Lecture Notes on

General Relativity

Kevin Zhou
kzhou7@gmail.com

These notes cover general relativity. The primary sources were:

- Harvey Reall's General Relativity and Black Holes lecture notes. A crystal clear introduction to the subject. Parts of the Black Holes notes are adapted from Wald, and contain somewhat less detail but more discussion.
- David Tong's General Relativity lecture notes. A fun set of notes that takes a lot of detours, diving into all the questions one might have on a *second* pass through relativity, and emphasizing links with theoretical physics at large.
- Schutz, *A First Course in General Relativity*. An introductory book which spends its first quarter very clearly reviewing special relativity, vectors, and tensors.
- Carroll, *Spacetime and Geometry*. The canonical "friendly" general relativity book. Has either the advantage or disadvantage of moving most of the math to appendices, allowing the main text to be casual and conversational, including discussions of philosophical topics such as the meaning of the equivalence principle.
- Wald, *General Relativity*. The canonical "unfriendly" general relativity book. Covers the foundations of differential geometry and general relativity within the first 100 pages, then moves onto advanced topics such as the singularity theorems and spinors in curved spacetime.
- Zee, *Einstein Gravity in a Nutshell*. A huge, chatty book written along the same lines as Zee's quantum field theory text. Gradually moves from flat space to curved space to flat spacetime to curved spacetime throughout the first two thirds, hence introducing many important concepts multiple times. The final chapter contains interesting speculations on topics such as twistors, the cosmological constant problem, and quantum gravity.
- Mukhanov and Winitzki, *Introduction to Quantum Effects in Gravity*. Introduces QFT in curved spacetime at the undergraduate level, without even requiring QFT as a prerequisite, by seamlessly routing around the usual technical difficulties; for instance, every spacetime considered is conformally flat. Also contains enlightening conceptual discussions.

The most recent version is here; please report any errors found to kzhou7@gmail.com.

Contents

1 Preliminaries

1.1 Coordinate Transformations

We begin by establishing conventions for Lorentz transformations.

- Vector components transform as
$$x^{\mu'} = \Lambda^{\mu'}{}_{\nu} x^{\nu}.$$
In order to write this in matrix form, we consider x^μ as the elements of a column vector and Λ as a matrix with ij entry $\Lambda^i{}_j$, giving $x \to \Lambda x$.

- Generally, primes on Greek letters denote another coordinate system. We always put primes on the components, and not on the geometric objects themselves, as those don't transform.

- We denote the inverse Lorentz transform with the same letter, but with the unprimed index on top. In particular,
$$\Lambda^{\mu'}{}_{\nu} \Lambda^{\nu}{}_{\rho'} = \delta^{\mu'}_{\rho'}.$$
We always write Lorentz transformations with the first index up. Note that Λ is not a tensor; it defines a transformation between frames, not a frame-independent geometric quantity.

- Since vector/covector contractions are invariant, covector components transform "oppositely" as
$$\omega_{\mu'} = \Lambda^{\nu}{}_{\mu'} \omega_{\nu}$$
which can be written in matrix notation as $\omega \to \omega \Lambda^{-1}$, where ω is a row vector, or alternatively $\omega^T \to (\Lambda^{-1})^T \omega^T$. In matrix notation, contractions transform as $\omega v \to \omega \Lambda^{-1} \Lambda v$ as desired.

- Since vectors $V = V^\mu \hat{e}_{(\mu)}$ are invariant, the basis vectors $\hat{e}_{(\mu)}$ transform like covector components,
$$\hat{e}_{(\mu')} = \Lambda^{\nu}{}_{\mu'} \hat{e}_{(\nu)}.$$
Similarly,
$$\hat{\theta}^{(\mu')} = \Lambda^{\mu'}{}_{\nu} \hat{\theta}^{(\nu)}.$$

- For more general tensors, the same pattern holds for each index. Lorentz transformations are defined to be coordinate transformations which keep the metric components the same, so
$$\eta_{\mu'\nu'} = \eta_{\mu\nu}, \quad \eta_{\mu'\nu'} = \Lambda^{\rho}{}_{\mu'} \Lambda^{\sigma}{}_{\nu'} \eta_{\rho\sigma}.$$
This is the only equation we'll see where the indices on both sides don't match. In matrix notation this condition is $\eta = \Lambda^T \eta \Lambda$.

- The above condition also means that in special relativity, we work only with Cartesian coordinates in inertial frames. This is the common definition, though we can also say that special relativity takes place in Minkowski space under any coordinates; this allows curvilinear coordinates and noninertial frames.

- In general relativity, we consider general coordinate transformations, for which we replace
$$\Lambda^{\mu'}{}_{\nu} \to \frac{\partial x^{\mu'}}{\partial x^{\nu}}.$$

Otherwise, all transformation laws stay the same. This generalization actually makes things easier: using the chain rule, it's obvious what the inverse is, and how basis vectors ∂_μ and covectors dx^μ transform.

- A useful trick to guess expressions is to replace tensors with products of rank 1 tensors. Consider a rank 2 tensor in the product form $dx^\mu dx^\nu$. Its transformation is obvious given the transformation for dx^μ. The rule then extends to all rank 2 tensors by linearity.

1.2 Equivalence Principles

In general relativity, gravity is described as the curvature of spacetime, not as an additional field propagating through spacetime. This is motivated by the equivalence principle, explained below.

- In Newtonian mechanics, the inertial mass is defined by $F = m_i a$ and the gravitational mass is defined by $F_g = -m_g \nabla\Phi$.
- The Weak Equivalence Principle (WEP) states that

$$m_i = m_g.$$

  This implies that the behavior of freely-falling test particles does not depend on their mass.
- One consequence of the WEP is that the motion of freely-falling particles is the same in a uniform gravitational field and a uniformly accelerated frame. (For a nonuniform field, we could of course tell the two apart by tidal effects.)
- The WEP is also surprisingly powerful. A hydrogen atom's mass is not equal to the sum of the mass of a proton and electron, due to the binding energy. Thus the WEP implies that gravity must couple to the EM field so that $m_i = m_g$ continues to hold.
- The Einstein Equivalence Principle (EEP) generalizes this statement to all local experiments. It's a bit tricky to think of a theory that violates the EEP but satisfies the WEP. One example is a situation where gravity caused non-point particles to start rotating as they fall.
- The EEP tells us that there are no "gravitationally neutral" objects with respect to which we can measure g. Thus we instead define unaccelerated/inertial frames to be freely falling, which motivates us to describe gravity as not a force, but a property of spacetime.
- Since gravitational fields are not homogeneous, global inertial frames don't exist. In particular, we have to abandon the SR picture of "networks of clocks and rulers", and coordinates become much harder to interpret physically.
- Sometimes, we make a distinction between gravitational and nongravitational interactions, and take the WEP to only include the latter. The Strong Equivalence Principle (SEP) then generalizes the WEP to include gravitational self-interactions. A direct test of the SEP (but not the WEP) would involve measuring the accelerations of objects with significant gravitational binding energy. For example, the Nordtvedt effect is the fact that if the SEP were violated, the relative motion between the Earth and Moon would be affected, because the two have different ratios of gravitational binding energy to inertial mass. This has been tested using lunar laser ranging experiments.

- From a particle physics perspective, tests of the equivalence principle are most useful for constraining additional, non-gravitational long-range forces, which would generically violate the WEP. Specific tests of the SEP are also useful. For instance, if one unluckily had a long-range force that coupled equally to protons and neutrons, there would be little apparent WEP violation, since protons and neutrons have almost equal mass, but there would be significant SEP violation. It is also useful to do tests at a variety of scales (i.e. tabletop to astronomical), because the force could have a finite range μ, which would be more apparent at distances $1/\mu$.
- Deviations from the equivalence principles, particularly the SEP, could of course also be explained by modifying the structure of gravity itself. However, at some level "adding new forces" and "modifying gravity" aren't all that different, because in most cases they both boil down to adding new fields and couplings to the Lagrangian.
- To date, all experimental tests are consistent with all equivalence principles, though specific tests of the SEP are less stringent.

Example. The EEP predicts gravitational redshift. Consider two rockets in space a distance z apart, with acceleration a and velocities $v \ll c$. The trailing rocket emits a photon of wavelength λ_0, which reaches the other rocket after time $\Delta t \approx z/c$. The receiving rocket has picked up an additional velocity $\Delta v = a\Delta t = az/c$, so that the photon is redshifted by

$$\frac{\Delta\lambda}{\lambda_0} = \frac{\Delta v}{c} = \frac{az}{c^2}.$$

Above we worked to first order in the velocities to avoid extra special relativistic effects.

By the EEP, the same thing should happen if a photon is instead emitted upward a distance z in a uniform gravitational field, giving

$$\frac{\Delta\lambda}{\lambda_0} = \frac{gz}{c^2}.$$

In general relativity, the metric in a weak time-independent gravitational potential ϕ is

$$c^2 d\tau^2 = (1 + 2\phi/c^2)c^2\, dt^2 - (1 - 2\phi/c^2)\, d\mathbf{r}^2, \quad \phi/c^2 \ll 1$$

Now consider two pulses of light sent between points A and B separated by a time Δt. Since the gravitational field is time-independent, the paths taken by the pulses are identical, so they also arrive separated in coordinate time Δt. Converting to proper time,

$$\Delta\tau_A^2 = (1 + 2\phi_A/c^2)\, \delta t^2, \quad \Delta\tau_B^2 = (1 + 2\phi_B/c^2)\, \delta t^2$$

so that a redshift of

$$\frac{\Delta\lambda}{\lambda_0} = \frac{\Delta\Phi}{c^2}$$

is observed, where we are working to lowest order in ϕ/c^2. This is in agreement with the EEP result. This trick of using the separation time for pulses also lets us avoid the geodesic equation.

1.3 Physical Differences

We quickly review what changes when moving from special relativity (SR) to general relativity (GR). Note that when we refer to SR, we are referring to inertial frames in Minkowski space with Cartesian coordinates.

- In SR, we considered vectors as 'free', with a base point that could be moved. In GR, this must be performed by parallel transport, and the result generally is not unique.
- In SR, we considered spacetime events as vectors x^μ. This is only possible because we identified spacetime itself with the tangent space at the origin, which we can't do in GR.
- In SR, inertial frames were defined over all spacetime. In GR, they can't due to tidal effects. In fact, we usually can't define a global system of coordinates at all, as spacetime is a general manifold which may require multiple charts.
- The time measured by a moving clock is still $\tau = \int \sqrt{-ds^2}$.
- Suppose a particle located at the origin of some coordinate system has momentum p^μ. If this coordinate system locally corresponds to the frame of an observer at the origin, they still measure the energy of the particle to be p^0, and so on.
- In SR, the Levi–Civita symbol
$$\widetilde{\epsilon}_{\mu\nu\rho\sigma} = \text{sign}(\mu\nu\rho\sigma)$$
is a pseudotensor: it transforms properly under connected Lorentz transformations and picks up an extra sign from T and P. In GR, the Levi–Civita symbol is not a tensor at all.
- In SR, the partial derivative takes (r, s) tensors to $(r, s+1)$ tensors. In GR, we instead must use the covariant derivative. (This also holds when we broaden the frames allowed in SR, as we get a nontrivial connection.)

Example. If two expressions agree in some frame, then they must agree in all frames. This allows us to find general results with very little work. For example, the energy of a particle in some frame is p^0, and the velocity of an observer in its own frame is $u^\mu = (1, 0, 0, 0)$. Therefore $E = p^\mu u_\mu$ in this frame, and hence it is true in all frames.

As another example, consider the Lie derivative $\mathcal{L}_V W$ where $V = \partial_0$. Then
$$\mathcal{L}_V W = (\partial_0 W^i)\partial_i.$$
The right-hand side happens to be equal to the commutator $[V, W]$ in these coordinates, so in general the Lie derivative is the commutator.

Example. Newtonian gravity in index notation. The equation of motion for a particle is $\ddot{x}_i = g_i$. However, in the spirit of the EEP, we note that this acceleration can always be set to zero in a falling frame, so instead we focus on tidal effects. Consider two particles separated by δx. Then
$$\delta \ddot{x}_i = \delta x_j \partial_j g_i + O(\delta x^2).$$
To simplify this we define the tidal tensor
$$E_{ij} = -\partial_j g_i, \quad \delta \ddot{x}_i + E_{ij}\delta x_j = 0.$$
Since the gravitational field in Newtonian theory is curl-free, there exists a potential ϕ so that
$$g_i = -\partial_i \phi$$

which implies $E_{ij} = E_{ji}$. In addition, since mixed partials commute, we have

$$E_{i[j,k]} = 0.$$

Finally, the field is sourced by matter by Poisson's equation,

$$\partial_i \partial_i \phi = E_{ii} = 4\pi G \rho.$$

Similar equations appear in general relativity, where the tidal tensor corresponds to the Riemann tensor, and our identities for it correspond to the symmetries of the Riemann tensor and the Bianchi identity. The potential roughly corresponds to the metric, and Poisson's equation corresponds to the Einstein field equation.

1.4 Manifolds

We review the basics of differential geometry, considering structures that can be defined without using a metric. We begin with the tangent space.

- Consider an n-dimensional manifold M. A scalar field $f\colon M \to \mathbb{R}$ is smooth if $F = f \circ \phi^{-1}$ is smooth for all charts ϕ. For example, coordinate functions themselves are smooth because transition functions are smooth by the definition of a smooth manifold.

- A smooth curve is a smooth function $\lambda\colon I \to M$ where I is an open interval in $\mathbb{R}$.

- If $f\colon M \to \mathbb{R}$ is smooth, then $f \circ \lambda\colon I \to \mathbb{R}$ is smooth, and in particular it has a derivative. We thus define the tangent vector to λ at p as the map

$$X_p(f) = \left(\frac{d}{dt} f(\lambda(t))\right)_{t=0}$$

  where $\lambda(0) = p$. Then X_p is a linear map on smooth functions and a derivation,

$$X_p(fg) = X_p(f)g(p) + f(p)X_p(g).$$

  The set of tangent vectors of p forms a vector space in the usual way.

- We may also write the tangent vector in components, by noting

$$f \circ \lambda = (f \circ \phi^{-1}) \circ (\phi \circ \lambda)$$

  which gives

$$X_p(f) = \left(\frac{\partial F(x)}{\partial x^\mu}\right)_{x=\phi(p)} \left(\frac{dx^\mu(\lambda(t))}{dt}\right)_{t=0}$$

  where we used the chain rule.

- Now we check that the tangent space T_p is an n-dimensional vector space.

  - First, we check that it is closed under linear combinations. Note that if $\lambda(t)$ and $\kappa(t)$ give vectors X_p and Y_p, then

$$\nu(t) = \phi^{-1}\left(\alpha(\phi(\lambda(t)) - \phi(p)) + \beta(\phi(k(t)) - \phi(p)) + \phi(p)\right)$$

    has tangent vector $\alpha X_p + \beta Y_p$, as desired.

  - Next, we check that T_p has dimension n. The expression above shows that any $X_p(f)$ can be written as a linear combination of $\partial F(x)/\partial x^\mu$, so the vectors $\partial/\partial x^\mu$ are a complete set; they correspond to paths that only change the coordinate x^μ.
  - The vectors $\partial/\partial x^\mu$ are independent because if $\alpha^\mu \partial_\mu F = 0$ for all F, then choosing $F = x^\nu$ gives $\alpha^\nu = 0$. Thus they are a basis, giving the result.

  The basis $\partial/\partial x^\mu$ depends on the coordinate chart. It is defined inside the entire patch and forms a coordinate basis in the patch. A general vector is written $X = X^\mu e_\mu$ for a basis e_μ.

- To see how coordinate bases change under coordinate change, let ϕ and ϕ' give coordinates x and x'. Then formally,

$$(\partial_\mu)(f) = \frac{\partial}{\partial x^\mu}(f \circ \phi^{-1}) = \frac{\partial}{\partial x^\mu}((f \circ \phi'^{-1}) \circ (\phi' \circ \phi^{-1}))$$

  where ∂_μ is an abstract vector. We can now use the chain rule, giving

$$\frac{\partial F}{\partial x^\mu} = \frac{\partial F}{\partial x'^\nu}\frac{\partial x'^\nu}{\partial x^\mu}, \quad \partial_\mu = \frac{\partial x'^\nu}{\partial x^\mu}\partial'_\nu.$$

  More casually, we can heuristically derive this result by writing

$$\frac{\partial f(x'(x))}{\partial x} = \frac{\partial f}{\partial x'}\frac{\partial x'}{\partial x}$$

  where we implicitly identified a few quantities. By a very similar argument,

$$X'^\mu = X^\nu \frac{\partial x'^\mu}{\partial x^\nu}.$$

  This leaves vectors $X^\mu \partial_\mu$ invariant since the transformation factors cancel.

Next, we define covectors.

- The dual space V^* of a vector space V is the set of linear maps $V \to \mathbb{R}$. Given a basis e_μ of V there is a dual basis f^μ of V^* defined by $f^\mu(e_\nu) = \delta^\mu_\nu$.

- There is no natural isomorphism between V and V^*, though there is one between V and V^{**} by 'shuffling parentheses'.

- Define the cotangent space $T_p^*(M)$ as the dual space of the tangent space. Given a smooth function f, we may define a covector df, called the gradient of f, by

$$(df)_p(X) = X(f)_p.$$

  In particular, dx^μ is the dual basis to ∂_μ.

- Writing a covector as $\omega = \omega_\mu dx^\mu$, we have the transformation laws

$$dx^\mu = \frac{\partial x^\mu}{\partial x'^\nu}\, dx'^\nu, \quad \omega'_\mu = \frac{\partial x^\nu}{\partial x'^\mu}\, \omega_\nu.$$

  Again, these follow by 'lining up the derivatives'.

Note. We use Greek indices in equations that are only true in a particular coordinate system, and Latin indices in equations that are always true. For example, for a vector X, we can write $X^\mu = \delta^\mu_0$ in some coordinate system, but generally $df(X) = df_a X^a$. Equations in Latin should be interpreted as 'component-free', with the indices only indicating where the parentheses go.

