expect to somehow "sew together" the two solutions along the $r = 2m$ line. But things are a bit fishy with that "line", and closer examination shows that it is not a line at all, but a point that is just pretending. The real line that belongs there is hidden off at infinity in these coordinate charts. To study this possibility, which is basic to all of this, we study first the analogous problem in Euclidean geometry.

Polar Coordinates

Look at the two–dimensional space with metric

$$\mathcal{E} = du^2 + u^2\, dv^2 \qquad u > 0.$$

over the region (open set)

$$\begin{aligned} u &> 0, \\ 0 < v &< 2\pi. \end{aligned}$$

This could be described by the orthonormal frame

$$\begin{aligned} \omega^u &= du, \\ \omega^v &= u\, dv. \end{aligned}$$

The metric figures are ellipses that fit inside these unit 1–forms. See the first figure.

The game we wish to play is to pretend that we do not know what space is really represented by the above metric. Instead we try to discover what is going on only from internal evidence. The first

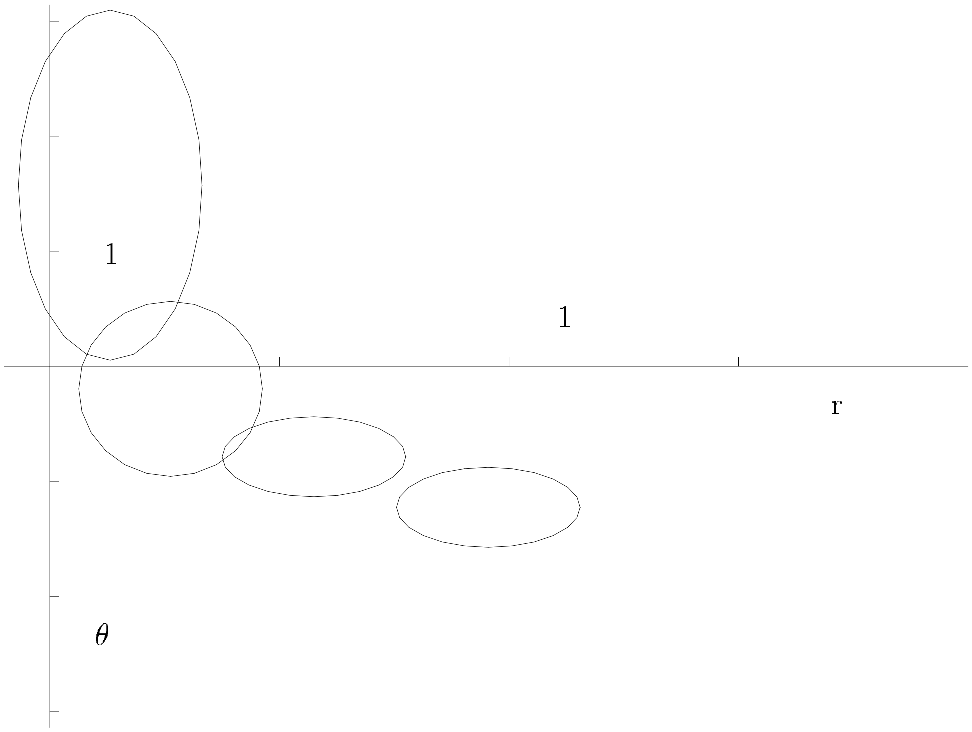
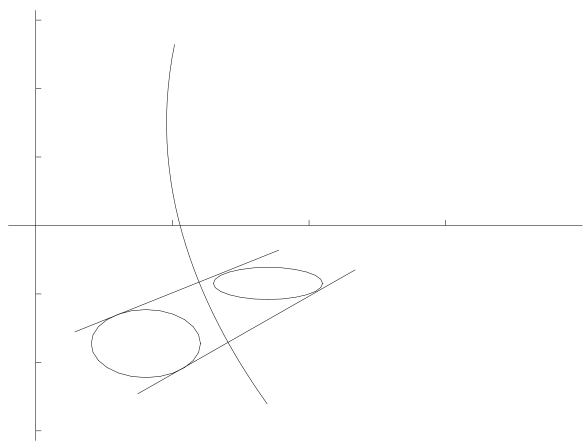


Figure 18-1. The metric of Euclidean space in polar coordinates.

Figure 18-2. Some geodesics sketched in polar coordinates.



observation is that the spacetime has an edge at $u = 0$. But, we can reach this edge in a finite distance. Just look at the metric figures. If we sketch geodesics in the metric, however, we see that they all avoid this edge. See figure 2.

Finally we notice the metric figures blowing up in the vertical direction as we approach the edge. This is a clue that a single point has been expanded and is now pretending that it is a line.

Let us make a change of coordinates designed to be poorly behaved in such a way that the bad behavior of the $u = 0$ "line" is compensated. We try a map ψ to new coordinates (x, y)

$$\psi(u, v) = (u\ \cos v, u\ \sin v) = (x, y).$$

This map is not a diffeomorphism for $u = 0$. The metric in the new coordinates can be found from the transformation of 1–forms. But remember, 1–forms can be pulled back, not pushed forward. The pull backs are

$$dx \mapsto \cos v\, du - u \sin v\, dv,$$
$$dy \mapsto \sin v\, du + u \cos v\, dv.$$

Since the coordinate transformation is a diffeomorphism over the original coordinate chart, we can invert these to

[Remember that a coordinate chart does not contain its edges. It is an open set.]

$$u\, du = x\, dx + y\, dy,$$
$$u^2\, dv = x\, dy - y\, dx.$$

These come from differentiating

$$x^2 + y^2 = u^2,$$

and

$$\frac{y}{x} = \tan v.$$

The result of this is that the metric in the new coordinates is given by the usual Euclidean metric

$$\mathcal{E} = dx^2 + dy^2.$$

This is no surprise, but what coordinate region is covered by this chart? The strip in the right half–plane in (u, v) space maps into the full plane minus the origin and the positive x–axis. Since the above metric in terms of (x, y) coordinates is perfectly well behaved at the missing points, a covering space is just the full plane with the same metric. Note how the "line" for $u = 0$ has been squashed into a point in the covering space. The extension of a Minkowski spacetime proceeds in a similar fashion, but with one important physical difference.