Finally, we introduce tensors.

- A tensor of type (r, s) at p is a multilinear map which takes r covectors and s vectors to $\mathbb{R}$. For example, covectors are tensors of type $(0, 1)$ and vectors are tensors of type $(1, 0)$. Also, defining $\delta(\omega, X) = \omega(X)$, δ is a $(1, 1)$ tensor.
- Choosing a basis of vectors $\{e_\mu\}$ with dual basis $\{f^\mu\}$, the components of a tensor are

  $$T^{\mu_1 \ldots \mu_r}{}_{\nu_1 \ldots \nu_s} = T(f^{\mu_1}, \ldots, e_{\nu_1}, \ldots).$$

  For example, the components of δ are

  $$\delta^\mu{}_\nu = \delta(f^\mu, e_\nu) = f^\mu(e_\nu) = \delta^\mu_\nu.$$

  The set of tensors at p is a vector space with dimension n^{r+s}.
- Now we consider how tensor components change under a general change of coordinates,

  $$f^{\mu'} = A^{\mu'}{}_\nu f^\nu.$$

  The same arguments as before tell us that

  $$e_{\mu'} = (A^{-1})^\nu{}_{\mu'} e_\nu, \quad X^{\mu'} = A^{\mu'}{}_\nu X^\nu, \quad \eta_{\mu'} = (A^{-1})^\nu{}_{\mu'} \eta_\nu.$$

  Plugging these results in, a tensor transforms as, e.g.

  $$T^{\mu'\nu'}{}_{\rho'} = A^{\mu'}{}_\sigma A^{\nu'}{}_\tau (A^{-1})^\lambda{}_{\rho'} T^{\sigma\tau}{}_\lambda.$$

  In the special case of a coordinate transformation, $A^{\mu'}{}_\nu = \partial x^{\mu'} / \partial x^\nu$. In the even more special case of a Lorentz transformation, A is Λ.
- Given an (r, s) tensor, we can construct an $(r-1, s-1)$ tensor by contracting two indices. This is done by plugging in a basis and dual basis,

  $$T(f^\mu, e_\mu, \ldots) = S(\ldots).$$

  This is basis independent because the left-hand side transforms as $T \to AA^{-1}T = T$.
- We can also construct tensors by the tensor product. For example,

  $$(S \otimes T)(\omega, X) = S(\omega)T(X)$$

  with the same pattern holding for arbitrary tensors. The components simply multiply.

- Finally, we may symmetrize and antisymmetrize tensors. For example, given a $(0,2)$ tensor T, its symmetric and antisymmetric parts are

$$S(X,Y) = \frac{T(X,Y)+T(Y,X)}{2}, \quad A(X,Y) = \frac{T(X,Y)-T(Y,X)}{2}$$

which, in index notation, reads

$$S_{\mu\nu} = \frac{T_{\mu\nu}+T_{\nu\mu}}{2}, \quad A_{\mu\nu} = \frac{T_{\mu\nu}-T_{\nu\mu}}{2}.$$

We will also use vertical bars to denote exclusion from (anti)symmetrization. For example,

$$T_{(\mu|\nu\rho|\sigma)} = \frac{T_{\mu\nu\rho\sigma}+T_{\sigma\nu\rho\mu}}{2}.$$

The most useful property is that contractions of symmetric and antisymmetric tensors vanish.

- Similarly we may define vector and tensor fields on M. A vector field X is smooth if $X(f)$ is a smooth function for all smooth f, with other definitions similar.

Next, we review some geometric objects derived from vector fields.

- Given a vector field X, an integral curve of X through p is a curve through p whose tangent at every point is X. Taking coordinates, this means that

$$\frac{dx^\mu(t)}{dt} = X^\mu(x(t)), \quad x^\mu(0) = x^\mu|_p.$$

Note that $X^\mu(x(t))$ means X^μ evaluated at the point $x(t)$, not acting on anything.

- If f is a function satisfying $X(f) = 0$, then f is conserved on integral curves.

- Flow along integral curves generates a one-parameter group of diffeomorphisms $\phi_t \colon M \to M$ by flowing along the integral curves for time t. Conversely, ϕ_t gives a vector field by differentiation at $t = 0$.

- The commutator of two vector fields is

$$[X,Y](f) = X(Y(f)) - Y(X(f)).$$

It turns out to be a vector field, since the second derivatives cancel, with components

$$\boxed{[X,Y]^\mu = X^\nu \partial_\nu Y^\mu - Y^\nu \partial_\nu X^\mu.}$$

The commutator operation turns the set of vector fields on M into a Lie algebra, whose corresponding Lie group is the set of diffeomorphisms of M.

- Note that we can define the components of a tensor with respect to any set of vector fields $\{e_\mu\}$ that form a basis at every point. By Frobenius' theorem, we have $[e_\mu, e_\nu] = 0$ if and only if the e_μ are a coordinate basis, i.e. $e_\mu = \partial_\mu$. Most of the time we'll work in a coordinate basis, but we'll try to point out what extra terms appear outside such a basis.

Note. More examples of Lie algebras.

- An explicit basis for the Lie algebra $\mathfrak{diff}(\mathbb{R})$ is

$$X_\alpha = -x^{\alpha+1}\partial_x, \quad \alpha \in \mathbb{Z}, \quad [X_\alpha, X_\beta] = (\alpha - \beta)X_{\alpha+\beta}.$$

- There exist only two two-dimensional Lie algebras,

$$[X, Y] = 0 \text{ and } [X, Y] = Y.$$

  The latter is the Lie algebra of affine transformations of the line.

- The Euclidean group $E(2)$ acts on $M = \mathbb{R}^2$ by rotations and translations,

$$\mathbf{x} \to R(\theta)\mathbf{x} + \begin{pmatrix} a \\ b \end{pmatrix}.$$

  Then $E(2)$ is a three-dimensional Lie group, parametrized by θ, a, and b. We can assign a vector field to every infinitesimal transformation (alternatively, every one-parameter subgroup gives a one-parameter family of diffeomorphisms), giving

$$X_a = \partial_x, \quad X_b = \partial_y, \quad X_\theta = x\partial_y - y\partial_x.$$

  which form a basis for $\mathfrak{e}(2)$ with

$$[X_a, X_b] = 0, \quad [X_a, X_\theta] = X_b, \quad [X_b, X_\theta] = -X_a.$$

  More generally, the set of Killing vectors will form a Lie algebra.

Note. With index notation, we simultaneously speak about tensors and their components; however, this leads to ambiguity, especially when working with covariant derivatives, and is a bit inelegant to mathematicians because we always need to specify a coordinate system. If necessary, we will use abstract indices, which only mean the former. For instance, $X^a f_a$ is simply a shorthand for $X(f)$ and does not indicate a coordinate system; the "abstract index" a does not take numeric values.

2 Riemannian Geometry

2.1 The Metric

- The metric tensor $g_{\mu\nu}$ is a nondegenerate symmetric $(0,2)$ tensor. Since the metric is nondegenerate, $g = |g_{\mu\nu}| \neq 0$. Then there exists an inverse metric, a $(2,0)$ tensor satisfying
$$g^{\mu\nu} g_{\nu\sigma} = g_{\sigma\lambda} g^{\lambda\mu} = \delta^\mu_\sigma.$$
For example, the trace of g is $g^{\mu\nu} g_{\mu\nu} = \delta^\mu_\mu = 4$, in any signature.

- Unlike in special relativity, index placement now matters (it cannot be restored at the end of the calculation) because the metric has a nontrivial derivative. For example, since
$$\partial_\lambda (g^{\mu\nu} g_{\nu\sigma}) = (\partial_\lambda g^{\mu\nu}) g_{\nu\sigma} + g^{\mu\nu} (\partial_\lambda g_{\nu\sigma}) = 0$$
we conclude that
$$\partial_\lambda g^{\mu\nu} = -g^{\mu\sigma} g^{\nu\rho} \partial_\lambda g_{\sigma\rho}.$$
The minus sign is the same one as in $(1/f)' = -f'/f^2$.

- The metric is extremely useful: we use it to raise and lower indices, and compute path lengths and proper times, giving geodesics. It determines causality and locally inertial frames. It is the generalization of both the Newtonian dot product and the Newtonian gravitational potential.

- The metric is in canonical form if it is diagonal, with p and q elements equal to 1 and -1 respectively. Sylvester's theorem states that this can always be done at any given point, with p and q unique. By continuity, the signature (p, q) is the same throughout the manifold.

- If $q = 0$, the metric is called Euclidean/Riemannian, and if $q = 1$ (as in relativity) the metric is called Lorentzian/pseudo-Riemannian; the canonical form is the Minkowski metric.

- The metric takes two vectors and gives a number, so it may be written as
$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu.$$
Here, ds^2 is the metric tensor in component-free form, and $dx^\mu dx^\nu$ is a tensor product. Since the metric is symmetric, we use symmetrized tensor products so that $dxdy = dydx$.

- The length of a spacelike curve is
$$s = \int \sqrt{g(V,V)}\, dt$$
where $V^\mu(t) = dx^\mu(t)/dt$. Similarly, the proper time along a timelike curve is
$$\tau = \int \sqrt{-g(V,V)}\, dt.$$
Note that t is not time, but just an arbitrary parameter. If we parametrize by proper time, then $V^\mu = dx^\mu/d\tau$ is the four-velocity, giving $g(V,V) = -1$ just as in special relativity.

Note. Parameter counting for coordinate transformations. Let $x^\mu(p) = x^{\hat{\mu}}(p) = 0$, and expand

$$g_{\hat{\mu}\hat{\nu}} = \frac{\partial x^\mu}{\partial x^{\hat{\mu}}} \frac{\partial x^\nu}{\partial x^{\hat{\nu}}} g_{\mu\nu}$$

in a Taylor series in $x^{\hat{\mu}}$ about p. Expanding both sides to second order in the $x^{\hat{\mu}}$, dropping constants and indices and evaluating all derivatives at p, we have

$$\begin{aligned}(\hat{g}) + (\hat{\partial}\hat{g})\hat{x} + (\hat{\partial}\hat{\partial}\hat{g})\hat{x}\hat{x} = \left(\frac{\partial x}{\partial \hat{x}}\frac{\partial x}{\partial \hat{x}} g\right) + \left(\frac{\partial x}{\partial \hat{x}}\frac{\partial^2 x}{\partial \hat{x}\partial \hat{x}} g + \frac{\partial x}{\partial \hat{x}}\frac{\partial x}{\partial \hat{x}}\hat{\partial} g\right)\hat{x} \\ + \left(\frac{\partial x}{\partial \hat{x}}\frac{\partial^3 x}{\partial \hat{x}\partial \hat{x}\partial \hat{x}} g + \frac{\partial^2 x}{\partial \hat{x}\partial \hat{x}}\frac{\partial^2 x}{\partial \hat{x}\partial \hat{x}} g + \frac{\partial x}{\partial \hat{x}}\frac{\partial^2 x}{\partial \hat{x}\partial \hat{x}}\hat{\partial} g + \frac{\partial x}{\partial \hat{x}}\frac{\partial x}{\partial \hat{x}}\hat{\partial}\hat{\partial} g\right)\hat{x}\hat{x}\end{aligned}$$

Now let's consider this expression order-by-order in $\hat{x}$.

- At zeroth order, we get the transformation law at p. There are 16 parameters in the matrix $\partial x^\mu / \partial x^{\hat{\mu}}$, but the metric only has 10, since it's symmetric. Therefore we can always bring the metric into canonical (i.e. Minkowski) form at a point, and the extra 6 degrees of freedom are the Lorentz transformations.
- At first order, we have 40 numbers on the left-hand side, from 4 derivatives of 10 metric components. On the right-hand side, the $(\partial x / \partial \hat{x})^2$ term gives nothing since it was used up at zeroth order, but $\partial^2 x^\mu / (\partial x^{\hat{\mu}_1} \partial x^{\hat{\mu}_2})$ has 40 parameters, since the second derivative is symmetric. Then we have just enough freedom to set $\hat{\partial}\hat{g}$ to zero.
- At second order, we have 100 numbers on the left-hand side, since $\hat{\partial}\hat{\partial}$ is symmetric. On the right-hand side, we can only set $\partial^3 x / \partial \hat{x}\partial \hat{x}\partial \hat{x}$, which has 4 choices in the numerator and 20 in the denominator, so we're short by 20 degrees of freedom. These tell us about the curvature of the manifold; we will see later the Riemann tensor has 20 independent components.

Note. As motivated above, at any point p, there should exist a coordinate system x^μ with

$$g_{\mu\nu}(p) = \eta_{\mu\nu}, \quad \partial_\sigma g_{\mu\nu}(p) = 0.$$

Such coordinates are called locally inertial coordinates, or Riemann normal coordinates, and the associated basis vectors constitute a local Lorentz frame. These frames are associated with freely falling observers, as they see no effects of gravity besides tidal effects, which only appear to second order. Later, we will construct such coordinates using geodesics.

Locally inertial coordinates are useful for extracting general expressions. While a calculation in curved spacetime may be difficult, we can always go into locally inertial coordinates at a point and simplify using $g \sim \eta$ and $\partial g = 0$. As long as we phrase our final answers in terms of tensorial quantities, they must hold in all coordinate systems.

Note. We can always choose a basis at every point so that the metric is in canonical form at every point. However, such a set of bases generally does not mesh together to form a coordinate system.

Example. In a more elementary treatment, we think of dx^μ as an infinitesimal displacement and ds as an infinitesimal length. For example, consider the metric

$$ds^2 = -dt^2 + t^{2q} dx^2.$$

We want to find the null paths $x^\mu(\lambda)$ followed by light. The tangent vector is

$$V = \frac{dx^\mu}{d\lambda}\partial_\mu.$$

Then the null paths must satisfy

$$ds^2(V,V) = -dt^2(V,V) + t^{2q}dx^2(V,V) = 0.$$

Now, working very explicitly, we have

$$dt^2(V,V) = [dt(V)]^2 = \left(\frac{dt}{d\lambda}\right)^2, \quad 0 = -\left(\frac{dt}{d\lambda}\right)^2 + t^{2q}\left(\frac{dx}{d\lambda}\right)^2.$$

We now have an ordinary differential equation, which simplifies to

$$\frac{dx}{dt} = \pm t^{-q}, \quad t = (1-q)^{1/(1-q)}(\pm x - x_0)^{1/(1-q)}.$$

In the more elementary view, we could have "set $ds = 0$" and then "divided by dt^2" and "taken the square root". This more casual procedure always gives the same result.

2.2 Geodesics

In this section, we derive the geodesic equation.

- In general relativity, we postulate that free massive particles follow paths of maximum proper time, called geodesics. This corresponds to the Lagrangian

$$L = \sqrt{-g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}, \quad S = \int L\, d\lambda$$

  for paths $x^\mu(\lambda)$, where the dot indicates differentiation with respect to λ.

- Physically, this postulate makes sense because such paths are locally straight through spacetime, just like segments of minimum length are straight lines in Euclidean space.

- By the chain rule, the action above is invariant under reparametrization. Given a path, a useful choice is the proper time τ along the curve, since $L = 1$. Now, given a variation $\delta x^\mu(\tau)$ about this path, the Lagrangian varies by

$$\delta\left(\sqrt{1+\epsilon}\right) = \epsilon/2.$$

  If we instead used the Lagrangian $-L^2$, the variation would be $-\epsilon$. Since these are proportional, the two actions have the same stationary points.

- The new action $-\int L^2\, d\tau$ is not reparametrization invariant. If we switch to a new parameter λ, the integrand is multiplied by $d\lambda/d\tau$. To maintain the same stationary points, this must be a constant, so our results will only be valid for parameters affinely related to τ, i.e. $\lambda = a\tau + b$.

- We can find geodesics with the Euler-Lagrange equations, but in this case it's easier to directly plug in the variation,

$$x^\mu \to x^\mu + \delta x^\mu, \quad g_{\mu\nu} \to g_{\mu\nu} + (\partial_\sigma g_{\mu\nu})\delta x^\sigma.$$

  We then work to first order and integrate by parts as usual, taking care to include derivatives of the metric (i.e. $dg_{\mu\nu}/d\tau = (\partial_\sigma g_{\mu\nu})dx^\sigma/d\tau$), to bring out a factor of δx^μ.

- Finally, solving for the acceleration gives the geodesic equation,

$$\boxed{\ddot{x}^\mu + \Gamma^\mu_{\nu\rho}\dot{x}^\nu\dot{x}^\rho = 0}$$

  where the Christoffel symbols are

$$\boxed{\Gamma^\mu_{\nu\rho} = \frac{1}{2}g^{\mu\sigma}(\partial_\nu g_{\rho\sigma} + \partial_\rho g_{\sigma\nu} - \partial_\sigma g_{\nu\rho}).}$$

  This only holds for affine parameters; otherwise an extra term appears. We have symmetrized the Christoffel symbols in the lower two indices, as $\dot{x}^\nu\dot{x}^\rho$ is symmetric.

- We will also use the comma notation for partial derivatives,

$$\Gamma^\mu_{\nu\rho} = \frac{1}{2}g^{\mu\sigma}(g_{\rho\sigma,\nu} + g_{\sigma\nu,\rho} - g_{\nu\rho,\sigma}).$$

  If there are multiple indices after the comma, they indicate higher derivatives.

We now make some remarks about this result.

- Sometimes, it's easiest to compute the Christoffel symbols by using the geodesic equation in reverse, explicitly varying the proper time integral to compute $\ddot{x}^\mu$. Another shortcut method when many metric components are zero is to just read them off from the four Euler-Lagrange equations, which will each be simple.

- We never used the fact that the metric was Lorentzian, so the geodesic equation also can be used to find, e.g., shortest paths in space. It can also be used to find shortest spacelike paths in relativity; in this case we have $L = \sqrt{g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu}$ and parametrize by the proper length s so that $L = 1$, and the rest of the derivation goes through as before.

- To see that the path is a maximum of proper time and not a minimum, note that we can always approximate a timelike path with many lightlike paths, which have zero proper time.