In Euclidean space we saw that geodesics avoided the "edge". This seems reasonable, because the edge is a single point, and how are we to hit it dead on? This is not a good argument, however. If you look at what goes on with a Minkowski metric, geodesics are sucked

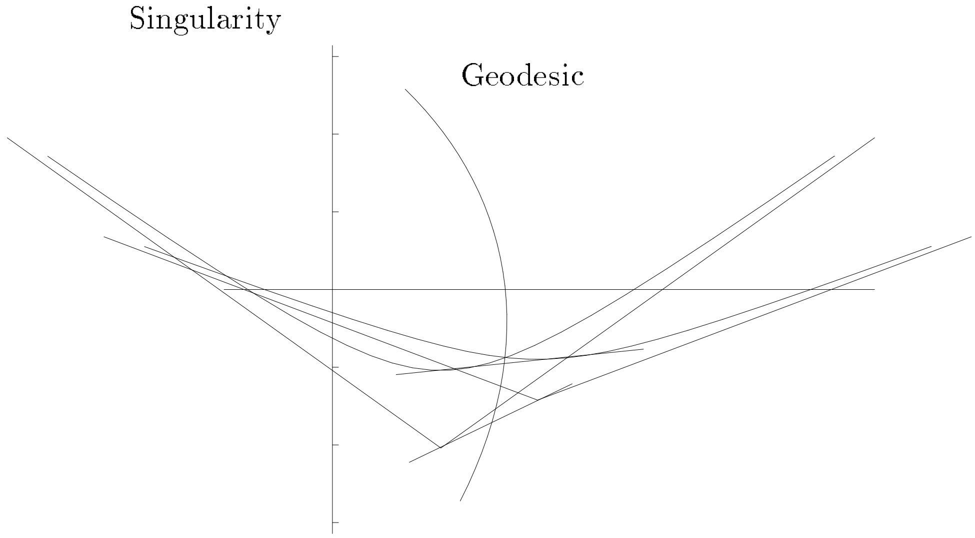


Figure 18-3. Geodesics being sucked into a coordinate singularity in spacetime.

into such a singularity rather than being repelled. This makes the issue even less ignorable. See figure 3.

Extending Schwarzschild

If we look at the metric in the $r > 2m$ region, we can sketch the metric again using an orthonormal frame. What we see is the now familiar point pretending to be a line, but now complicated by the squashing of the frames horizontally.

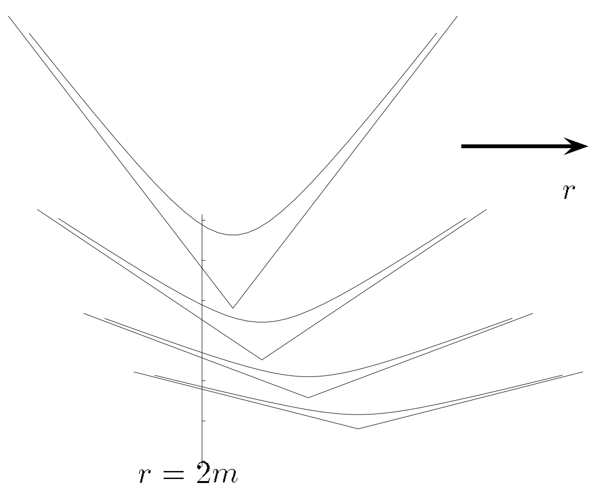


Figure 18-4. Outer Schwarzschild coordinates.

Near $r = 2m$ this has the behavior

[First we look at an approximation.]

$$\mathcal{G} \sim -\frac{u - 2m}{2m}\, dv^2 + \frac{2m}{u - 2m}\, du^2 + 4m^2\, d\Omega^2.$$

We see that there are two kinds of singularities here. The infinite coefficient in the du^2 term we will see is no problem at all. It just means that we have brought infinity in to where we can see it. On the other hand, it is the zero in the dv^2 term, which causes all the trouble.

Again we will try to rectify the situation by squashing the line $r = 2m$ down to a point. We guess a transformation in the form

$$\psi(u, v) = (f(u)\cosh\alpha v, f(u)\sinh\alpha v) = (x, y).$$

We leave the spherical coordinates unchanged. The function f is in there to handle the minor problems with the infinity in the metric coefficients. The constant α will be adjusted when we need it.

With this transformation we have

$$\begin{aligned} x^2 - y^2 &= f^2, \\ \frac{y}{x} &= \tanh\alpha v. \end{aligned}$$

and the metric in these new coordinates becomes

$$\mathcal{G} = -\frac{u-2m}{u\alpha^2 f^4}(x\,dy - y\,dx^2) + \frac{u}{(u-2m)f^2(f')^2}(x\,dx - y\,dy)^2 + r^2\,d\Omega^2.$$

To cancel the cross terms we must pick

$$\frac{f'}{f} = \frac{\alpha u}{u-2m}.$$

This can be integrated:

$$\log f = \alpha\big[(u-2m) + 2m\log\mid u-2m\mid\big].$$

and so

$$f = (u-2m)^{2m\alpha}\exp\big(\alpha(u-2m)\big).$$

The metric is now

$$\mathcal{G} = A(x^2 - y^2)(-dy^2 + dx^2) + r^2\,d\Omega^2$$

with the factor A

$$A = \frac{(u-2m)\exp\big(-4\alpha(u-2m)\big)}{u\alpha^2(u-2m)^{8m\alpha}}$$

The choice of $\alpha = 1/4m$ removes most of the junk here.