- If we add forces, they will appear on the right-hand side of the geodesic equation; for example, the electromagnetic force would appear as $qF^{\mu\nu}u_\nu$.

- As we'll see later, geodesics are paths that parallel transport their own tangent vector $dx^\mu/d\lambda$, which implies that its norm remains fixed. Thus a geodesic timelike at one point is timelike everywhere, and so on. This can also be shown directly from the squared Lagrangian; the norm of the four-velocity is the conserved quantity associated with τ-translation invariance.

- Null geodesics are paths obeying the geodesic equation that are everywhere null. Our derivation above doesn't work for massless particles, but we can show using the einbein action (as introduced in the notes on String Theory) that massless particles follow null geodesics.

- As for timelike geodesics, null geodesics have affine freedom in their parametrizations, but there is no canonical choice like the proper time. One typical choice is

$$p^\mu = \frac{dx^\mu}{d\lambda}$$

  so that the velocity is the momentum. For a timelike geodesic, the velocity is the momentum per unit mass, so the null geodesic parameter λ is essentially τ/m in the limit $m \to 0$.

- Finally, there are spacelike geodesics, which we parametrize by proper length. These would appear, for example, as the paths of taut strings.

2.3 Covariant Derivatives

The partial derivative ∂_μ does not map tensors to tensors, so equations involving it are typically not valid in general in curved spacetime. We fix this problem by replacing the partial derivative with the covariant derivative.

- We define the covariant derivative with the following postulates.
  - ∇ is a map from (k, l) tensor fields to $(k, l+1)$ tensor fields.
  - Linearity: $\nabla(T+S) = \nabla T + \nabla S$.
  - Leibniz (product) rule: $\nabla(T \otimes S) = (\nabla T) \otimes S + T \otimes (\nabla S)$.
  - Reduces to partial derivative for scalars: $(\nabla f)_\mu = \partial_\mu f$.
  - Compatible with contractions: $\nabla(\delta^\mu_\nu) = 0$.

  Such a structure is independent of a metric.
- The third postulate is reasonable since ∇ should act like a derivative. The fourth follows because there's no issue with differentiation of scalars, as their values don't depend on local basis choices. The fifth is sensible because the identity map shouldn't change. Combining it with the Leibniz rule gives, for example, $\nabla_\nu(A^\mu B_\mu) = \nabla_\nu(A^\mu)B_\mu + A^\mu \nabla_\nu(B_\mu)$.
- We can also define directional covariant derivatives, i.e. $W^\mu \nabla_\mu V^\nu$ is the rate of change of V along W. We write $W^\mu \nabla_\mu = \nabla_W$, so that ∇_W is a map from (k, l) tensor fields to (k, l) tensor fields. This is a more general starting point, as we may replace the fourth postulate with $\nabla_W f = W(f)$, which works without a coordinate basis.
- There are several notations for covariant derivatives,

  $$(\nabla_\nu V)^\mu = \nabla_\nu V^\mu = V^\mu{}_{;\nu} = (\nabla V)^\mu{}_\nu.$$

  The second notation is ambiguous because it can also be thought of as the covariant derivative of the component function V^μ, which would simply be $\partial_\nu V^\mu$. The correct meaning must be inferred from context. Generally, V^μ stands for the component function alone when there is a corresponding vector ∂_μ somewhere else in the equation, as in $(\nabla V^\mu)\partial_\mu$. Alternatively, one can use abstract index notation, where $\nabla_a V^b$ is unambiguously equal to $V^a{}_{;b}$.
- When there are multiple indices after the semicolon, the derivative immediately after the semicolon is taken first, e.g. $V^\mu{}_{;\rho\sigma} = \nabla_\sigma \nabla_\rho V^\mu$.

We now construct the covariant derivative of a tensor field explicitly.

- Given a basis $\{e_\mu\}$ at every point, not necessarily a coordinate basis, we define

  $$\nabla_\nu e_\mu = \Gamma^\rho_{\mu\nu} e_\rho$$

  where the Γ's are the connection coefficients; these are simply a set of numbers that depends on the basis.

- Therefore the covariant derivative of a general vector is

$$\nabla_\nu W = \nabla_\nu (W^\mu e_\mu) = (\nabla_\nu W^\mu) e_\mu + W^\mu (\nabla_\nu e_\mu) = (\partial_\nu W^\rho + \Gamma^\rho_{\mu\nu} W^\mu) e_\rho.$$

Thus, relapsing into our sloppy notation,

$$\boxed{\nabla_\nu W^\rho = \partial_\nu W^\rho + \Gamma^\rho_{\mu\nu} W^\mu.}$$

- We can compute the transformation of the connection coefficients using our first definition. Denoting the new basis with a prime and the Jacobian as J,

$$\Gamma^{\rho'}_{\mu'\nu'} e_{\rho'} = \nabla_{\nu'} e_{\mu'} = J^\nu_{\nu'} \nabla_\nu (J^\mu_{\mu'} e_\mu) = J^\nu_{\nu'} \nabla_\nu (J^\mu_{\mu'}) e_\mu + J^\nu_{\nu'} J^\mu_{\mu'} \nabla_\nu e_\mu.$$

Recognizing the unprimed connection coefficients on the right, and rearranging gives

$$\Gamma^{\rho'}_{\mu'\nu'} = J^{\rho'}_\rho J^\nu_{\nu'} J^\mu_{\mu'} \Gamma^\rho_{\mu\nu} + J^{\rho'}_\rho \partial_{\nu'} J^\rho_{\mu'}.$$

We get the expected tensorial term, plus an extra term that is independent of Γ.

- This shows that the connection coefficients are not a tensor, but the difference of connection coefficients is, so we define the torsion tensor

$$T_{\mu\nu}{}^\rho = \Gamma^\rho_{\mu\nu} - \Gamma^\rho_{\nu\mu}.$$

We say a connection is torsion free if the torsion vanishes.

- Another way to see this result is to consider the map $\nabla \colon X, Y \mapsto \nabla_X Y$. This is not a $(1,2)$ tensor because it is not linear in Y,

$$\nabla_X (fY) = f \nabla_X Y + Y \nabla_X f.$$

However, the difference of two covariant derivatives $\nabla - \nabla'$ is a $(1,2)$ tensor because the extra term $Y \nabla_X f = Y \nabla'_X f = Y X(f)$ cancels out.

- Above, the Jacobian is defined as

$$J^{\mu'}_\nu = \frac{\partial x^{\mu'}}{\partial x^\nu}, \quad J^\mu_{\nu'} = \frac{\partial x^\mu}{\partial x^{\nu'}}.$$

Then the chain rule says that

$$J^{\mu'}_\nu J^\nu_{\rho'} = \delta^{\mu'}_{\rho'}$$

and differentiating both sides gives

$$(\partial_{\sigma'} J^{\mu'}_\nu) J^\nu_{\rho'} + J^{\mu'}_\nu (\partial_{\sigma'} J^\nu_{\rho'}) = 0.$$

- The extra transformation term for Γ makes $\nabla_\nu W^\rho$ into a tensor. Note that

$$\partial_{\nu'} W^{\rho'} = J^\nu_{\nu'} J^{\rho'}_\rho \partial_\nu W^\rho + W^\rho \partial_{\nu'} (J^{\rho'}_\rho).$$

Applying our Jacobian identity above shows that the extra nontensorial pieces cancel.

- Applying the Leibniz rule gives the covariant derivative for covectors,

$$\boxed{\nabla_\rho \eta_\mu = \partial_\rho \eta_\mu - \Gamma^\nu_{\mu\rho}\eta_\nu.}$$

The covariant derivative for a rank (r, s) tensor T is similar. Besides the partial derivative term, we get $r + s$ terms, each of which have a factor of Γ contracted to one index, where downstairs indices T get terms with minus signs. In general, one of the left two indices in each Γ is contracted with T, while the bottom right index is the index on ∇.

Next, we consider some properties of torsion-free connections.

- A general, coordinate-free definition of a torsion-free connection is one which satisfies

$$\nabla_a \nabla_b f = \nabla_b \nabla_a f$$

in any basis. In general we would have

$$\nabla_a \nabla_b f - \nabla_b \nabla_a f = -T^c{}_{ab} \nabla_c f.$$

- It's straightforward to show that in any basis,

$$f_{;\mu\nu} = f_{,\mu\nu} - \Gamma^\rho_{\mu\nu} f_{,\rho}.$$

Antisymmetrizing both sides and restricting to a coordinate basis, $f_{,[\mu\nu]} = 0$, which gives $\Gamma^\rho_{[\mu\nu]} = 0$, our earlier definition of a torsion-free connection.

- Geometrically, this tells us that in the absence of torsion, parallel transporting in a square of side a gives a closed curve, up to $O(a^3)$ terms due to the curvature.
- As another example, note that

$$\nabla_X Y - \nabla_Y X = X^\nu \nabla_\nu Y^\mu - Y^\nu \nabla_\nu X^\mu = X^\nu \partial_\nu Y^\mu - Y^\nu \partial_\nu X^\mu + 2\Gamma^\mu_{[\rho\nu]} X^\nu Y^\rho.$$

Therefore, for any torsion-free connection, and any vector fields X and Y,

$$\nabla_X Y - \nabla_Y X = [X, Y].$$

This is a nice result, since we now can compute $[X, Y]$ without intermediate nontensorial quantities. More generally, we can compute the Lie derivative and exterior derivative, for a torsion-free connection, by replacing all partial derivatives by covariant ones.

We now define the Levi–Civita connection used in general relativity.

- A connection is metric compatible if $\nabla g = 0$. We claim that there exists a unique torsion-free metric compatible connection, called the Levi–Civita connection, where the connection coefficients are the Christoffel symbols.
- The metric compatibility condition is

$$\partial_\alpha g_{\beta\gamma} = \Gamma^\rho_{\beta\alpha} g_{\rho\gamma} + \Gamma^\rho_{\gamma\alpha} g_{\beta\rho}.$$

To show that the Christoffel symbols obey it, we simply plug them in. To show they are the only things that obey it, note that we can add this equation to itself to get Christoffel symbols on the left-hand side, and simplifying the right-hand side gives the connection coefficients.

- Metric compatibility allows us to raise and lower indices through the covariant derivative,

$$g_{\mu\lambda}\nabla_\rho V^\lambda = \nabla_\rho V_\mu.$$

  We may also define an upper covariant derivative $\nabla^\mu = g^{\mu\nu}\nabla_\nu$. Without metric compatibility, this would be different from $\nabla_\nu g^{\mu\nu}$, but with metric compatibility they are perfectly equivalent.

Note. Philosophically, the torsion-free criterion is so natural that, when considering theories with torsion, we prefer to think of the connection as remaining torsion-free but with the torsion tensor acting as a new matter field.

Note. We repeat the derivation of the Levi–Civita connection without indices. Note that

$$X(g(Y,Z)) = \nabla_X(g(Y,Z)) = g(\nabla_X Y, Z) + g(Y, \nabla_X Z).$$

Adding copies of this equation to itself with the indices permuted gives

$$\begin{aligned}X(g(Y,Z)) + Y(g(Z,X)) - Z(g(X,Y))&\\ = g(\nabla_X Y + \nabla_Y X, Z) - g(\nabla_Z X - \nabla_X Z, Y) &+ g(\nabla_Y Z - \nabla_Z Y, X).\end{aligned}$$

Since the connection is torsion-free, we use our expression for the commutator to isolate the covariant derivative, resulting in the Koszul formula

$$2g(\nabla_X Y, Z) = X(g(Y,Z)) + Y(g(Z,X)) - Z(g(X,Y)) + g([X,Y],Z) + g([Z,X],Y) - g([Y,Z],X).$$

Since the metric is nondegenerate, this determines $\nabla_X Y$ and hence the connection. To recover our usual expression, we specialize to a coordinate basis; the commutator terms vanish and we read off the Christoffel symbols. For completeness, we should check that $\nabla_X Y$ defined above really does obey the properties of a connection, such as $\nabla_{fX} Y = f\nabla_X Y$. This check is long but straightforward.

2.4 Parallel Transport and Geodesics

Next, we define parallel transport and connect it to geodesics.

- Consider a parametrized curve with tangent vector $d/d\lambda$. Then a tensor T is parallel transported, or parallel propagated along the path if

$$\nabla_{d/d\lambda} T = \frac{dx^\mu}{d\lambda}\nabla_\mu T = 0.$$

  Intuitively, this means the tensor is kept constant along the curve. Concretely, expanding the definition gives an ordinary differential equation for the components of T.

- Technically, we have only defined the covariant derivative with respect to a vector field, not a parametrized curve. However, one can show that one can extend $d/d\lambda$ to a vector field in a small neighborhood, and the result is independent of the extension.

- Now recall the geodesic equation reads

$$\frac{d^2x^\alpha}{d\lambda^2} + \Gamma^\alpha_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = 0.$$

The velocity is $U^\alpha = dx^\alpha/d\lambda$, and writing in terms of it gives

$$0 = \frac{dU^\alpha}{d\lambda} + \Gamma^\alpha_{\mu\nu} U^\mu U^\nu = U^\nu(\partial_\nu U^\alpha + \Gamma^\alpha_{\mu\nu} U^\mu) = \nabla_U U^\alpha.$$

Therefore, a geodesic is a path that parallel transports its own tangent vector; these are the generalizations of straight lines. As we saw earlier, not every parametrization works; the vector $dx^\alpha/d\kappa$ is only parallel transported if κ and λ are affinely related; otherwise we get artificial "acceleration".

- Parallel transport preserves inner products, because $(D/D\lambda)(g_{\mu\nu} V^\mu W^\nu)$ splits into three terms by the Leibniz rule, all of which are zero. In particular, the norm of the velocity of a geodesic is preserved.

We give some simple examples of connections and parallel transport.

- In flat space and Cartesian coordinates, the Christoffel symbols are zero; parallel transport is done by keeping vector components constant.
- However, we must have nonzero connection coefficients in polar coordinates because "sliding a vector around" doesn't keep its polar components constant. This shows that the connection coefficients are not tensorial.
- In a curved space, we can always make the Christoffel symbols vanish at a point by using locally inertial coordinates, since they only depend on metric derivatives.
- In a manifold embedded in $\mathbb{R}^n$, parallel transport under the Levi–Civita connection simply means shifting a vector over in $\mathbb{R}^n$, then projecting it back down to the tangent plane.
- Given velocities at two distinct points in spacetime, the natural way to compute a relative velocity is to parallel transport one velocity to the other and subtract. This is possible exactly when the space is flat, in which case this reduces to the usual vector subtraction in Cartesian coordinates. When the space is curved, the parallel transport is path-dependent.

Note. Intuition for the torsion tensor T. There are several ways of visualizing torsion as an extra twist. For example, when the torsion vanishes, we have

$$T(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y] = 0.$$

The Lie bracket tells us how much the vector fields X and Y twist as they flow along each other. Thus when there is no torsion, the covariant derivative part has no extra twist.

Another intuition comes from geodesics. If we imagine making a geodesic on a surface by pulling a string taut, torsion corresponds to twisting the string. With torsion, a rigid body in $\mathbb{R}^3$ whose center of mass follows a geodesic rotates about the center of mass.

To formalize this, we demand the displacement vectors from the center of mass to points on the body are parallel transported. Then we define the body to not be rotating if those other points also follow geodesics. This makes sense dynamically, since rotating rigid bodies have internal stresses so that points in them experience external forces. **(clarify)**

Next, we explicitly construct locally inertial coordinates using geodesics.

- At every point p, there is a unique geodesic through p with tangent vector $dx/d\lambda = X_p$, by the existence and uniqueness theorems applied to the geodesic equation.

- Define the exponential map $\exp\colon T_p \to M$ as the map that takes X_p to the point of this geodesic at $\lambda = 1$. It can be shown that this map is bijective for X_p in a neighborhood of the origin of T_p, and hence defines a system of coordinates by $\exp(X) \mapsto X^\mu$.

- Note that geodesics in normal coordinates are simply linear in the coordinates, $X^\mu(t) = tX_p^\mu$. Therefore the geodesic equation reduces to $\Gamma^\mu_{\nu\rho}(X(t))X_p^\nu X_p^\rho = 0$. In particular, at $t = 0$

$$\Gamma^\mu_{\nu\rho}(p)X_p^\nu X_p^\rho = 0, \quad \Gamma^\mu_{(\nu\rho)}(p) = 0$$

  because X_p is arbitrary. If the connection is torsion-free, then $\Gamma^\mu_{[\nu\rho]} = 0$, so the connection coefficients vanish at p. (This doesn't work with torsion, because the torsion is a tensor, and hence is nonzero independent of coordinates.)

- Furthermore, in the case of the Levi–Civita connection, we can solve for ∂g in terms of Γ, which shows that $\partial g = 0$ at p.

- If we defined normal coordinates using a basis $\{e_\mu\}$ of T_p, then the $\{e_\mu\}$ are the coordinate basis vectors $\{\partial_\mu\}$ of the normal coordinates at p. Thus the metric in normal coordinates is determined by the inner products of the $\{e_\mu\}$.

- A locally inertial frame at p is a set of normal coordinates at p with $g = \eta$. Such frames physically correspond to observers in special relativity: the time and space coordinates are laid down by spatial and timelike geodesics, which correspond physically to rigid straight rods and unaccelerated clocks.

2.5 The Riemann Curvature Tensor

We have seen that second covariant derivatives commute for scalars. The extend to which they don't commute for vectors defines the Riemann curvature tensor.

- We define the Riemann curvature tensor $R^a{}_{bcd}$ of a connection ∇ by $R^a{}_{bcd}Z^bX^cY^d = (R(X,Y)Z)^a$ where, for any vector fields X, Y, and Z,

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

  To verify this is a tensor, we must check linearity in X, Y, and Z. For linearity in X, we have

$$\begin{aligned} R(fX,Y)Z &= \nabla_{fX}\nabla_Y Z - \nabla_Y\nabla_{fX}Z - \nabla_{[fX,Y]}Z \\ &= f\nabla_X\nabla_Y Z - \nabla_Y f\nabla_X Z - \nabla_{f[X,Y]-Y(f)X}Z \\ &= f\nabla_X\nabla_Y Z - Y(f)\nabla_X Z - f\nabla_Y\nabla_X Z - f\nabla_{[X,Y]}Z - Y(f)\nabla_X Z \end{aligned}$$

  where we use $\nabla_{fX}Y = f\nabla_X Y$ repeatedly, and the extra terms cancel as desired.