The final metric form is

$$\mathcal{G} = \frac{16m^2}{r} \exp\left[-\frac{r-2m}{2m}\right](dx^2 - dy^2) + r^2\, d\Omega^2,$$

and the function $r(x,y)$ here is found from

$$\sqrt{x^2 - y^2} = \sqrt{r-2m}\exp\frac{r-2m}{4m}.$$

Unfortunately, this is a transcendental equation.

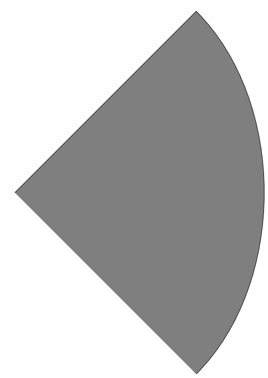
The diffeomorphism is

$$(x,y) =$$

$$\left(\sqrt{u-2m}\exp\frac{u-2m}{4m}\cosh\frac{v}{4m}, \sqrt{u-2m}\exp\frac{u-2m}{4m}\sinh\frac{v}{4m}\right).$$

The range of this in the (x,y) plane can be deduced as follows. The $r =$ constant curves are the hyperbolae

$$x^2 - y^2 = \text{constant}.$$



The boundary $r = 2m$ corresponds to

$$x^2 - y^2 = 0.$$

Thus we are in the right–hand open quadrant, from $-45°$ to $+45°$. The curves $v =$ constant go into lines through the origin:

$$y/x = \text{constant}.$$

We have indeed squashed $r = 2m$ down to a point. What is more important, though, is that we have brought in an entire line from infinity, in fact two of them, the top and bottom edges of the quadrant.

Now what? Well, you can guess that the other solution is going to be handled similarly. The end result of a similar calculation is the transformed metric:

$$\mathcal{G} = \frac{16m^2}{r} \exp(-\frac{r-2m}{2m})(dx^2 - dy^2) + r^2\, d\Omega^2.$$

Notice how the signs have all straightened themselves out, and we can insert this solution into the upper quadrant, smoothly and analytically even crossing the $r = 2m$ line. In the next figure we show how this

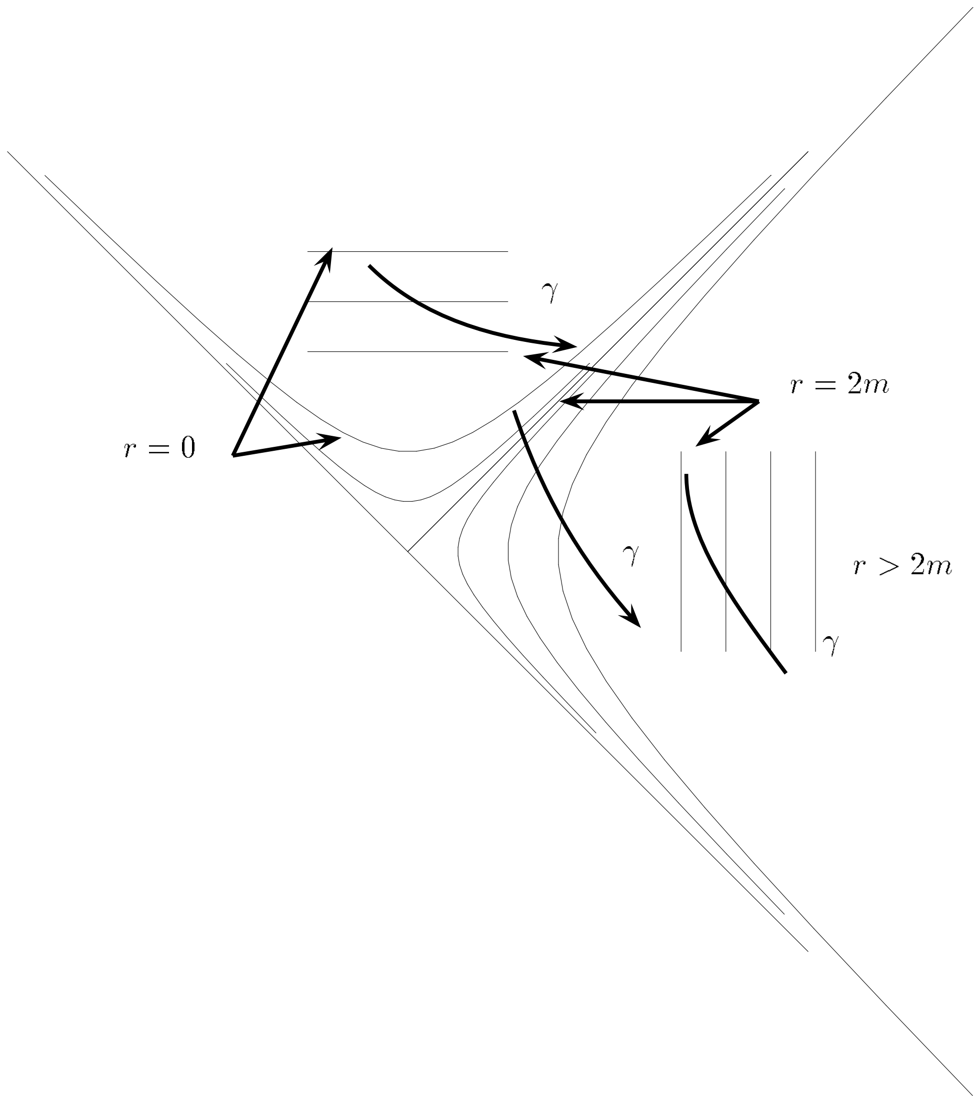


Figure 18-5. Note how pathological the curve γ appears in the original charts.

covering space reveals details of curves that were obscured by the $r = 2m$ coordinate singularity in the original charts.

Note that in these new coordinates radial light signals travel along 45° lines.

The Complete Collapsing Dust Cloud

The metric still has edges and other problems. Some of these will be eliminated by going to a more realistic physical situation. We can try to extend the collapsing dust cloud, for starters.

Note that we have cut off the dust cloud solution for times before

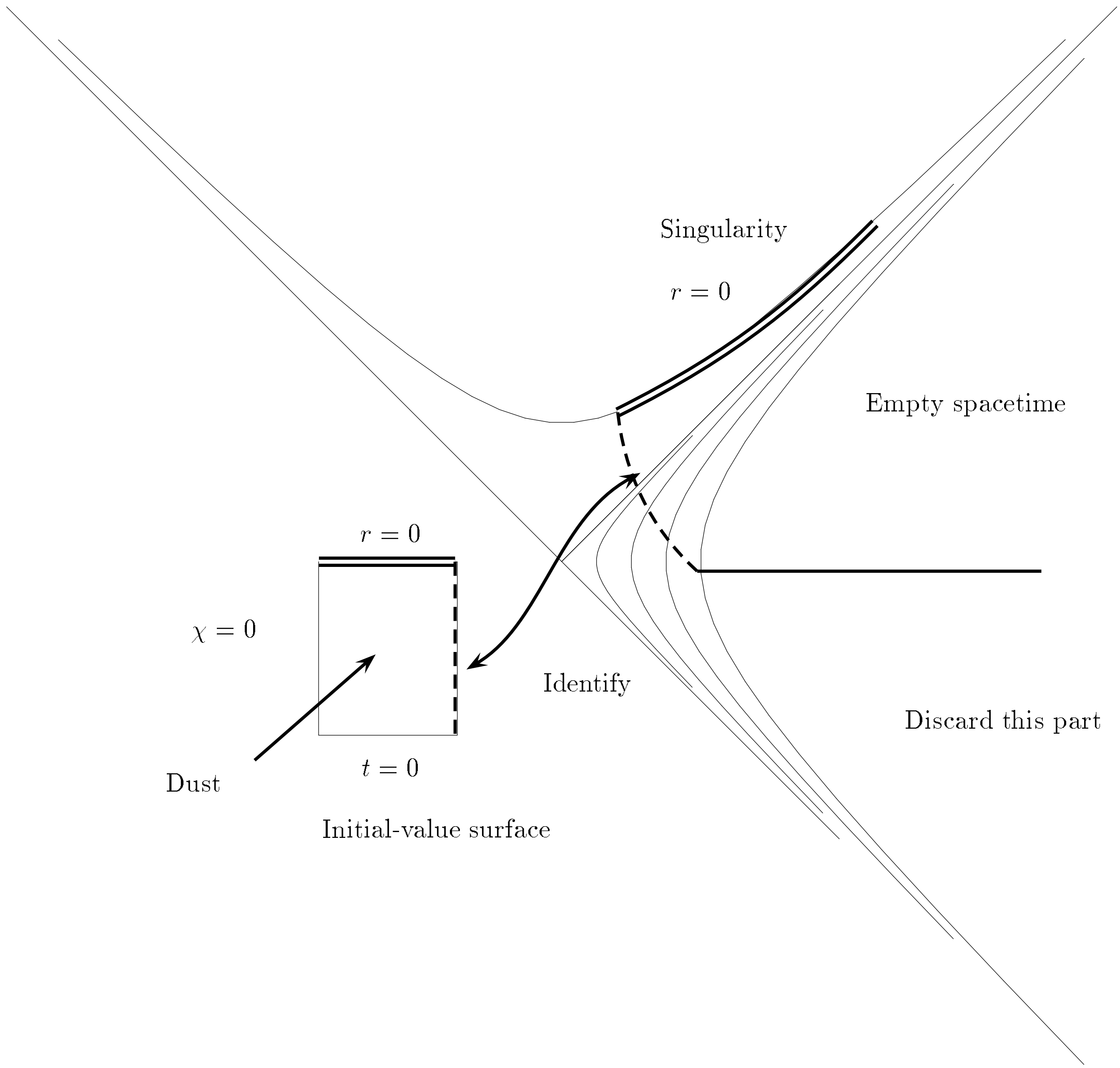


Figure 18-6. The maximal extension of the collapsing dust cloud.

the initial conditions. Otherwise we will have to answer the question: what was the dust cloud doing beforehand? How was it set up? If astrophysicists with dump trucks are involved, then their $T^{\mu\nu}$ must be included.

Bizarre Spacetimes

We can make a complete picture of the dust cloud, even into the distant past if we use the complete cosmological solution. Then the dust cloud was created in an initial singularity, a little bang if you will. To get enough of the Schwarzschild to patch it onto we need another copy of the spacetime for $r < 2m$ with a sign reversed so that it can be slipped into the bottom quadrant. It will end up being the same metric form as above. The picture for this is shown in figure 18-8.