- We may also specialize to a coordinate basis. Let $e_\mu = \partial/\partial x^\mu$ and $\nabla_\mu = \nabla_{e_\mu}$. Then

$$R(e_\rho, e_\sigma)e_\nu = \nabla_\rho\nabla_\sigma e_\nu - \nabla_\sigma\nabla_\rho e_\nu = \nabla_\rho(\Gamma^\tau_{\nu\sigma}e_\mu) - \nabla_\sigma(\Gamma^\tau_{\nu\rho}e_\tau)$$

  since the commutator term is zero, and carrying out the covariant derivatives gives

$$\boxed{R^\gamma{}_{\rho\alpha\beta} = \partial_\alpha\Gamma^\gamma_{\rho\beta} - \partial_\beta\Gamma^\gamma_{\rho\alpha} + \Gamma^\gamma_{\mu\alpha}\Gamma^\mu_{\rho\beta} - \Gamma^\gamma_{\mu\beta}\Gamma^\mu_{\rho\alpha}.}$$

- If we further specialize to torsion-free connections, we have the Ricci identity

$$\boxed{\nabla_\alpha \nabla_\beta V^\gamma - \nabla_\beta \nabla_\alpha V^\gamma = R^\gamma{}_{\rho\alpha\beta} V^\rho}$$

  This is intuitive: the Riemann tensor takes in a vector (index ρ) and two directions (α and β), and returns the difference between parallel transporting along α and then β, and parallel transporting in the reverse order (index γ). Thus the Riemann tensor tells us about the path-dependence of parallel transport. The extra twist of torsion would add another term.

- To prove the Ricci identity in a coordinate basis, we may simply expand

$$\nabla_\alpha \nabla_\beta V^\gamma = \partial_\alpha(\nabla_\beta V^\gamma) - \Gamma^\rho_{\beta\alpha} \nabla_\rho V^\gamma + \Gamma^\gamma_{\rho\alpha} \nabla_\beta V^\rho.$$

  We expand the outer covariant derivative first; we do it in the opposite order, we have to take care to avoid writing covariant derivatives of connection coefficients, which are not tensors.

- Since the torsion vanishes, the second term vanishes upon antisymmetrizing α and β, as does the $\partial_\alpha \partial_\beta$ term. All terms involving derivatives acting on V cancel between the first and third term, recovering our coordinate basis expression for the Riemann tensor.

- The Ricci identity also holds outside of a coordinate basis,

$$\nabla_c \nabla_d Z^a - \nabla_d \nabla_c Z^a = R^a{}_{bcd} Z^b.$$

  To show this, we start from our original definition. The torsion-free condition gives

$$R(X,Y)Z = \nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - (\nabla_{\nabla_X Y} Z - \nabla_{\nabla_Y X} Z).$$

  On the other hand, we have

$$\nabla_X \nabla_Y Z = \nabla_X Y^d \nabla_d Z = (\nabla_X Y)^d \nabla_d Z + Y^d \nabla_X \nabla_d Z = \nabla_{\nabla_X Y} Z + X^c Y^d \nabla_c \nabla_d Z$$

  so we conclude that

$$X^c Y^d (\nabla_c \nabla_d Z - \nabla_d \nabla_c Z^a) = R^a{}_{bcd} Z^b X^c Y^d$$

  and since X and Y are arbitrary, this proves the Ricci identity.

- There is also a Ricci identity for covectors,

$$\nabla_c \nabla_d \eta_a - \nabla_d \nabla_c \eta_a = -R^b{}_{acd} \eta_b$$

  where again the only requirement is that the connection be torsion-free. Similarly, we have a Ricci identity for rank n tensors with n terms.

Example. Explicitly relating the Riemann tensor to parallel transport. Take two commuting vector fields X and Y. Then they define part of a coordinate system, $X = \partial/\partial s$ and $Y = \partial/\partial t$, which we will use to parametrize geodesics. Take a point p and work in normal coordinates so that the connection coefficients vanish at p.

Now consider a vector $Z_p \in T_p(M)$, and parallel transport it by δs along X to point q. The geodesic equation reads

$$\nabla_X Z = 0, \quad \frac{dZ^\mu}{ds} = -\Gamma^\mu_{\nu\rho} Z^\nu X^\rho, \quad \frac{d^2 Z^\mu}{ds^2} = -(\Gamma^\mu_{\nu\rho} Z^\nu X^\rho)_{,\sigma} X^\sigma.$$

Then expanding to second order gives

$$Z^\mu_q = Z^\mu_p - \frac{1}{2}(\Gamma^\mu_{\nu\rho,\sigma} Z^\nu X^\rho X^\sigma)|_p \delta s^2$$

where only the term involving a derivative of Γ survives, since we used normal coordinates. Next, we parallel transport by δt along Y to point r. Expanding to second order again,

$$Z^\mu_r = Z^\mu_q + \left(\frac{dZ^\mu}{dt}\right)_q \delta t + \frac{1}{2}\left(\frac{d^2 Z^\mu}{dt^2}\right)_q \delta t^2.$$

This expression can be simplified by Taylor expanding the connection coefficients at q, as

$$\Gamma|_q = \Gamma|_p + (X^\sigma \partial_\sigma \Gamma)|_p \, \delta s + O(\delta s^2).$$

By expanding to second order in δs and δt we find

$$Z^\mu_r = Z^\mu_p - \frac{1}{2}\Gamma^\mu_{\nu\rho,\sigma} Z^\nu (X^\rho X^\sigma \delta s^2 + Y^\rho Y^\sigma \delta t^2 + 2 Y^\rho X^\sigma)\Big|_p \delta s \delta t$$

We may also compute Z^r_μ by parallel transporting in the opposite order. The difference is

$$\Delta Z^\mu_r = Z'^\mu_r - Z^\mu_r = \Gamma^\mu_{\nu\rho,\sigma} Z^\nu (Y^\rho X^\sigma - X^\rho Y^\sigma)\delta s \delta t = (\Gamma^\mu_{\nu\sigma,\rho} - \Gamma^\mu_{\nu\rho,\sigma}) Z^\nu X^\rho Y^\sigma \delta s \delta t = R^\mu{}_{\nu\rho\sigma} Z^\nu X^\rho Y^\sigma \delta s \delta t$$

where we used our coordinate expression for the Riemann tensor at p. Now we're almost done, but the left-hand side contains a tensor at r and the right-hand side contains a tensor at p. If we evaluate the Riemann tensor at r instead, the error terms will be higher order in δs and δt and hence can be ignored. Therefore we have the tensorial expression

$$R^a{}_{bcd} Z^b X^c Y^d|_r = \lim_{\delta s, \delta t \to 0} \frac{\Delta Z^a_r}{\delta s \, \delta t}.$$

This justifies our intuitive interpretation. Here, we implicitly assumed that there was no torsion. If torsion were present, then parallel transporting along δs and then δt would land us at a different point (at order $\delta s \, \delta t$) than parallel transporting along δt and then δs, specifically differing by the torsion tensor $\delta p^\rho = T_{\mu\nu}{}^\rho X^\mu Y^\nu \delta s \delta t$.

The Riemann tensor has some important symmetries. Note that its definition makes no mention of the metric. Thus we begin with statements that only require a connection.

- Ignoring the matrix indices, the Riemann tensor has the simple form

  $$R^{\cdot}{}_{\cdot\mu\nu} = \partial_\mu \Gamma_\nu - \partial_\nu \Gamma_\mu + [\Gamma_\mu, \Gamma_\nu],$$

  which is similar to the field strength tensor in Yang-Mills. The analogy is

  $$A_\mu \leftrightarrow \Gamma_\mu, \quad F_{\mu\nu} \leftrightarrow R_{\mu\nu}.$$

  That is, both A and Γ are connections on a fibre bundle and F and R are their curvatures. The analogy is not perfect; in general relativity Γ is a derivative of the fundamental field g, but in Yang-Mills A is itself the fundamental field.

- By definition, we have antisymmetry in the last two indices,

$$R^a{}_{b(cd)} = 0.$$

  This holds even for connections with torsion.

- For a torsion-free connection, we have

$$R^a{}_{[bcd]} = 0.$$

  To see this, work in normal coordinates at p, so that

$$R^\mu{}_{\nu\rho\sigma} = \partial_\rho \Gamma^\mu_{\nu\sigma} - \partial_\sigma \Gamma^\mu_{\nu\rho}$$

  at p. Since the torsion is zero, $\Gamma^\mu_{[\nu\rho]} = 0$ everywhere. Then antisymmetrizing over ν, ρ, and σ gives $R^\mu{}_{[\nu\rho\sigma]} = 0$. Since this is a tensorial equation, it holds in all coordinate systems.

- For a torsion-free connection, we have the Bianchi identity

$$R^a{}_{b[cd;e]} = 0.$$

  We work in normal coordinates, so the covariant derivative reduces to the partial derivative. We have $\partial R \sim \partial(\partial\Gamma + \Gamma\Gamma) = \partial\partial\Gamma + \Gamma\partial\Gamma = \partial\partial\Gamma$ by normal coordinates, so

$$R^\mu{}_{\nu\rho\sigma;\tau} = \partial_\tau \partial_\rho \Gamma^\mu_{\nu\sigma} - \partial_\tau \partial_\sigma \Gamma^\mu_{\nu\rho}$$

  and antisymmetrizing gives the result, by the symmetry of mixed partials. Equivalently,

$$R^a{}_{bcd;e} + R^a{}_{bde;c} + R^a{}_{bec;d} = 0$$

  where we used antisymmetry in the last two indices. The Bianchi identity looks quite similar to the Jacobi identity and can be proven using it.

Next, we derive the geodesic deviation equation.

- Physically, we expect the Riemann tensor to be related to the relative acceleration of nearby geodesics, because the Riemann tensor describes the curvature of spacetime, which physically appears as a gravitational tidal force.

- Mathematically, suppose we have two initially parallel nearby geodesics, where "parallel" is defined by parallel transporting the velocity of one to the other. We propagate both geodesics for a small time, by parallel transporting them along their velocity vectors. Then their relative acceleration is proportional to their new relative velocity, which is again determined by parallel transport. We've essentially just described the commutator of two parallel transports, which is exactly what the Riemann tensor is.

- To formalize this, define a one-parameter family of geodesics to be a diffeomorphism

$$\gamma \colon I \times I' \to M, \quad (s,t) \mapsto \gamma(s,t)$$

  where I and I' are intervals, so that for fixed s, $\gamma(s,t)$ is a geodesic with affine parameter t.

- This defines a surface on M with coordinates (s, t), which we extend to coordinates for an open set containing the surface, and we define $S^\mu = \partial x^\mu / \partial s$ and $T^\mu = \partial x^\mu / \partial t$. Note that S and T are commuting vector fields since they are derived from coordinates.
- When evaluated on the surface, S^μ points from one geodesic to its neighbor, while T^μ gives the local geodesic velocity. Therefore the relative velocity of geodesics is $\nabla_T S$, and the relative acceleration of geodesics is $\nabla_T \nabla_T S$.
- Since the torsion vanishes,

  $$\nabla_T S - \nabla_S T = [T, S] = 0.$$

  Therefore we have

  $$\nabla_T \nabla_T S = \nabla_T \nabla_S T = \nabla_S \nabla_T T + R(T, S)T = R(T, S)T$$

  where we used the definition of the Riemann tensor and $\nabla_T T = 0$. This is the geodesic equation, relating the relative acceleration to the Riemann tensor.
- In abstract index notation, the geodesic deviation equation reads

  $$\boxed{T^c \nabla_c (T^b \nabla_b S^a) = R^a{}_{bcd} T^b T^c S^d.}$$

  It is useful because it gives us a direct way to measure the Riemann tensor. Since

  $$R^a{}_{bcd} = \frac{2}{3}(R^a{}_{(bc)d} - R^a{}_{(bd)c})$$

  we may use geodesic deviation to measure all of the components of the Riemann tensor.

2.6 Curvature of the Levi–Civita Connection

In this section, we consider additional properties of the Riemann tensor when it is derived from the Levi–Civita connection on a manifold with a metric.

Prop. On a simply-connected manifold, the Riemann tensor associated with the Levi–Civita connection vanishes if and only if there is a coordinate system where the metric components are constant.

Proof. The backwards direction is easy: if $\partial g = 0$ everywhere, the connection vanishes, so the Riemann tensor vanishes as well. To go forwards, take a basis of one-forms $\hat{\theta}^{(a)}$ at a point p, so

$$ds^2(p) = \eta_{ab} \hat{\theta}^{(a)} \otimes \hat{\theta}^{(b)}$$

where η_{ab} is arbitrary. Now extend the one-forms to one-form fields by parallel transport, which is unique because the Riemann tensor vanishes. Then by metric-compatibility, we have

$$ds^2 = \eta_{ab} \hat{\theta}^{(a)} \otimes \hat{\theta}^{(b)}$$

at every point, where η_{ab} is constant in this field of frames. We now must show that the $\hat{\theta}^{(a)}$ are derived from a coordinate system. Note that if a one-form ω is parallel transported,

$$\nabla_\mu \omega_\nu = 0$$

and antisymmetrizing both sides gives

$$\nabla_{[\mu}\omega_{\nu]} = \partial_{[\mu}\omega_{\nu]} \propto d\omega = 0$$

where we used torsion-freeness. On a simply-connected manifold, the first cohomology group is trivial, so all closed one-forms are exact, $\hat{\theta}^{(a)} = dy^a$, giving the desired coordinates.

Next, we consider additional symmetry properties of the Riemann tensor.

- Since we now have a metric, we work with the lowered form R_{abcd}. We claim that

$$R_{abcd} = R_{cdab}.$$

  To show this, work in normal coordinates at p, so $\partial g = 0$. Then $\partial g^{-1} = 0$ as well, as can be shown formally by considering $\partial\delta = \partial(gg^{-1})$. Then

$$\partial_\rho \Gamma^\tau_{\nu\sigma} = \frac{1}{2} g^{\tau\mu}(g_{\mu\nu,\sigma\rho} + g_{\mu\sigma,\nu\rho} - g_{\nu\sigma,\mu\rho})$$

  and plugging this into our expression for the Riemann tensor gives

$$R_{\mu\nu\rho\sigma} = \frac{1}{2}\left(g_{\mu\sigma,\nu\rho} + g_{\nu\rho,\mu\sigma} - g_{\nu\sigma,\nu\rho} - g_{\mu\rho,\nu\sigma}\right)$$

  where we lower the index only after the connection coefficients are substituted away. The conclusion follows using the symmetry of the metric and mixed partials.

- This result can be combined with our earlier results for the corollaries

$$R_{(ab)cd} = 0, \quad \nabla_{[e} R_{ab]cd}.$$

  By combining some of these results, we further have

$$R_{[abcd]} = 0, \quad \nabla_e R_{abcd} + \nabla_a R_{becd} + \nabla_b R_{eacd} = 0.$$

Next, we define special contractions of the Riemann tensor and write down the Einstein equation.

- In terms of group theory, tensors with the symmetry properties of the Riemann tensor form a representation of the Lorentz group, and we would like to decompose it into irreps. When this is done to a general rank 2 tensor, for example, we get the trace, the antisymmetric part, and the traceless symmetric part.

- We define the Ricci tensor by

$$R_{ab} = R^c{}_{acb}.$$

  The Ricci tensor is symmetric, and its trace is called the Ricci scalar,

$$R = g^{ab} R_{ab}.$$

  The Ricci tensor is the only independent contraction of the Riemann tensor; all others are either proportional to zero.

- We define the Einstein tensor by

$$\boxed{G_{ab} = R_{ab} - \frac{1}{2}Rg_{ab}.}$$

By contracting the Bianchi identity twice, we have

$$\boxed{\nabla^b R_{ab} = \frac{1}{2}\nabla_a R}$$

which implies that

$$\nabla^a G_{ab} = 0.$$

We say the Einstein tensor is covariantly conserved.

- In general relativity, we postulate the Einstein field equation

$$\boxed{G_{ab} = 8\pi G T_{ab}}$$

where the constant is fixed by Newtonian gravity. Taking the trace gives $R = -8\pi G T$, so

$$R_{ab} = 8\pi G\left(T_{ab} - \frac{1}{2}Tg_{ab}\right).$$

Hence R_{ab} is completely determined by T_{ab}. In vacuum, $R_{ab} = 0$.

- The alternate equation $R_{ab} \propto T_{ab}$ is unacceptable because covariant conservation of R implies $\nabla^a R_{ab} = 0$, which implies $\nabla_a R = 0$ by the Bianchi identity. Then R is constant which implies T is constant, but $T = 0$ in vacuum and $T \neq 0$ in matter.

- The remaining degrees of freedom are in the Weyl tensor, which is completely traceless, i.e. all of its contractions vanish. It is defined by

$$R_{abcd} = C_{abcd} + \frac{2}{n-2}(g_{a[c}R_{d]b} - g_{b[c}R_{d]a}) - \frac{2}{(n-1)(n-2)}Rg_{a[c}g_{d]b}$$

so that the Weyl tensor is essentially the Riemann tensor with all contractions subtracted off.

- The Weyl tensor has the same symmetries as the Riemann tensor, and vanishes in dimension $n < 4$. It represents gravitational degrees of freedom, i.e. the components of the Riemann tensor which are not determined by the Einstein field equations, which are exemplified by gravitational waves. Analogously, Maxwell's equations with sources don't uniquely specify the fields; we can always add on electromagnetic radiation.

- The Weyl tensor is invariant under conformal transformations $g \to \Omega^2(x)g$, while the Riemann tensor is not. We say a metric is conformally flat if it related to a flat metric by a conformal transformation; one can show that conformal flatness is equivalent to vanishing Weyl tensor.