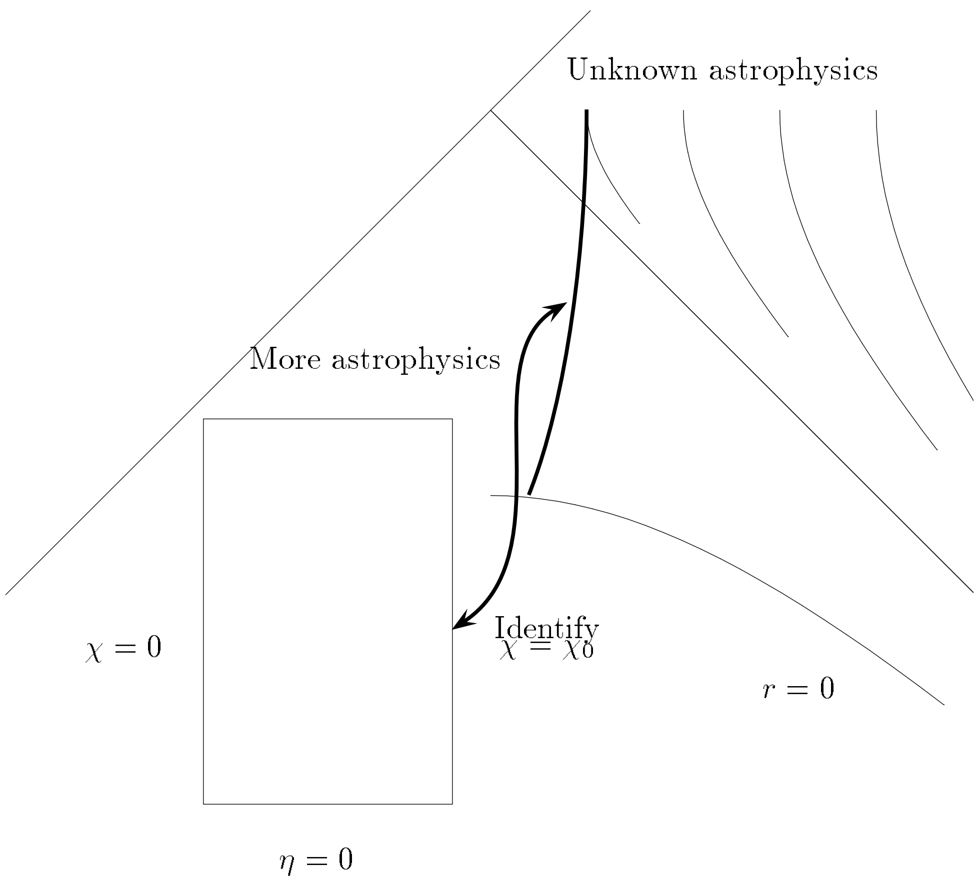


Figure 18-7. A Dust Cloud Expanding from a Little Bang.

The lower part of this diagram is often called a *white hole*. It represents a delayed exploding singularity. There was a time when such a desperate solution would be required to explain quasars. I say desperate because normal physics and normal physical laws do not tell us what is going on with a singularity.

Most Bizarre Spacetime (so far)

To fully compete the Schwarzschild solution, we need to fill in the far left quadrant. This requires another whole universe. Let's not think small.

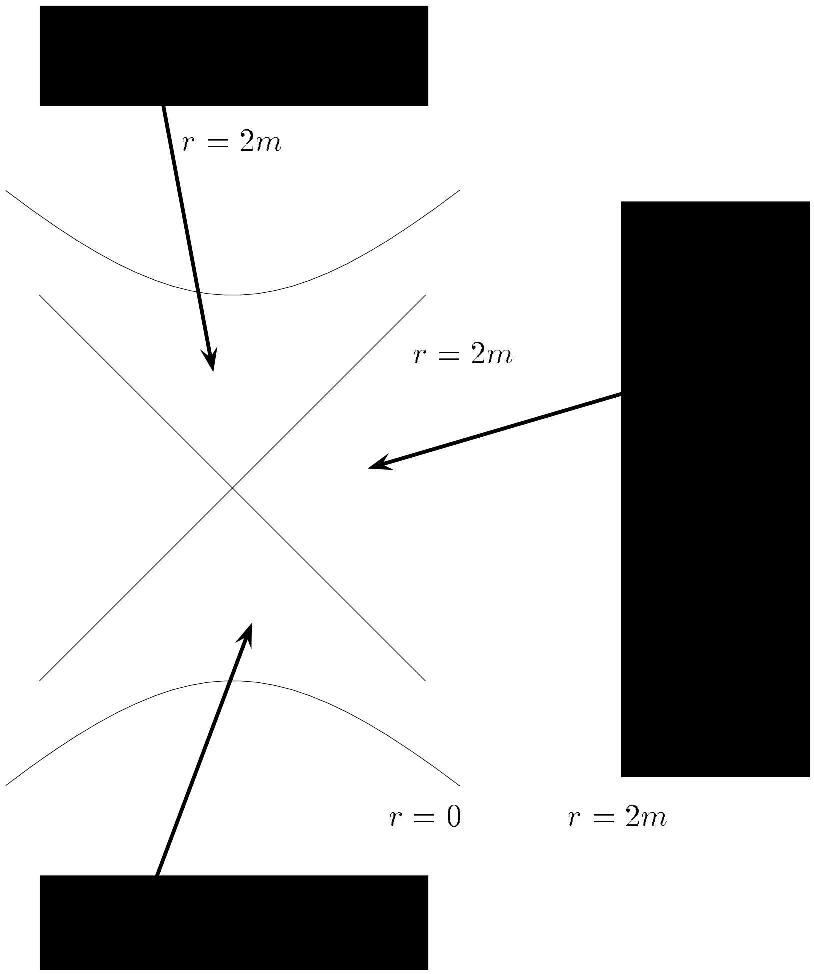


Figure 18-8. The Wormhole Spacetime.

This spacetime is often said to have a wormhole in it, a connection to another universe called an Einstein–Rosen bridge. You can see by looking at the solution that it is not possible to get through the wormhole without exceeding the speed of light.

This spacetime has further unsavory features: it has a naked singularity. From the outer, normal regions of the universe, either one, it is possible to trace a null geodesic back to a singularity. Since we do not know what initial conditions to impose on this line, we cannot predict what radiation will be seen, or what.

Event Horizon

These spacetimes have a special geometric feature called an event horizon. This concept is only defined for spacetimes which have a flat infinity tucked off somewhere. If we look at the worldline of something falling in to the future singularity, and look at light signals that it sends out, we see that after a finite amount of time, at an unremarkable point in spacetime, the light signals themselves are trapped by the singularity. The surface separating the universe into causally disconnected pieces is called an event horizon.

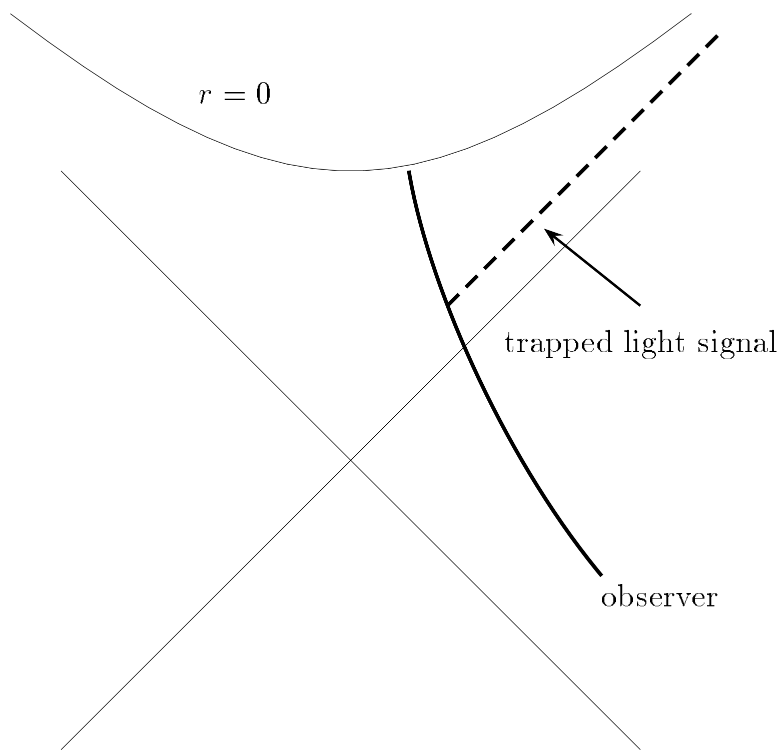


Figure 18-9. Event horizon for an observer falling into a Schwarzschild singularity.

19. Global Structure

These issues arise from a consideration of causality and the initial–value problem. The basic idea is that no information or causal influence can propagate faster than the speed of light. This follows from the hyperbolic nature of the basic equations of physics. It is not really quite so straightforward because of the complications of gauge invariance, as we saw when discussing gravitational waves.

The situation in Minkowski spacetime is sketched in figure 19-1.

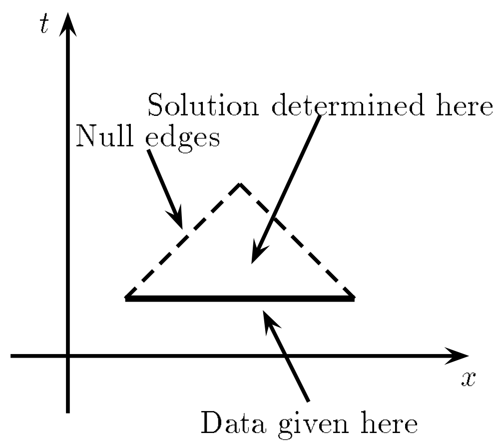


Figure 19-1. Causal structure in Minkowski spacetime.

Data given on the initial hypersurface determines the solution within a triangle with null edges. What if you have to go all the way out to infinity? When would this be important? Well, in a wave problem you might have the condition that there are no incoming waves.

If infinity is important, then you should bring it in to where you can see it. Then you can do a further step, similar to extending the Schwarzschild spacetime by finding a covering space. You can attach new points at infinity, called ideal points. This is called compactification. There is not a unique way to compactify a spacetime. Some physical properties are lost, like distance, and some preserved, like causality. You might try to preserve straight lines as well, and this would be called a projective compactification. There are no existence theorems that guarantee that any particular type of compactification is possible. The situation is quite different from the situation in Euclidean space, where compactification involves only adding a single point at infinity.

Conformal Structure

If we ignore everything about a metric except its null geodesics, then we are talking about the conformal structure of a manifold. We say an application of conformal equivalence when we discussed the redshift in an expanding universe.

If we rescale the metric by a factor that can depend upon position, then obviously null lines are transformed into other null lines. If the original lines are geodesics, then we have to check that the transformed lines are also geodesics. This is routine, and shows that while the curve remains a geodesic, the parameter on the curve will not remain a special affine parameter. That is, the final curve may only satisfy

$$\sigma^{\mu}{}_{;\nu}\sigma^{\nu} = \phi\sigma^{\mu},$$

rather than

$$\sigma^{\mu}{}_{;\nu}\sigma^{\nu} = 0.$$

This shows that the causal structure, the null geodesics, is a conformal invariant.