Note. Counting the degrees of freedom of the Riemann tensor. Since the first two indices are antisymmetric, they have $n(n-1)/2$ degrees of freedom, as do the last two indices. The fact that the Riemann tensor is symmetric under their interchange means that we are left with the same number of degrees of freedom as an $m \times m$ symmetric matrix, $m(m+1)/2$, with $m = n(n-1)/2$.

The final constraint is $R_{[abcd]} = 0$. To understand what this constraint means, consider writing

$$R_{abcd} = A_{abcd} + R_{[abcd]}$$

which implies $A_{[abcd]} = 0$. Then every tensor can be written as the sum of a totally antisymmetric tensor $R_{[abcd]}$ plus a tensor whose totally antisymmetric part vanishes. None of the identities we used above place any constraints on $R_{[abcd]}$, so $R_{[abcd]} = 0$ contains $n(n-1)(n-2)(n-3)/4!$ new independent constraints. The final count is

$$D(n) = \frac{1}{2}\left(\frac{n(n-1)}{2}\right)\left(\frac{n(n-1)}{2}+1\right) - \frac{n(n-1)(n-2)(n-3)}{24} = \frac{n^2(n^2-1)}{12}$$

degrees of freedom in the Riemann tensor. We have

$$D(n) = \begin{cases} 0 & n = 1, \\ 1 & n = 2, \\ 6 & n = 3, \\ 20 & n = 4. \end{cases}$$

Then the Riemann tensor is always trivial in $n = 1$, determined by the Ricci scalar in $n = 2$, and determined by the Ricci tensor in $n = 3$. In $n = 4$, there are 10 extra degrees of freedom, captured in the Weyl tensor.

Note. Intrinsic versus extrinsic curvature. Intrinsic curvature is measured by the Riemann tensor; it can be detected by observers living on the manifold through parallel transport loops. Extrinsic curvature is defined by surfaces embedded in a higher-dimensional space, and measures the deviation of the embedding from a flat plane.

- The intrinsic curvature of a 1D manifold is always zero, since there are no nontrivial parallel transport loops. However, as a curve embedded in a larger space, it can have extrinsic curvature, measured by its radius of curvature.

- In 2D, the intrinsic curvature is determined by the Ricci scalar. For example, the torus, thought of as $\mathbb{R}^2/\mathbb{Z}^2$, has zero intrinsic curvature. However, the standard embedding of the torus in $\mathbb{R}^3$ as a donut has both nonzero intrinsic and extrinsic curvature.

- The intrinsic curvature is linked to topology by

  $$\chi(M) = \frac{1}{4\pi}\int_M R\sqrt{|g|}d^n x, \quad \chi(M) = 2(1-g).$$

  Here, $\chi(M)$ is the Euler characteristic, a topological invariant of the space, and g is the genus, which is one for the torus.

- A piece of paper rolled up into a cylinder, viewed as a surface in $\mathbb{R}^3$, has extrinsic curvature but no intrinsic curvature. The reason is that rolling up the paper without stretching it preserves lengths between points along the paper, so the intrinsic curvature stays the same.

- A two-sphere with metric

  $$ds^2 = a^2(d\theta^2 + \sin^2\theta d\phi^2)$$

  has constant intrinsic curvature; the Ricci scalar is $R = 2/a^2$. If it is embedded in $\mathbb{R}^3$ in the standard way, it also has constant extrinsic curvature.

In general relativity, we will almost always be concerned with intrinsic curvature.

Note. Intuition for the pieces of the Riemann tensor. Consider a small ball of free particles, in a coordinate system where all of them are initially at rest. Then the Ricci tensor tells us how the volume of this ball changes over time. To see this, work in normal coordinates and recall the geodesic deviation equation says

$$a^\mu = -R^\mu{}_{00\nu} S^\nu$$

where S is the separation between geodesics and a is the acceleration. Since $R_{\mu 00\nu} = R_{\nu 00\mu}$, we can choose a coordinate system aligned with the principle axes of the ellipsoid that the ball will deform into by the spectral theorem. Then

$$\text{second derivative of principle axis } i \propto -R^i{}_{00i}$$

with no summation, and since $R^0{}_{000} = 0$, we have

$$\text{second derivative of volume} \propto -R^\mu{}_{00\mu} = -R_{00}.$$

This is Raychaudhuri's equation. Thus, the Ricci tensor tells us about how small volumes changes over time. Einstein's equations say that

$$R_{00} = 4\pi G(\rho + p_x + p_y + p_z)$$

so contraction is caused by both energy density and pressure. (The idea that pressure is repulsive is a red herring, since a uniform pressure isn't repulsive even nonrelativistically.)

For ordinary, nonrelativistic matter, the pressure is negligible, so we recover Newtonian gravity. A photon gas has $p = \rho/3$, which implies that its gravitational effect is twice as large as one would guess from its energy density alone. Dark energy has $p = -\rho$, implying that positive dark energy has a repulsive gravitational effect. Finally, there's no expansion or contraction effect when $p = -\rho/3$. In cosmology, this kind of contribution roughly corresponds to spatial curvature.

More generally, for a ball with velocity v^μ, the result above becomes

$$\text{second derivative of volume} \propto -R_{\mu\nu} v^\mu v^\nu$$

so we can recover all components of the Ricci tensor by volume change. The Weyl tensor, on the other hand, represents degrees of freedom which don't change the volume. For example, a gravitational wave can stretch a ball in one direction and contract in another.

2.7 Non-Riemannian Geometries

We give a tour of some exotic, non-Riemannian geometries.

Example. Newton–Cartan geometries. Consider the Newtonian equation of motion $\ddot{\mathbf{x}} = -\nabla\phi(\mathbf{x})$. We can interpret these as the geodesic equation

$$\ddot{x}^\mu + \Gamma^\mu_{\nu\rho} \dot{x}^\nu \dot{x}^\rho = 0$$

with an appropriate connection. We parametrize the geodesics by t, so the dot indicates differentiation with respect to t, and set

$$\Gamma^i_{00} = \delta^{ij} \partial_j \phi$$

with others zero. Parametrizing by t is legal since, in the nonrelativistic limit, $t = \tau$. Then

$$\ddot{x}^i + \Gamma^i_{00}\dot{t}\dot{t} = 0$$

which gives the result since $\dot{t} = 1$. However, this connection is not a Levi–Civita connection with respect to any metric!

More generally, we would like to formulate Newtonian gravity in a geometrical way. We define a Newton–Cartan structure as a triple (h, θ, ∇) so that

- h is a degenerate symmetric rank 2 tensor $h = h^{\mu\nu}\partial_\mu \otimes \partial_\nu$ so that $h^{\mu\nu}$ has rank $n-1$. We call h the spatial metric.
- θ is a one-form in the kernel of h, $h^{ab}\theta_b = 0$, called the clock.
- ∇ is a torsion-free connection which preserves both h and θ, i.e. $\nabla h = \nabla\theta = 0$.

Since ∇ is torsion-free, we have $\nabla_{[a}\theta_{b]} = 0$, which implies $d\theta = 0$, so $\theta = dt$ locally. This gives a time for every point in the spacetime, and h restricted to a surface of fixed t is nondegenerate.

For example, if $h^{ij} = \delta^{ij}$ with all other elements zero, then $\theta = dt$, and ∇ with $\Gamma^i_{00} \neq 0$, recovering Newtonian geodesics $\mathbf{x}(t)$. A Newton–Cartan structure has absolute time, since we can always say if two points have the same t, but not absolute space.

Example. Projective structures. Two torsion-free connections ∇, $\hat{\nabla}$ are projectively equivalent if they share the same unparametrized geodesics. For a general parameter, the geodesic equation reads

$$\ddot{x}^\mu + \Gamma^\mu_{\nu\rho}\dot{x}^\nu\dot{x}^\rho = f\dot{x}^\mu.$$

In particular, we can get unparametrized geodesics by using one of the coordinates x^n as the parameter. One can show that ∇ and $\hat{\nabla}$ are projectively equivalent if there is a one-form ω so

$$\hat{\Gamma}^\mu_{\nu\rho} = \Gamma^\mu_{\nu\rho} + \delta^\mu_\nu\omega_\rho + \delta^\mu_\rho\omega_\nu.$$

Projective equivalence defines an equivalence relation on connections, and a projective structure is one of these equivalence classes. One can also define curvatures of projective structures.

As an example, consider a hemispherical bowl sitting on a plane. Projecting from the center of the sphere defines a map between the half-sphere and $\mathbb{R}^2$, where great circles map to lines. Hence the Levi–Civita connections of the standard metric on the half-sphere yields a connection on $\mathbb{R}^n$ that is projectively equivalent to the standard one.

Example. Magnetic geometries and Kaluza–Klein reduction. Consider a three-dimensional Riemannian manifold with coordinates (x, y, z) and metric $g = dx^2 + dy^2 + (dz - x dy)^2$. The geodesic Lagrangian is

$$L = \frac{1}{2}\dot{x}^2 + \frac{1}{2}\dot{y}^2 + \frac{1}{2}(\dot{z} - x\dot{y})^2.$$

The Euler-Lagrange equations read

$$\ddot{x} = -\dot{y}(\dot{z} - x\dot{y}), \quad \frac{d}{dt}(\dot{y} - x(\dot{z} - x\dot{y})) = 0, \quad \dot{z} - x\dot{y} = c.$$

Plugging in the final equation, we find

$$\ddot{x} = -c\dot{y}, \quad \ddot{y} = c\dot{x}$$

which, if we project away the z direction, describes a particle in two dimensions in a magnetic field!

More generally, magnetism in two dimensions is described by

$$\ddot{x}^i + \Gamma^i_{jk}\dot{x}^j\dot{x}^k = F^i{}_j\,\dot{x}^j, \quad F_{ij} = -c\,\epsilon_{ij}$$

The Kaluza–Klein model does the same for electromagnetism in four dimensions. In general, given a manifold Σ with metric h and F with $F = dA$, we look at a manifold with one higher dimension with metric $g = h + (dz + A)^2$.

3 Equations in Curved Spacetime

3.1 Minimal Coupling

Physical laws should exhibit general covariance; they should not depend on the choice of coordinates, and should be ideally written in a manifestly coordinate-invariant form.

- In special relativity, we wrote equations that were manifestly invariant under Lorentz transformations. We can often make these equations generally covariant using the minimal coupling procedure:

  1. Replace the Minkowski metric with a general metric.
  2. Replace partial derivatives with covariant derivatives associated with the Levi–Civita connection; this is called minimal coupling.
  3. To write the result in abstract indices, replace Greek letters (referring to inertial frames only) with Latin ones.

  To go back, we simply work in a locally inertial frame, which reverses the above steps.

- The minimal coupling prescription is not unique, because we can always add terms depending on the curvature, which vanish in flat spacetime. For instance, the action for a scalar field can contain the term $\xi R\phi^2$, where ξ is a dimensionless free parameter. (However, higher powers of the curvature would generally be expected to be suppressed by powers of M_{pl}.)

- As another warning, you can't consistently generalize *every* equation in a theory with minimal coupling, because partial derivatives always commute, while covariant derivatives don't. If one generalizes the basic equations of motion of a theory using minimal coupling, then derived equations higher order in derivatives will generally contain commutators of covariant derivatives, and hence explicit curvature terms.

Example. A massless scalar field satisfies the equation

$$\eta^{\mu\nu}\partial_\mu\partial_\nu\phi = 0.$$

Applying minimal coupling gives the result

$$g^{ab}\nabla_a\nabla_b\phi = \nabla^a\nabla_a\phi = \phi_{;a}{}^a = 0.$$

The operator $\nabla^a\nabla_a$ is the Laplacian in curved spacetime, also called the Laplace-Beltrami operator. To get a more explicit expression, note that for the Levi–Civita connection

$$\nabla_\mu V^\mu = \partial_\mu V^\mu + \Gamma^\mu_{\mu\lambda}V^\lambda, \quad \Gamma^\mu_{\mu\lambda} = \frac{1}{2}g^{\mu\rho}\partial_\lambda g_{\rho\mu}.$$

The last term is heuristically $(1/2)\operatorname{tr} g^{-1}\partial g$. To simplify this, note that for a matrix A,

$$\det(A+\epsilon B) = \exp\operatorname{tr}\log(A+\epsilon B) = \exp\operatorname{tr}(\log A + \epsilon A^{-1}B) = (\det A)\exp(\epsilon\operatorname{tr}(A^{-1}B))$$

from which we conclude

$$\frac{\partial(\det A)}{\partial(\delta A)} = (\det A)\operatorname{tr}(A^{-1}\delta A).$$

Here we set $A = g$ and $\delta A = \partial_\lambda g$. Multiplying both sides by -1 because $g \equiv \det g_{\mu\nu}$ is negative,

$$\partial_\lambda \log(-g) = g^{\mu\rho}\partial_\lambda g_{\rho\mu}.$$

Then the divergence of a vector and Laplacian of a scalar become

$$\nabla_\mu V^\mu = \frac{1}{\sqrt{-g}}\partial_\mu(\sqrt{-g}V^\mu), \quad \nabla_\mu \nabla^\mu \phi = \frac{1}{\sqrt{-g}}\partial_\mu(\sqrt{-g}\partial^\mu\phi)$$

where we used the product rule, and the index on ∂^μ is raised with the metric g. These expressions are useful since they only involve the metric and its determinant.

Example. The formulas above give quick shortcuts to compute the divergence and Laplacian in curvilinear coordinates. For example, in spherical coordinates on $\mathbb{R}^3$, $g = r^4 \sin^2\theta$, so

$$\begin{aligned}\nabla \cdot \mathbf{F} &= \frac{1}{r^2 \sin\theta}\Big(\partial_r(r^2 \sin\theta\, F^r) + \partial_\theta(r^2\sin\theta\, F^\theta) + \partial_\phi)(r^2 \sin\theta\, F^\phi)\Big) \\ &= \frac{1}{r^2}\partial_r(r^2 F^r) + \frac{1}{\sin\theta}\partial_\theta(\sin\theta\, F^\theta) + \partial_\phi F^\phi.\end{aligned}$$

Example. In flat spacetime, the source-free Maxwell equations take the form

$$\eta^{\mu\nu}\partial_\mu F_{\nu\rho} = 0, \quad \partial_{[\mu}F_{\nu\rho]} = 0.$$

Thus in curved spacetime we have

$$g^{ab}\nabla_a F_{bc} = 0, \quad \nabla_{[a}F_{bc]} = 0.$$

The Lorentz force law for a particle of charge q and mass m is

$$\frac{du^\mu}{d\tau} = \frac{q}{m}\eta^{\mu\nu}F_{\nu\rho}u^\rho.$$

Noting as before that the left-hand side is $u^\nu\partial_\nu u^\mu$, performing minimal coupling gives

$$u^b\nabla_b u^a = \frac{q}{m}g^{ab}F_{bc}u^c.$$

As expected, if the field vanishes we simply get back the geodesic equation.

Example. A single particle. Consider a particle with four-momentum p^μ and an observer which has four-velocity w^μ. In the case where $w^\mu = (1, 0, 0, 0)$, the energy is p^0, so in general it must be

$$E = -\eta_{\mu\nu}w^\mu p^\nu$$

in flat spacetime. Moreover, the particle's mass is $m^2 = -\eta_{\mu\nu}p^\mu p^\nu$. These equations become

$$m^2 = -g_{ab}p^a p^b, \quad E = -g_{ab}w^a p^b$$

in curved spacetime. Note that in general relativity, it is only possible for an observer to measure a particle's energy if their locations coincide.

3.2 The Stress-Energy Tensor

We now introduce the stress-energy tensor, initially focusing on flat spacetime.

- The stress-energy tensor is a rank two symmetric tensor $T^{\mu\nu}$. With mixed indices, $T^{\mu}{}_{\nu}$ describes the flux of four-momentum p^μ across a surface of constant x^ν.
  - $T^0{}_0$ is the flux of energy across a surface of constant t, or in other words, the energy density. Similarly, $T^i{}_0$ is the momentum density.
  - Therefore, in a general frame, the energy-momentum current measured by an observer with four-velocity u^a is $j^a = T^a{}_b u^b$.
  - $T^0{}_i$ is the flux of energy across a surface of constant x^i. It is equal to $T^i{}_0$ because momentum corresponds to the spatial flow of energy, which is shown in the notes on Quantum Field Theory by applying Noether's theorem to boosts. (Note this doesn't hold in curved spacetime, and we shouldn't expect it to, since boosts are no longer a symmetry.)
  - $T^i{}_j$ is the flux of momentum p^i across a surface of constant x^j, i.e. the i-component of the force on a surface of constant x^j. It is also called the stress tensor and used in nonrelativistic continuum mechanics.
  - In the rest frame of a fluid, the $T^i{}_i$ are pressures and the off-diagonal elements are shear stresses. For each piece of the fluid, we must have $T^i{}_j = T^j{}_i$ for internal torques to balance.
- In special relativity, conservation of the stress-energy tensor is expressed by $\partial^\mu T_{\mu\nu} = 0$. Then in general relativity, we postulate
$$\nabla^a T_{ab} = 0.$$
- Dust is a set of free particles with uniform density and uniform velocity u^μ. Then the number flux is $N^\mu = nu^\mu$ where n is the number density in the rest frame. In the rest frame, we have $T^{00} = \rho = mn$ with all other elements zero. To generalize to arbitrary frames, note that this agrees with the tensorial equation $T = N \otimes p$, where p^μ is the momentum of each particle, so
$$T^{\mu\nu} = \rho u^\mu u^\nu.$$
  This is the stress-energy tensor of dust, a pressureless fluid.
- More generally, we would like to consider a fluid with a pressure p in its rest frame. That is, in the rest frame, $T^{\mu\nu} = \mathrm{diag}(\rho, p, p, p)$. This generalizes to
$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu + p\eta^{\mu\nu}$$
  and in general relativity η is simply promoted to g. An alternative definition of a perfect fluid is one which has no viscosity or heat conduction. In the rest frame in flat spacetime, these conditions set $T^{ij} = 0$ for $i \neq j$ and $T^{0i} = 0$ respectively, and isotropy sets $T^{11} = T^{22} = T^{33}$.
- Finally, dark energy is a perfect fluid which we demand is Lorentz invariant, i.e. it must look the same in all inertial reference frames. Then we cannot use u^μ to build $T^{\mu\nu}$, so
$$T^{\mu\nu} = -\rho\eta^{\mu\nu}$$
  which again generalizes by promoting η to g.