Conformal Minkowski Spacetime

Let us look into the equivalence class of spacetimes that are conformally equivalent to Minkowski spacetime. We start from

$$\mathcal{G} = dx^2 - dt^2,$$

and consider maps

$$\psi : (t,x) \mapsto (u,v) = (U(t,x), V(t,x)),$$

such that the pullback metric

$$\psi^*(\mathcal{G}) \propto \mathcal{G}.$$

We have

$$du = u_x\, dx + u_t\, dt,$$
$$dv = v_x\, dx + v_t\, dt.$$

The new metric will not have cross terms if we make the choice (this is not unique)

$$u_{tt} - u_{xx} = 0,$$
$$v_{tt} - v_{xx} = 0,$$
$$u_x = v_t,$$
$$u_t = v_x.$$

In this case the metric goes to

$$dv^2 - du^2 = (u_x^2 - u_t^2)(dt^2 - dx^2),$$

and they are related by a conformal factor, as we hoped.

Two–Dimensional Minkowski Spacetime

First look at an example of what we mean by compactification.

Example: Look at the infinite line, and the map

$$\theta \mapsto \tan\theta.$$

This maps the open interval $-\pi/2 < \theta < \pi/2$ onto the entire real line. This is the same trick that we want to do on Minkowski spacetime using the above type of conformal transformation.

To do this to Minkowski spacetime, we need to guess solutions of the wave equation similar in form to the above. It works to use

$$u = \tan^{-1}(t + x) + \tan^{-1}(t - x),$$
$$v = \tan^{-1}(t + x) - \tan^{-1}(t - x).$$

These are obviously solutions of the wave equation, and they also satisfy the other conditions.

We need to study the map. What happens to the t–axis?

$$(s, 0) \mapsto (2\tan^{-1} s, 0).$$

So the entire time axis is mapped into the finite open interval of the t–axis $(-\pi, \pi)$. The same thing happens to the x–axis. What about a null geodesic?

$$(s, s) \mapsto (\tan^{-1} 2s, \tan^{-1} 2s).$$

This is a 45° segment, ending at $(\pi/2, \pi/2)$. What about an off–center null geodesic?

$$(s, s+a) \mapsto (\tan^{-1}(2s+a) + \tan^{-1} a, \tan^{-1}(2s+a) - \tan^{-1} a).$$

These all have future endpoints on the line connecting the end of the t–segment with the end of the x–segment. These curves fill in a diamond–shaped region.

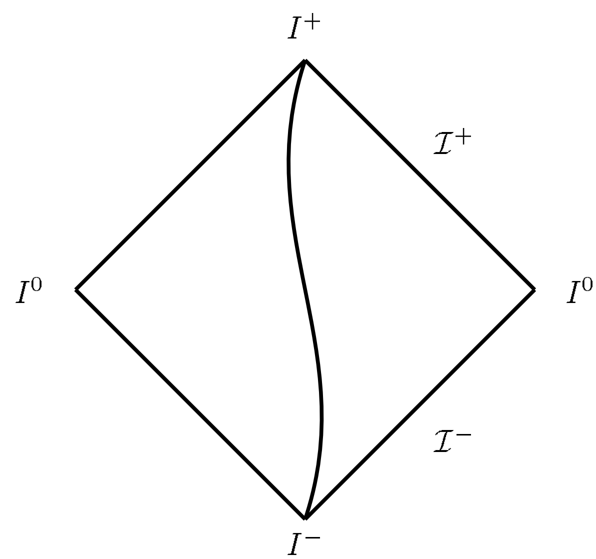


Figure 19-2. The compactification of 2D Minkowski Spacetime.

What about time–like lines?

$$(s, a) \mapsto \big(\tan^{-1}(s+a) + \tan^{-1}(s-a), \tan^{-1}(s+a) - \tan^{-1}(s-a)\big).$$

The endpoints are again at $(\pm\pi, 0)$. This point is called *future timelike infinity*.

More interesting curves are finite–velocity time–like geodesics. We plot their behavior next, considering

$$(s, vs) \mapsto \big(\tan^{-1}(1+v)s + \tan^{-1}(1-v)s, \tan^{-1}(1+v)s - \tan^{-1}(1-v)s\big).$$

These also end up at future time–like infinity.

The result of all of this is that the diamond

$$| u + v | < \pi,$$
$$| u - v | < \pi.$$

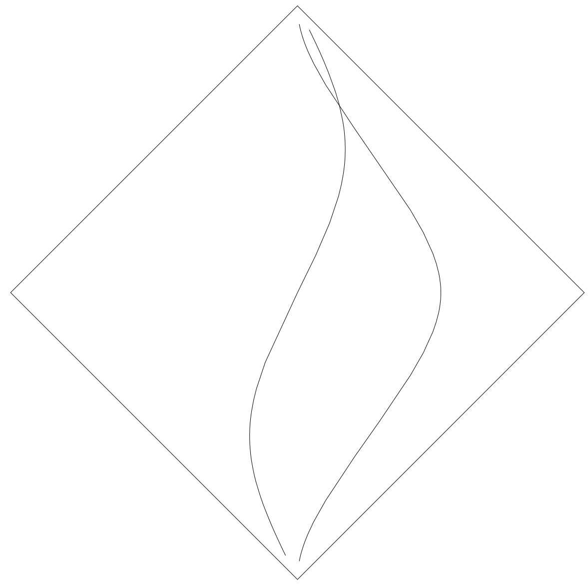


Figure 19-3. Timelike curves in 2D Minkowski spacetime.

is conformally equivalent to Minkowski spacetime.

A conformal extension can be made now by just adding in the boundary points of this open diamond. The top and bottom points of the diamond are called future and past time–like infinity. There are also two space–like infinite points. There are also four open null lines, that are called null infinity. This will change just a little bit when we go and extend four dimensional Minkowski spacetime.