- The relationship between p and ρ is called the equation of state. In cosmology, we often have $p = w\rho$. Dust has $w = 0$, a photon gas has $w = 1/3$, and dark energy has $w = -1$. On the other hand, in stellar astronomy, we often have a "polytropic" equation of state $p \propto \rho^\gamma$.

Example. Conservation of the stress-energy tensor for a perfect fluid gives

$$\nabla_\alpha T^{\alpha\beta} = (\partial_\alpha \rho + \partial_\alpha p)u^\alpha u^\beta + (\rho + p)(u^\beta \nabla_\alpha u^\alpha + u^\alpha \nabla_\alpha u^\beta) + (\partial_\alpha p) g^{\alpha\beta} = 0.$$

Now we contract both sides with u_β, noting that

$$u_\beta u^\beta = -1, \quad u_\beta \nabla_\alpha u^\beta = \nabla_\alpha (u_\beta u^\beta) = 0$$

which yields

$$u^\alpha \nabla_\alpha \rho + (\rho + p)\nabla_\alpha u^\alpha = \nabla_\alpha(\rho u^\alpha) + p \nabla_\alpha u^\alpha = 0.$$

Taking the Newtonian limit, the pressure term is negligible, so ρu^α is a conserved current. Therefore we have the law of mass conservation. Plugging this result back into our original equation for one of the $\rho + p$ terms, we find

$$(\rho + p) u^\alpha \nabla_\alpha u^\beta = -(g^{\alpha\beta} + u^\alpha u^\beta) \nabla_\alpha p$$

which reduces to the Euler equation in the Newtonian limit.

Example. Energy is not conserved in general relativity. As an explicit example, consider an isotropic universe filled with a perfect fluid, so

$$g_{\mu\nu} = \mathrm{diag}(-1, a^2, a^2, a^2)$$

in comoving coordinates. In the rest frame of the fluid, the mass conservation law above becomes

$$\dot{\rho} = -3\frac{\dot{a}}{a}(\rho + p)$$

which implies

$$\rho \propto a^{-3(1+w)}.$$

However, the spatial volume scales as $\sqrt{-g} = a^3$, so the energy is not conserved. This is to be expected, since the time variation of a breaks time-translational invariance.

Note. One can think of the energy as going into gravitational field energy. However, this is difficult to make precise. In the Newtonian limit, the energy is proportional to $(\nabla\phi)^2$, so we expect it is proportional to $(\partial g)^2$ here. But we can always set $\partial g = 0$ at any point by working in normal coordinates. Hence it is impossible to unambiguously define gravitational field energy locally.

There are two options at this point. One option is to only define the total energy globally, which can be done assuming asymptotic flatness, leading to the Komar mass or ADM mass, or other variants. The other option is to give up on a tensorial definition. One can define energy-momentum pseudotensors that are coordinate-dependent but locally conserved in all coordinates, which establishes that energy cannot "teleport". However, since they are strongly coordinate-dependent their precise physical meaning is controversial.

Note. How is energy non-conservation consistent with the conservation of the stress-energy tensor? For a conserved vector, $\nabla_\mu J^\mu = 0$, we have a conserved quantity by the divergence theorem. For a conserved symmetric tensor such as $T^{\mu\nu}$, it is not possible to simply "set $\nu = 0$" to get a conserved vector. However, for a conserved symmetric rank r tensor it is generally true that we can form a conserved vector by contraction with $r - 1$ Killing vectors.

Example. The electromagnetic field. The energy density is

$$\epsilon = \frac{E_i E_i + B_i B_i}{8\pi}$$

and the energy flux density is given by the Poynting vector

$$S_i = \frac{1}{4\pi}\epsilon_{ijk}E_j B_k.$$

The stress tensor is

$$t_{ij} = \frac{1}{4\pi}\left(\frac{1}{2}(E_k E_k + B_k B_k)\delta_{ij} - E_i E_j - B_i B_j\right).$$

Then conservation of energy and momentum are written as

$$\frac{\partial \epsilon}{\partial t} + \partial_i S_i = 0, \qquad \frac{\partial S_i}{\partial t} + \partial_j t_{ij} = 0.$$

In special relativity, all of these results are combined into the stress-energy tensor

$$T_{\mu\nu} = \frac{1}{4\pi}\left(F_{\mu\rho}F_\nu{}^\rho - \frac{1}{4}F^{\rho\sigma}F_{\rho\sigma}\eta_{\mu\nu}\right), \qquad \partial_\mu T^{\mu\nu} = 0.$$

This can also be derived from Noether's theorem, but using the theorem directly will yield a result that is neither symmetric nor gauge invariant. The expression above is the "improved" version, which does not have these problems. It is easy to guess, because the two terms present are the only possible ones that are symmetric, gauge invariant, and bilinear in the field strength; their coefficients are fixed by the energy density $T^{00} = \epsilon$.

In general relativity, the stress-energy tensor above is the physically correct one; the only changes are that we must replace η with g, and raise and lower indices using g. Then Maxwell's equations imply $\nabla_a T^{ab} = 0$.

Example. A single particle of mass m, following a path $q^\mu(\tau)$. The number flux is

$$N^\mu(x) = \int d\tau\, \frac{dq^\mu}{d\tau}\delta(x - q(\tau)).$$

Weighting this by momentum, the stress-energy tensor is

$$T^{\mu\nu}(x) = \int d\tau\, \frac{dq^\mu}{d\tau}p^\nu \delta(x - q(\tau)) = \frac{1}{m}\int d\tau\, p^\mu p^\nu \delta(x - q(\tau)).$$

More explicitly, we can perform the integral by splitting the delta function as $\delta(t - q^0(\tau))\delta(\mathbf{x} - \mathbf{q}(\tau))$,

$$N^\mu(\mathbf{x}, t) = \frac{dq^\mu/d\tau}{dq^0/d\tau}\delta(\mathbf{x} - \mathbf{q}(\tau))\bigg|_{\tau=\tau_0} = \frac{dq^\mu}{dq^0}\delta(\mathbf{x} - \mathbf{q}(\tau))\bigg|_{\tau=\tau_0}.$$

Similarly, the stress-energy tensor is

$$T^{\mu\nu}(\mathbf{x}, t) = \frac{p^\mu p^\nu}{m\, dq^0/d\tau}\delta(\mathbf{x} - \mathbf{q}(\tau))\bigg|_{\tau=\tau_0} = \frac{p^\mu p^\nu}{E}\delta(\mathbf{x} - \mathbf{q}(\tau))\bigg|_{\tau=\tau_0}$$

where τ_0 is the proper time when $q^0 = t$. This final expression looks a bit unfamiliar, but it reduces to what we expect for a gas; the energy density averages over E while the pressure averages over $p^i p^j/E = (p^2/3E)\delta^{ij}$ as expected.

3.3 Einstein's Equation

In this section we cover some of the physics of the Einstein field equation.

- Einstein's equation is a set of ten second-order differential equations for g_{ab}. However, the contracted Bianchi identity $\nabla^a G_{ab} = 0$ reduces the number of independent equations to six. Physically, this is because any metric satisfying Einstein's equation should also satisfy it in any other coordinate system, using up four degrees of freedom.

- Einstein's equation is nonlinear: it does not obey a superposition principle. Upon quantization, gravitons couple to themselves. This self-interaction is required physically: consider two particles bound by gravity. The inertial mass of the system is less than the sum of the inertial masses of the particles because of the negative gravitational binding energy. For the gravitational mass to be equal to the inertial mass, gravity must couple to this binding energy.

- Einstein's equation makes the geodesic equation redundant; it can be shown that a test mass follows geodesics using only energy-momentum conservation. The simplest way to see this is to note that for dust, $T^{ab} = \rho U^a U^b$, and using $\nabla_a T^{ab} = 0$ produces $U^a \nabla_a U^b = 0$.

- We may also postulate a vacuum energy density ρ_v, which would have $T_{\mu\nu} = -\rho_v g_{\mu\nu}$. Then Einstein's equation becomes

  $$R_{ab} - \frac{1}{2} R g_{ab} = 8\pi G (T_{ab} - \rho_v g_{ab})$$

  where T_{ab} does not include the vacuum energy, or equivalently

  $$R_{ab} - \frac{1}{2} R g_{ab} + \Lambda g_{ab} = 8\pi G T_{ab}.$$

  Thus the cosmological constant can be thought of as either a property of spacetime or a form of matter in spacetime.

- Therefore, vacuum solutions of the Einstein field equations in $d = 4$ satisfy

  $$R_{ab} = \Lambda g_{ab}.$$

  Manifolds where the Ricci tensor is proportional to the metric are called Einstein manifolds.

- Lovelock's theorem states that the only symmetric, covariantly conserved tensor that depends on g, ∂g and $\partial^2 g$, in $d = 4$ with at most linear dependence on $\partial^2 g$, is a linear combination of G_{ab} and g_{ab}. Thus there is no further freedom to modify the Einstein equation.

Note. The Newtonian limit. Parametrize a geodesic with proper time, and assume it is moving slowly. Then to lowest order in v/c the geodesic equation is

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma^\mu_{00} = 0$$

Assuming the metric is static, so $\partial_0 g^{\mu\nu} = 0$, we have

$$\Gamma^\mu_{00} = -\frac{1}{2} g^{\mu\lambda} \partial_\lambda g_{00}.$$

Next, we assume the gravitational field is weak, so that we may perturb the metric about η,

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}.$$

In general, we can expand everything else as a perturbation series in h. For example,

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu} + O(h^2)$$

where the indices on h are raised using η, not g, as the error is second order in h. Returning to our computation, to lowest order in h, we have

$$\frac{d^2x^\mu}{d\tau^2} = \frac{1}{2}\eta^{\mu\nu}\partial_\nu h_{00}, \quad \frac{d^2x^i}{dt^2} = \frac{1}{2}\partial_i h_{00}.$$

This identifies the Newtonian gravitational potential Φ via

$$g_{00} = -(1+2\Phi).$$

For example, in the Schwarzschild metric, we have $\Phi = -GM/r$ as expected.

Note. Motivating Einstein's equation. We wish to generalize the equation

$$\nabla^2\Phi = 4\pi G\rho.$$

We know that ρ generalizes to $T^{\mu\nu}$, and we've seen that Φ is a metric component in the Newtonian limit, so the left-hand side should contain second derivatives of the metric. This is enough to motivate a guess of the form

$$G_{\mu\nu} = \kappa T_{\mu\nu}.$$

To pin down the constant κ, note the our equation is equivalent to

$$R_{\mu\nu} = \kappa\left(T_{\mu\nu} - \frac{1}{2}Tg_{\mu\nu}\right).$$

Now consider a spacetime with sparse dust. In the rest frame, $T^{00} = \rho u^0 u^0$ with all other elements zero. As above we write $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ and work to lowest order in h. We know u^0 is set by the normalization condition $g_{\mu\nu}u^\mu u^\nu = -1$, so to lowest order $u^0 = 1$ and

$$R_{00} \approx \frac{1}{2}\kappa\rho.$$

Next, we compute R_{00}. We may ignore products of connection coefficients because they are higher-order in h, and time derivatives of the metric vanish since the situation is static, so

$$R^i{}_{0j0} = \partial_j\Gamma^i_{00} - \partial_0\Gamma^i_{j0} + \Gamma^i_{j\lambda}\Gamma^\lambda_{00} - \Gamma^i_{0\lambda}\Gamma^\lambda_{j0} \approx \partial_j\Gamma^i_{00} = -\frac{1}{2}\partial_j\partial_i h_{00}$$

which implies that

$$R_{00} = R^i{}_{0i0} = -\frac{1}{2}\nabla^2 h_{00}.$$

Above, we identified $h_{00} = -2\Phi$, so the Newtonian limit works if $\kappa = 8\pi G$.

4 Further Geometry

4.1 Diffeomorphisms

We quickly review pushforward and pullback and how they relate to our operations.

- Given a smooth map $\phi\colon M \to N$ we define the pullback of a function

$$\phi^*(f) = f \circ \phi$$

  and the pushforward of a vector X and one-form η as

$$(\phi_* X)(f) = X(\phi^* f), \quad (\phi^*\eta)(X) = \eta(\phi_* X)$$

  where $f(p) = q$, and everything in M is evaluated at p and everything in N is evaluated at q. More generally, we can always pullback one-form fields but *can't* pushforward vector fields. This is due to an asymmetry in the definition of a function; for every input, there is exactly one output, but not vice versa.

- In components, we have

$$(\phi_* X)^\alpha\Big|_q = \frac{\partial y^\alpha}{\partial x^\mu} X^\mu\Big|_p, \quad (\phi^*\eta)_\mu\Big|_p = \frac{\partial y^\alpha}{\partial x^\mu}\eta_\alpha\Big|_q.$$

  One can remember these expressions by the chain rule. Schematically, we have $\eta^T(JX) = (J^T\eta)^T X$ so the factors are related by transpose.

- The pullback and pushforward of arbitrary tensors of type $(r, 0)$ or $(0, r)$ is defined by pushforward and pullback of their arguments, giving a Jacobian factor for every tensor index.

- As an example, suppose ϕ embeds M in N. Then we can define a metric on M by pulling back the metric on N.

- Pullback and pushforward are linear and commute with the tensor product and with contraction. Pullback commutes with the exterior derivative for functions,

$$\phi^* df = d(\phi^* f).$$

  Since the pullback is linear and commutes with tensor products (which include multiplying by an arbitrary scalar), it commutes with the exterior derivative and wedge product.

- Finally, we have $\phi^*[X, Y] = [\phi^* X, \phi^* Y]$. If we are willing to extend the definition of pushforward to an arbitrary differential operator, this follows from $\phi_*(XY) = (\phi_* X)(\phi_* Y)$.

Next, we specialize to the case where ϕ is a diffeomorphism.

- If ϕ is a diffeomorphism, we can define the pushforward and pullback of arbitrary tensor fields by applying $(\phi^{-1})_*$ and $(\phi^{-1})^*$ when needed. For example,

$$(\phi_* T)^\mu{}_\nu = \frac{\partial y^\mu}{\partial x^\rho}\frac{\partial x^\sigma}{\partial y^\nu} T^\rho{}_\sigma.$$

  These formulas contain a mix of Jacobians and inverse Jacobians. While we typically think of diffeomorphisms as active, in the passive view they are simply coordinate changes, explaining why the formulas look similar.

- Given a covariant derivative ∇ on M, we can define its pushforward $\tilde{\nabla}$ as
$$\tilde{\nabla}_X T = \phi_*(\nabla_{\phi^* X}(\phi^* T)).$$
That is, we pullback X and T, evaluate the covariant derivative on M, then pushforward the result. One can show this respects our previous results: the Riemann tensor of $\tilde{\nabla}$ is the pushforward of the Riemann tensor of ∇, and if ∇ is the Levi–Civita connection of a metric g on M, then $\tilde{\nabla}$ is the Levi–Civita connection of the metric $\phi_*(g)$ on N.

- We say a diffeomorphism $\phi\colon M \to M$ is a symmetry transformation of a tensor field T if $\phi_* T = T$. A symmetry transformation of the metric is called an isometry.

Example. We have to be careful with coordinates when $M = N$. Consider $\phi\colon \mathbb{R} \to \mathbb{R}$, where both copies of $\mathbb{R}$ are described by the coordinate x, so we are considering an active transformation. Let ϕ map the point p with coordinate x to the point q with coordinate $f(x)$. Then if $\eta(x) = g(x)\,dx$,
$$(\phi^*\eta)|_p = f' g|_q\,dx, \quad (\phi^*\eta)(x) = f'(f(x))g(f(x))\,dx$$
by our earlier formulas. As a simple example, if $f(x) = 2x$, then we have
$$(\phi^*\eta)(x) = 2g(2x)\,dx.$$
Note that the argument of g does not match with that of $\phi^*\eta$. Since everything here is active, there is no notion of a coordinate change.

On the other hand, suppose we used different coordinates for the two copies of $\mathbb{R}$, i.e. we could let ϕ map x to $f(y)$, where $y = h(x)$. In order to compute the pullback of $\eta(x) = g(x)\,dx$, the first step would be to change coordinates to y,
$$\eta(x) = g(x)\,dx = g(x)\,\frac{dx}{dy}\,dy = g(h^{-1}(y))\,\frac{dh^{-1}}{dy}\bigg|_{y=h(x)} dy.$$
Then we proceed with the formula above. While this is a bit conceptually clearer, it's technically messier, so we usually prefer to use a single coordinate system.

Diffeomorphisms are the gauge symmetries of general relativity.

- A physical situation is specified by a set of tensor fields on a manifold; if two of these sets are related by a diffeomorphism, they are physically equivalent. That is, if we have $\phi\colon M \to N$ with tensor fields Φ on M, including the metric, then (M, Φ) is physically the same as $(N, \phi_*\Phi)$.

- For example, the statement that two geodesics intersect at a point p in a manifold is not gauge-invariant, because we can map p to any other point. However, the proper time experienced by a geodesic between two collisions is gauge-invariant.

- Thus, diffeomorphism invariance tells us that individual points on a manifold have no physical meaning. Once we have accepted this, the only additional thing diffeomorphism invariance says is that the theory should be coordinate-independent, which we already know.

- It is possible to derive general relativity by starting with a massless spin 2 field on flat spacetime with a gauge symmetry derived from diffeomorphism invariance.

- The vacuum Einstein equation appears to be ten equations for ten components of the metric. However, four degrees of freedom of the metric are redundant due to diffeomorphism invariance (i.e. coordinate transformations). The resolution is that four of the Einstein equations are redundant by the contracted Bianchi identity.

4.2 The Lie Derivative

Given a vector field X, let $\phi_t \colon M \to M$ be the diffeomorphism resulting from following integral curves for parameter t. This gives a one-parameter group of diffeomorphisms.

- We define the Lie derivative of a tensor field T as

$$(\mathcal{L}_X T)_p = \lim_{t\to 0} \frac{((\phi_{-t})_* T)_p - T_p}{t}.$$

  We could also write this with a pullback, $(\phi_t)^*$, but we'll use pushforward by the inverse because we want to consider vectors explicitly first. The Lie derivative is clearly linear and maps (r, s) tensors to (r, s) tensors.

- To guess an expression for the Lie derivative, we pick a hypersurface with coordinates x^i so that X is nowhere tangent to the hypersurface. We then define the coordinate t so that $t = 0$ on the hypersurface and (x^i, t) is the point reached by flowing from $(x^i, 0)$ for parameter t. Then in these coordinates, $X = \partial/\partial t$ and we can explicitly compute $\mathcal{L}_X T = \partial T/\partial t$.

- Therefore, in this coordinate system, the Lie derivative commutes with contraction and obeys the Leibniz rule

$$\mathcal{L}_X(S \otimes T) = (\mathcal{L}_X S) \otimes T + S \otimes \mathcal{L}_X T.$$

  It also clearly obeys

$$\mathcal{L}_X(f) = X(f), \quad \mathcal{L}_X Y = [X, Y].$$

  Since these results are tensorial, they hold in all coordinates.

- Applying the Leibniz rule gives the Lie derivative of a one-form,

$$(\mathcal{L}_X \omega)_\mu = X^\nu \partial_\nu \omega_\mu + \omega_\nu \partial_\mu X^\nu.$$

  Note the sign of the second term. Intuitively, a flow that slows down into a "traffic jam" makes vectors smaller, but one-forms bigger. The expression for a general tensor is similar,

$$\boxed{\mathcal{L}_X T^{\mu_1 \cdots}{}_{\nu_1 \cdots} = X^\sigma \partial_\sigma T^{\mu_1 \cdots}{}_{\nu_1 \cdots} - (\partial_\lambda X^{\mu_1}) T^{\lambda \cdots}{}_{\nu_1 \cdots} - \ldots + (\partial_{\nu_1} X^\lambda) T^{\mu_1 \cdots}{}_{\lambda \cdots} + \ldots .}$$

  That is, we have the intuitive $(X \cdot \partial) T$ term, along with a $(\partial X) \cdot T$ for each index on T.

- The Lie derivative obeys the identity

$$\mathcal{L}_{[X,Y]} = [\mathcal{L}_X, \mathcal{L}_Y].$$

  For scalars, it follows by the definition of the commutator; for vectors, it follows from the Jacobi identity. Using the same method as above, we can show it holds for one-forms, and since both sides obey the Leibniz rule it holds for all tensors.

Now we see how the Lie derivative interacts with our other operations.

- For a torsion-free connection, we may replace all partial derivatives in the equations above with covariant derivatives, as all the extra terms produced cancel. This gives a totally coordinate-free definition of the Lie derivative.

- The Lie derivative of the metric is

$$(\mathcal{L}_X g)_{\mu\nu} = X^\rho \nabla_\rho g_{\mu\nu} + g_{\mu\rho}\nabla_\nu X^\rho + g_{\rho\nu}\nabla_\mu X^\rho = \nabla_\mu X_\nu + \nabla_\nu X_\mu.$$

  If ϕ_t is a one-parameter group of isometries of g, then $\mathcal{L}_X g = 0$ and we say X is a Killing vector field. Then Killing vectors obey the Killing equation

$$\boxed{\nabla_a X_b + \nabla_b X_a = 0.}$$

  Given a Killing vector X, we may choose coordinates so that $X = \partial/\partial t$ as above, which implies the metric components are independent of t.

- Let X^a be a Killing vector and let V^a be the tangent vector for an affinely-parametrized geodesic. Then

$$\frac{d}{d\tau}(X_a V^a) = V(X_a V^a) = \nabla_V (X_a V^a) = V^a V^b \nabla_b X_a + X_a V^b \nabla_b V^a.$$

  The first term vanishes by Killing's equation and the second vanishes by the geodesic equation. Therefore, Killing vectors yield conserved quantities.

- Finally, given a symmetric conserved energy-momentum tensor T_{ab} and a Killing vector X_a, we have a conserved current

$$J^a = T^a{}_b X^b, \quad \nabla_a J^a = 0.$$

  That is, given a Killing vector, covariantly conserved tensors yield conserved currents. Furthermore, given a conserved current and a timelike Killing vector, we can identify a conserved charge, i.e. a quantity that is constant on a family of spacelike hypersurfaces.

- It can also be shown that

$$\nabla_\mu \nabla_\sigma K^\rho = R^\rho{}_{\sigma\mu\nu} K^\nu.$$

  and contracting this identity gives

$$\nabla_\mu \nabla_\sigma K^\mu = R_{\sigma\nu} K^\nu.$$

  Applying the Bianchi identity and Killing's equation, we have

$$K^\lambda \nabla_\lambda R = 0.$$

  That is, the Ricci scalar does not change along a Killing vector field. This is another reflection of the fact that geometry is invariant along a Killing vector field.

4.3 Maximally Symmetric Spaces

Next, we investigate spacetimes that are as symmetric as possible.

- Clearly, $\mathbb{R}^n$ is maximally symmetric. It has n translational symmetries and $n(n-1)/2$ rotational symmetries, for a total of $n(n+1)/2$ independent Killing vectors.

- Since one can check if a vector field is Killing at a point using only local information, this should remain the maximum possible number of Killing vectors for curved spacetimes. We say a spacetime is maximally symmetric if it has this number.

- For example, Minkowski space is maximally symmetric, though some of the rotations become boosts. The sphere S^n is also maximally symmetric; embedding it in $\mathbb{R}^{n+1}$, the symmetries are the $n(n+1)/2$ rotations in $n+1$ dimensions. Note that for a fixed point p on the sphere, all but n of these rotations fix p, so these n rotations of $\mathbb{R}^{n+1}$ act like translations locally.
- Intuitively, the curvature in a maximally symmetric spacetime must be homogeneous and isotropic. This implies that, if we go to locally inertial coordinates, the Riemann tensor must be invariant under Lorentz transformations, and hence must be built out of the Minkowski metric and the Levi–Civita tensor.
- The only possibility that has the appropriate symmetries is
$$R_{\mu\nu\rho\sigma} \propto g_{\mu\rho}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\rho}$$
since $g_{\mu\nu} = \eta_{\mu\nu}$ in this frame. Since the expression is tensorial,
$$R_{abcd} = \frac{R}{n(n-1)}(g_{ac}g_{bd} - g_{ad}g_{bc}).$$
Conversely, if the Riemann tensor has this form, the space is maximally symmetric.
- We see a maximally symmetric space is specified by its dimension, signature, R, and possibly discrete global topological information.

Example. Consider two-dimensional maximally symmetric spaces in Euclidean signature; since the Riemann tensor has only one component the only condition is that the Ricci scalar be constant. For positive and zero R, we get S^2 and $\mathbb{R}^2$ respectively.

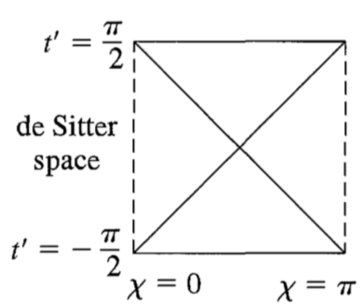
For negative R, we get a hyperboloid, which cannot be embedded isometrically into $\mathbb{R}^3$. It can be embedded as a hyperboloid in three-dimensional Minkowski space, with signature $(-++)$. Alternatively, it can be represented by the Poincare half plane, the set of points (x, y) with $y > 0$ and metric
$$ds^2 = \frac{a^2}{y^2}(dx^2 + dy^2).$$
Direct computation shows that $R = -2/a^2$, and geodesics are semicircles centered on the x-axis. Also note that the space does not have a boundary, which would contradict homogeneity, as the x-axis is an infinite distance away.

Note. We've been a little hasty above, because there is a distinction between requiring local or global existence of $n(n+1)/2$ Killing vectors. We only need the former for our formula for the Riemann tensor to hold, but then there is a larger variety of maximally symmetric spaces, obtained by quotienting the sphere, plane, or hyperboloid by finite groups. For example, we can quotient $\mathbb{R}^2$ by $\mathbb{Z}^2$ to get the torus, but then the rotation Killing vector is not globally defined.

The quotienting is not completely arbitrary, as the Gauss–Bonnet theorem relates local geometry with global topology. It states that for a two-dimensional compact boundaryless orientable manifold,
$$\chi(M) = \frac{1}{4\pi}\int_M R\sqrt{|g|}\, d^n x$$
where χ is the Euler characteristic, which in two dimensions satisfies
$$\chi(M) = 2(1-g)$$
where g is the genus of the surface. For example, we can't quotient the sphere to get a torus.

We now consider maximally symmetric spacetimes, in Lorentzian signature.

- The maximally symmetric spacetime with $R = 0$ is simply Minkowski space, at least locally. The spacetimes with $R > 0$ and $R < 0$ are de Sitter space and anti-de Sitter (AdS) space respectively.
- We construct de Sitter space by embedding it as a hyperboloid in five-dimensional Minkowski space,
$$ds_5^2 = -du^2 + dx^2 + dy^2 + dz^2 + dw^2, \quad -u^2 + x^2 + y^2 + z^2 + w^2 = \alpha^2.$$
We define coordinates $\{t, \chi, \theta, \phi\}$ by an analogue of hyperspherical coordinates,
$$u = \alpha \sinh(t/\alpha), \quad w = \alpha \cosh(t/\alpha)\cos\chi, \quad x = \alpha\cosh(t/\alpha)\sin\chi\cos\theta,$$
$$y = \alpha\cosh(t/\alpha)\sin\chi\sin\theta\cos\phi, \quad z = \alpha\cosh(t/\alpha)\sin\chi\sin\theta\sin\phi.$$
Then the metric on the hyperboloid is
$$ds^2 = -dt^2 + \alpha^2\cosh^2(t/\alpha)\left(d\chi^2 + \sin^2\chi\, d\Omega^2\right), \quad d\Omega^2 \equiv d\theta^2 + \sin^2\theta d\phi^2.$$
Therefore, in these coordinates, de Sitter space describes a spatial sphere S^3 that initially shrinks, hitting a minimum size at $t = 0$, then expands again.
- One can check that these coordinates cover the entire manifold, since geodesics never terminate in finite affine parameter. The topology of de Sitter space is thus $\mathbb{R} \times S^3$.
- It is simple to find the conformal diagram; by substituting $\cosh(t/\alpha) = 1/\cos(t')$, we have
$$ds^2 = \frac{\alpha^2}{\cos^2(t')} d\bar{s}^2, \quad d\bar{s}^2 = -dt'^2 + d\chi^2 + \sin^2\chi d\Omega^2.$$
Hence de Sitter space is conformally related to the Einstein static universe with conformal factor $\alpha/\cos(t')$. Now, the range of the new time coordinate is $t \in (-\pi/2, \pi/2)$, while $\chi \in [0, \pi]$, as is true for all hyperspherical coordinates except for the final one, $\phi \in [0, 2\pi]$.
- Thus the conformal diagram of de Sitter space is a square, shown below.



Note that the left and right edge of the square do not represent spatial infinity; they are simply the North and South poles of the S^3, and lie a finite distance away in de Sitter space. All other points on the diagram are copies of S^2.

- AdS space can be constructed similarly. We let

$$ds_5^2 = -du^2 - dv^2 + dx^2 + dy^2 + dz^2, \quad -u^2 - v^2 + x^2 + y^2 + z^2 = -\alpha^2$$

  and again take the analogue of hyperspherical coordinates $\{t', \rho, \theta, \phi\}$ by

$$u = \alpha \sin t' \cosh \rho, \quad v = \alpha \cos t' \cosh \rho, \quad x = \alpha \sinh \rho \cos \theta,$$

$$y = \alpha \sinh \rho \sin \theta \cos \phi, \quad z = \alpha \sinh \rho \sin \theta \sin \phi.$$

  Then the metric on the hyperboloid is

$$ds^2 = \alpha^2 \left(-\cosh^2 \rho \, dt'^2 + d\rho^2 + \sinh^2 \rho \, d\Omega^2\right).$$

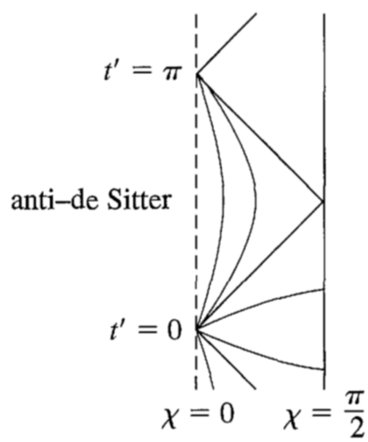
- Note that t' is a timelike coordinate, but it is also periodic, indicating the existence of closed timelike curves. To fix this, we recall that we've only constructed AdS space locally; we define the true AdS space by passing to the universal cover, allowing t' to range from $-\infty$ to ∞, so that our current embedding is merely a quotient.

- Another way to picture the quotiented AdS space in $1+1$ dimensions is as a hyperboloid, where the direction wrapping radially around the hyperboloid is the timelike direction.

- To get the conformal diagram, we define $\cosh \rho = 1/\cos \chi$ to find

$$ds^2 = \frac{\alpha^2}{\cos^2 \chi} d\bar{s}^2.$$

  The axes on the conformal diagram are $\chi \in [0, \pi/2)$ and $t' \in (-\infty, \infty)$, as shown.



  Since χ only goes up to $\pi/2$, AdS space is mapped onto only half of the Einstein static universe; the spatial slices are hemispheres. Spatial infinity at $\chi = \pi/2$ has the topology of S^2 and indicates the initial value problem is not well-defined, as information can come from spatial infinity.

Finally, we relate these spaces to cosmology.

- By isotropy, the Einstein tensor is proportional to $g_{\mu\nu}$, and the Einstein field equation is

$$T_{\mu\nu} = -\frac{3}{8\pi G}\frac{R}{n(n-1)}g_{\mu\nu}.$$

  This corresponds to an FRW spacetime dominated by vacuum energy.

- In particular, it can be shown that de Sitter space corresponds to an FRW universe with $k = 0$ and exponential scale factor; the apparent discrepancy in the behavior of the scale factor is because the FRW coordinates only cover half of de Sitter space. Locally, it is a good model for our universe during the era of inflation and the far future.

- By contrast, AdS space corresponds to an open universe with $R > 0$. It is unrealistic, but interesting for theoretical reasons.

Note. These spaces may also be understood as coset spaces. As review, let a Lie group G act transitively on a set $\mathcal{M}$. Then if H is the stabilizer of a point in $\mathcal{M}$, then $\mathcal{M} \cong G/H$ by the orbit-stabilizer theorem. Now let $\mathfrak{g} = \mathfrak{h} \oplus \mathfrak{k}$. If $[\mathfrak{k}, \mathfrak{k}] \subset \mathfrak{h}$, then $\mathcal{M}$ is a symmetric space.

This idea may also be used in reverse. For example, a person living on the surface of the Earth observes the symmetry group $SO(2)$. If the entire Earth is assumed to be symmetric, then $\mathcal{M} \cong G/SO(2)$ for some Lie group G, which must be three-dimensional to make $\mathcal{M}$ two-dimensional. The two possibilities $G = E(2)$ and $G = SO(3)$ lead to a flat Earth and a sphere respectively.

Similarly, we observe that space locally has symmetry group $SO(3)$, so assuming spatial homogeneity, $\mathcal{M} = G/SO(3)$ where G is six-dimensional. The three possibilities are $G = E(3)$, $G = SO(4)$, and $SO(3,1)$, which correspond to Euclidean, spherical, and hyperbolic space. Note that the appearance of the Lorentz group here has nothing to do with relativity.

Finally, in the cosmological context, space locally has symmetry group $SO(3,1)$. If we assume spacetime homogeneity, then we need to find a ten-dimensional group G. The three possibilities are

$$E(3,1), \quad SO(4,1), \quad SO(3,2)$$

where $E(3,1)$ is just the Poincare group. These correspond to Minkowski, de Sitter, and anti de Sitter spacetime, respectively. However, the ordinary cosmological principle only assumes homogeneity in space, leading to the more general FLRW spacetimes. Here we have additionally assumed homogeneity in time (the perfect cosmological principle), which is why the result is not realistic.

4.4 Differential Forms

We quickly review the basics of differential forms and set conventions (distinct from those in the notes on Geometry and Topology), before moving on to new structures built from differential forms that require a metric or connection.

- The wedge product of a p-form X and a q-form Y is the $(p+q)$-form

$$(X \wedge Y)_{a_1 \ldots a_p b_1 \ldots b_q} = \frac{(p+q)!}{p!\, q!} X_{[a_1 \ldots a_p} Y_{b_1 \ldots b_q]}.$$

  Then we have $X \wedge Y = (-1)^{pq} Y \wedge X$ and the wedge product is associative. While the coefficients here look a bit nasty, they ensure calculations in terms of wedge products are nice.

- Given a basis of one-forms $\{f^\mu\}$, a p-form X may be expanded in components as

$$X = \frac{1}{p!} X_{\mu_1 \ldots \mu_p} f^{\mu_1} \wedge \ldots \wedge f^{\mu_p}.$$

- The exterior derivative of a p-form X is

$$(dX)_{\mu_1 \ldots \mu_{p+1}} = (p+1) \partial_{[\mu_1} X_{\mu_2 \ldots \mu_{p+1}]}$$

where the components are taken in the dual basis of the coordinate basis $\{\partial_\mu\}$. This definition is tensorial; the antisymmetrization cancels the extra terms by the symmetry of mixed partials.