Four Dimensional Minkowski Spacetime

We look at the map

$$\tau = \tan^{-1}(t+r) + \tan^{-1}(t-r),$$
$$\chi = \tan^{-1}(t+r) - \tan^{-1}(t-r),$$
$$\theta = \theta,$$
$$\phi = \phi.$$

which leads after the same sort of computation to the metric

$$\mathcal{G} = [\frac{1}{4\cos^2(\frac{\chi+\tau}{2})\cos^2(\frac{\chi-\tau}{2})}](-d\tau^2 + d\chi^2 + \sin^2\chi\, d\Omega^2).$$

Now a conformally equivalent spacetime will just eliminate the first factor. This spacetime

$$\mathcal{G} = -d\tau^2 + d\chi^2 + \sin^2\chi\, d\Omega^2,$$

is just our friend the Einstein static universe. Here we are just using it as a model spacetime in which to embed conformally the spacetime of interest. A picture of the embedding, in which the 3–spheres are represented by circles is shown next.

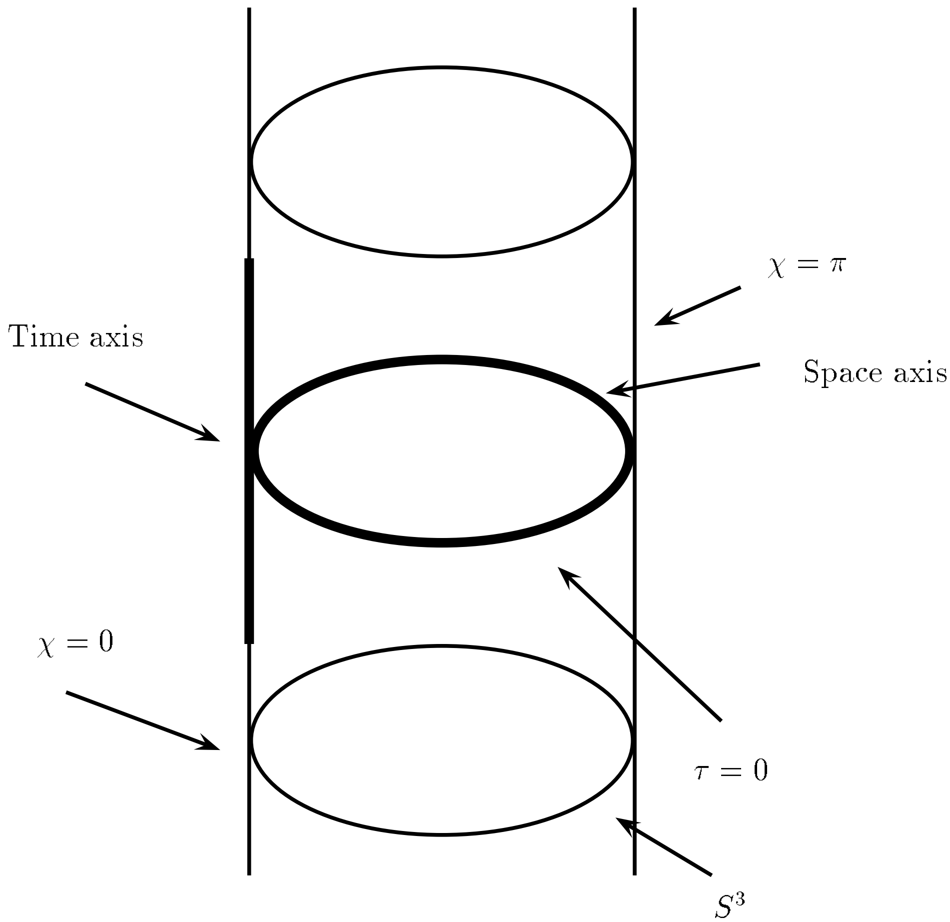


Figure 19-4. Embedding Minkowski spacetime in the Einstein static universe.

The only major change here is that spatial infinity has become a single point, called I_0 It is customary to represent this by the plane drawing in figure 19-5, which is to be “rotated over a 2–sphere, while keeping the χ axis and I_0 fixed.

This is the first view of what are often called *pocket spacetimes.* This is the tool that we need to delve into the intricacies of black holes in the next section.

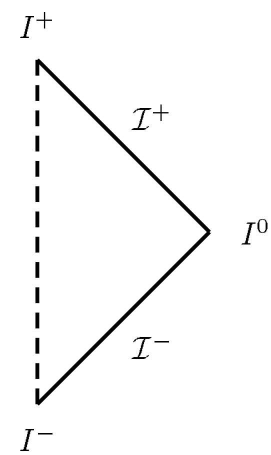


Figure 19-5. A 2D representation of compact Minkowski spacetime.

A1. Tensor Notation

For most of these notes we will use the usual tensor notation, where k^μ denotes the μ-th component of a contravariant tensor. Sometimes, however, we use the "abstract index" convention championed by Roger Penrose, and speak of "the tensor $g_{\mu\nu}$", while other times I will introduce a new symbol, and speak of "the tensor $\mathcal{G}$". The advantages of the abstract index notation lie in the automatic indication of the type of the tensor, and the nature of any contractions that are to be performed. It should be clear that

$$h_{\mu\nu} = (2m/r)(dt^2 + dr^2)$$

only makes sense if we are using the abstract index notation.

Sometimes the kernal symbol alone will be use to indicate the tensor, but only for vectors and covectors. For the above metric perturbation, one usually defines

$$h = h_{\alpha\beta}\, g^{\alpha\beta},$$

and for a metric tensor itself

$$g = \mathrm{Det}\, g_{\mu\nu}.$$

In old tensor a metric was denoted by the abstract symbol ds^2, which I happen to dislike because it is not the square of the d of any s.

If the metric tensor is diagonal, then you can easily work as if you were in an orthonormal basis by using what MTW call the "physical components". You move the indices midway between up and down, using

$$v^\mu \sqrt{g_{\mu\nu}}, \omega_\alpha \sqrt{g^{\alpha\alpha}}.$$