- In terms of wedge products, d is "$dx^\mu \wedge \partial_\mu$" where ∂_μ acts on coefficients, with no additional constants. For example, $d(y\,dx) = dy \wedge dx$.

- For a torsion-free connection, we may replace the partial derivatives by covariant derivatives,

$$(dX)_{\mu_1 \ldots \mu_{p+1}} = (p+1) \nabla_{[\mu_1} X_{\mu_2 \ldots \mu_{p+1}]}$$

where the extra terms cancel by antisymmetry. Alternatively, when the connection is torsion-free, the covariant derivative reduces to the partial derivative in normal coordinates. Then this equation holds in normal coordinates, so it holds in all coordinates.

- The exterior derivative obeys the properties

$$d(dX) = 0, \quad d(X \wedge Y) = (dX) \wedge Y + (-1)^p X \wedge dY, \quad d(\phi^* X) = \phi^*(dX)$$

for a p-form X, where the first is by the symmetry of mixed partials and the second is by anticommuting the derivative through p indices. The last fact can be proven by induction.

- The results above imply

$$\mathcal{L}_V(dX) = d(\mathcal{L}_V X).$$

Also note that since $\mathcal{L}$ commutes with tensor products, and the wedge product is just a tensor product, $\mathcal{L}$ satisfies the Leibniz rule

$$\mathcal{L}_V(X \wedge Y) = \mathcal{L}_V X \wedge Y + X \wedge \mathcal{L}_V Y.$$

- We say a form X is closed if $dX = 0$, and exact if $X = dY$. All exact forms are closed, and the Poincare lemma shows all closed forms are locally exact.

- We define $i_V X$ to be the result of contracting V with the first index of X,

$$(i_V X)_{\mu_2 \ldots \mu_p} = V^{\mu_1} X_{\mu_1 \ldots \mu_p}$$

which implies that, like the exterior derivative, it satisfies a graded Leibniz rule

$$i_V(X \wedge Y) = (i_V X) \wedge Y + (-1)^p X \wedge i_V Y.$$

This can be used to compute in terms of wedge products. For example, letting $V = \partial_y$,

$$i_V(dx \wedge dy) = i_V(dx) \wedge dy - dx \wedge i_V(dy) = V(dx)dy - V(dy)dx = -dx.$$

- Cartan's formula states that

$$\mathcal{L}_V X = (d i_V + i_V d) X.$$

It can be proven by induction, by noting that both sides are linear and obey the ungraded Leibniz rule.

4.5 Tetrads

Next, we introduce tetrads, a useful computational tool.

- An orthonormal basis of vector fields $\{e_\mu^a\}$, also called a tetrad or vielbein, obeys

$$g_{ab} e_\mu^a e_\nu^b = \eta_{\mu\nu}.$$

  There are two types of indices here: the Latin indices are abstract indices which are raised and lowered by the metric g_{ab}, while the Greek indices merely label which vector field we are talking about. Note that the vielbein need not form a coordinate basis; it may vary from point to point in any way.

- The dual basis $\{e_a^\mu\}$ is the set of one-forms satisfying

$$e_a^\mu e_\nu^a = \delta_\nu^\mu.$$

  This equation is satisfied if we define

$$e_a^\mu = \eta^{\mu\nu}(e_\nu)_a.$$

  Hence Greek indices are raised and lowered by the Minkowski metric.

- The dual basis vectors satisfy

$$\eta_{\mu\nu} e_a^\mu e_b^\nu = g_{ab}, \quad e_a^\mu e_\mu^b = \delta_a^b.$$

  To show the first result, contract both sides by e_ρ^b. The second follows as a corollary of the first. The upshot is that the two types of indices behave exactly as we would naively expect.

- We can write tensor components in the coordinate basis in the tetrad basis by raising and lowering with the vielbein, e.g.

$$V^a = e_\mu^a V^\mu, \quad V = V^a \partial_a = V^\mu e_\mu$$

  Our first identity above is just a special case of this for the metric. We can also have tensors in mixed Latin and Greek indices.

- Note that it may be impossible to define a tetrad globally, given some desired conditions. Thus, just like for coordinates, we may need multiple tetrads each defined in a patch of the manifold.

- In special relativity, we work with only orthonormal bases, and transfer between them by Lorentz transformations. In the tetrad formalism, we have an orthonormal basis at every point, and different tetrads are related by position-dependent Lorentz transformations. Hence Lorentz symmetry is made into a local symmetry. Note that this is completely independent of coordinate transformations, which act on the Latin indices.

(finish)

4.6 Integration

First, we define orientability and the volume form.

- An n-dimensional manifold is orientable if it admits an orientation, a nowhere-vanishing n-form $\epsilon_{a_1 \ldots a_n}$. Two orientations ϵ and ϵ' are equivalent if $\epsilon' = f\epsilon$ where f is positive.

- A coordinate chart x^μ on an oriented manifold is right-handed with respect to ϵ if $\epsilon = f(x) dx^1 \wedge \ldots \wedge dx^n$ with $f(x) > 0$. Equivalently, it is right-handed if $\epsilon(\partial_1, \ldots, \partial_n) > 0$.

- On an oriented manifold with a metric, the volume form or Levi–Civita tensor is defined by

$$\epsilon_{12\ldots n} = \sqrt{|g|}$$

  in any right-handed coordinate chart. This is a coordinate-independent definition, because $\sqrt{|g|}$ picks up a factor of the Jacobian on a coordinate transformation.

- In a right-handed coordinate chart, we have

$$\epsilon^{12\ldots n} = \pm \frac{1}{\sqrt{|g|}}$$

  where the lower sign applies in Lorentzian signature. To show this, note that $\epsilon_{a_1 \ldots a_n} \epsilon^{a_1 \ldots a_n} = \pm 1$ in an orthonormal basis, and this result is independent of basis.

- For the Levi–Civita connection, we have

$$\nabla_a \epsilon_{b_1 \ldots b_n} = 0.$$

  To show this, work in normal coordinates in p. Then the covariant derivative becomes the partial derivative, and the partial derivative vanishes because $\partial g_{\mu\nu} = 0$ so $\partial g = 0$.

- We also have the identity

$$\epsilon^{a_1 \ldots a_p c_{p+1} \ldots c_n} \epsilon_{b_1 \ldots b_p c_{p+1} \ldots c_n} = \pm p!(n-p)! \delta^{a_1}_{[b_1} \ldots \delta^{a_p}_{b_p]}.$$

  To prove this identity, note that the left-hand side is nonzero when the a_i and b_i are all distinct, and permutations of each other. The antisymmetrized delta function accounts for this. The factor of $(n-p)!$ is the number of c_i values that contribute.

- We define the Hodge dual of a p-form X as

$$(\star X)_{a_1 \ldots a_{n-p}} = \frac{1}{p!} \epsilon_{a_1 \ldots a_{n-p} b_1 \ldots b_p} X^{b_1 \ldots b_p}.$$

  Directly using the above identity, we have

$$\star(\star X) = \pm(-1)^{p(n-p)} X, \quad (\star d \star X)_{a_1 \ldots a_{p-1}} = \pm(-1)^{p(n-p)} \nabla^b X_{a_1 \ldots a_{p-1} b}.$$

  That is, preceding d with a Hodge star turns off the antisymmetrization.

- Another useful result is that for two p-forms X and Y, we have

$$(X \wedge \star Y)_{a_1 \ldots a_{2p}} = \frac{(-1)^{p(n-p)}}{p!} \epsilon_{a_1 \ldots a_{2p}} X_{b_1 \ldots b_p} Y^{b_1 \ldots b_p}.$$

  A useful corollary is that $X \wedge \star Y = Y \wedge \star X$.

Next, we define integration on a manifold.

- Let M be an oriented manifold of dimension n, let $\psi\colon \mathcal{O} \to \mathcal{U}$ be a right-handed coordinate chart with coordinate x^μ, and let X be an n-form. Then the integral of X over $\mathcal{O}$ is

$$\int_{\mathcal{O}} X = \int_{\mathcal{U}} dx^1 \dots dx^n \, X_{1\dots n}.$$

  This definition is chart-independent, as the two terms pick up canceling Jacobian factors.

- More generally, we can extend this definition to all of M by breaking it into coordinate patches. Specifically, for charts $\phi_\alpha\colon \mathcal{O}_\alpha \to \mathcal{U}_\alpha$, define a partition of unity to be a set of functions $h_\alpha\colon M \to [0,1]$ so that

$$h_\alpha(p) = 0 \text{ if } p \notin \mathcal{O}_\alpha, \quad \sum_\alpha h_\alpha(p) = 1.$$

  Then we define

$$\int_M X = \sum_\alpha \int_{\mathcal{O}_\alpha} h_\alpha X.$$

  It can be shown this definition is independent of the partition of unity.

- A diffeomorphism $\phi\colon M \to M$ is orientation preserving if $\phi^*(\epsilon)$ is equivalent to ϵ for any orientation ϵ. Then the integral is preserved,

$$\int_M \phi^*(X) = \int_M X.$$

  This is easiest to see in the passive view of diffeomorphisms as coordinate transformations, in which case it's just an extension of chart-independence.

- If M has a metric g and hence a metric volume form ϵ, we define the volume of M as

$$V = \int_M \epsilon.$$

  Moreover, we can define the integral of a function on M by

$$\int_M f \equiv \int_M f\epsilon = \int_M d^n x \, \sqrt{|g|} f.$$

  The latter is the physicist's preferred notation, since it is more explicit, but is technically incorrect since the manifold generally can't be covered by a single chart.

- Delta functions must also be normalized to be tensorial. In coordinates, we have

$$\int_{\mathcal{U}} dx^1 \dots dx^n \, f(x)\delta(x) = f(0)$$

  and the right-hand side is a scalar. To write the left-hand side in terms of tensorial quantities, the function must be multiplied by $\sqrt{|g|}$, so the delta function must be divided by $\sqrt{|g|}$.

Next, we define submanifolds and state Stokes' theorem.

- Let S and M be oriented manifolds of dimensions $m < n$. A smooth map $\phi\colon S \to M$ is an embedding if it is injective, and for any $p \in S$ there is a neighborhood $\mathcal{O}$ so that $\phi^{-1}\colon \phi[\mathcal{O}] \to S$ is smooth. Then we say $\phi[S]$ is an embedded submanifold of M.
- We will refer to embedded submanifolds as simply submanifolds. Then a hypersurface is a submanifold of dimension $n-1$.
- If $\phi[S]$ is an m-dimensional submanifold of M and X is an m-form on M, then the integral of X over $\phi[S]$ is defined as
$$\int_{\phi[S]} X = \int_S \phi^*(X).$$
- A manifold with boundary M is like a manifold, but the charts are maps $\phi_\alpha\colon \mathcal{O}_\alpha \to \mathcal{U}_\alpha$ where
$$\mathcal{U}_\alpha \text{ is an open set of } \{(x_1, \ldots, x^n) \in \mathbb{R}^n \mid x_1 \leq 0\}.$$
The boundary ∂M is the set of points for which $x^1 = 0$. Then ∂M is a submanifold of M. If M is oriented, the orientation of ∂M is fixed by saying that $(x^2, \ldots, x^n)$ is a right-handed chart on ∂M when $(x^1, \ldots, x^n)$ is a right-handed chart on M.
- Stokes' theorem states that
$$\int_N dX = \int_{\partial N} X$$
where ∂N is regarded as a submanifold of N. Most commonly, this theorem is used when N is a region of a larger manifold M, so that ∂N is a hypersurface in M.
- If we are dealing with fields on spacetime that fall off at infinity, then the right-hand side in Stokes' theorem vanishes. In particular, if X can be written in terms of wedge products, this gives an integration by parts formula for differential forms, which has extra signs.

Finally, we specialize to hypersurfaces.

- We say $X \in T_p(M)$ is tangent to $\phi[S]$ at p if X is the tangent vector at p of a curve that lies in $\phi[S]$. We say a one-form n is normal to $\phi[S]$ if $n(X) = 0$ for any vector X tangent to $\phi[S]$. Note that this definition does not require a metric.
- In the case of a hypersurface, the tangent space has dimension $n-1$, so the vector space of normals has dimension 1.
- If the manifold has a metric, we can calculate the norm of the normal one-forms using the inverse metric. We say a hypersurface on a Lorentzian manifold is
$$\begin{cases} \text{timelike} \\ \text{spacelike} \\ \text{null} \end{cases} \text{if the normal vectors are everywhere} \begin{cases} \text{spacelike} \\ \text{timelike} \\ \text{null.} \end{cases}$$
- Explicitly, if M is a manifold with boundary, then its boundary ∂M has the normal one-form dx^1 by definition. We can normalize the normals by
$$n_a = \frac{(dx^1)_a}{\sqrt{\pm g^{bc}(dx^1)_b (dx^1)_c}}, \quad g^{ab} n_a n_b = \pm 1$$
whenever n_a is not null.

- Note that the above definition of n_a is still ambiguous up to an overall sign. For a spacelike hypersurface, we choose the sign so that n^a points into M. However, we need n^a to point out of M when ∂M is timelike. (This can be done smoothly, because at the crossover, when ∂M is null, n^a is tangent to ∂M.)

- With the above sign prescription, we have the divergence theorem,

$$\boxed{\int_M d^n x \sqrt{|g|}\, \nabla_a X^a = \int_{\partial M} d^{n-1} x \sqrt{|h|}\, n_a X^a}$$

  where X^a is a vector field on M, ∇ is the Levi–Civita connection, and h_{ab} is the pullback of the metric g_{ab} to ∂M by inclusion. The integrand $n_a X^a$ makes sense because it is a scalar on M and hence can be pulled back to ∂M.

- In particular, a covariantly conserved vector $\nabla_a X^a = 0$ yields a conserved quantity. If M is bounded by hypersurfaces Σ and Σ', then

$$0 = \int_{\partial M} d^{n-1} x \, \sqrt{|h|} n_a X^a = \int_\Sigma d^{n-1} x \, \sqrt{|h|} n_a X^a - \int_{\Sigma'} d^{n-1} x \, \sqrt{|h|} n_a X^a.$$

  In the context of classical field theory, this means conserved charges can be equivalently computed using any hypersurface with the same boundary. For instance, one special case of this is the Lorentz invariance of electric charge.

Note. All of the above results are applied in the theory of electromagnetism. The action is

$$S = \int d^4 x \sqrt{-g} \left(-\frac{1}{4} F^{ab} F_{ab} + J^a A_a \right)$$

which, in terms of differential forms, is

$$S = \int -\frac{1}{2} F \wedge \star F + A \wedge \star J$$

where $F = dA$. Gauge transformations take the form $A \to A + d\alpha$, where F is gauge invariant because $d^2 = 0$. The action is gauge invariant provided that

$$\int d\alpha \wedge \star J = \int \alpha \wedge (d \star J) = 0$$

where we "integrated by parts", which means we require the current be conserved,

$$\nabla_a J^a = 0, \quad d \star J = 0.$$

The equation of motion is

$$\nabla^a F_{ab} = -J_b, \quad d \star F = \star J.$$

Applying d to both sides recovers current conservation. The other two Maxwell's equations form the Bianchi identity,

$$\nabla_{[a} F_{bc]} = 0, \quad dF = 0$$

which follow automatically from $F = dA$. To translate these equations to differential form, let Σ be a hypersurface in a spacetime M. Then we have

$$\int_{\partial \Sigma} \star F = \int_\Sigma d \star F = \int_\Sigma \star J = Q.$$

This is a generalization of Gauss's law. The left-hand side measures the flux through $\partial\Sigma$, and the right-hand side measures the charge inside Σ. One can also define the magnetic flux through a closed surface, but this vanishes automatically,

$$Q_m = \int_{\partial\Sigma} F = \int_\Sigma dF = 0.$$

However, this can be evaded in topologically more interesting spaces.

Note. A tiny amount of Hodge theory. In Euclidean space, let X be a one-form with corresponding vector field $\mathbf{X}$. Then the familiar vector calculus operations are

$$\nabla f = df, \quad \nabla \cdot \mathbf{X} = \star\, d \star X, \quad \nabla \times \mathbf{X} = \star\, dX. \quad \nabla^2 f = \star\, d \star df.$$

Moving away from vectors and focusing on forms, this tells us that d is an operation that generalizes the gradient and curl, while $\star d \star$ generalizes the divergence.

We are therefore motivated to define the adjoint operator to d,

$$d^\dagger = \pm(-1)^{np+n-1} \star d \star .$$

This is the adjoint in the sense that if we define an inner product on p-forms,

$$\langle \eta, \omega \rangle = \int_M \eta \wedge \star\omega$$

then "integration by parts" tells us that

$$\langle d\alpha, \omega \rangle = \langle \alpha, d^\dagger \omega \rangle,$$

where all the complicated sign factors are engineered to cancel out. We can thus define a Laplacian on p-forms,

$$\triangle = (d + d^\dagger)^2 = dd^\dagger + d^\dagger d$$

where we used $d^2 = {d^\dagger}^2 = 0$. Acting on functions (0-forms), this matches with our above expression for $\nabla^2 f$. In analogy with harmonic functions, we say a harmonic form $\gamma \in \mathrm{Harm}^p(M)$ obeys

$$\triangle\gamma = 0.$$

Since we have

$$\langle \gamma, \triangle\gamma \rangle = \langle d\gamma, d\gamma \rangle + \langle d^\dagger\gamma, d^\dagger\gamma \rangle = 0$$

and the inner product is positive definite, a harmonic form is both closed and co-closed, $d\gamma = 0$ and $d^\dagger\gamma = 0$. (Technically, this is stricter than the usual vector calculus notion of a harmonic function, as this would imply $df = 0$ and hence f must be constant on $\mathbb{R}^3$. The issue is that we integrated by parts above, which requires functions to fall off at infinity. Indeed, there are no harmonic functions on $\mathbb{R}^3$ that fall off at infinity.)

The Hodge decomposition theorem states that any form can be uniquely decomposed as

$$\omega = d\alpha + d^\dagger\beta + \gamma$$

which generalizes Helmholtz's theorem, which states that on $\mathbb{R}^3$, a vector field that falls off at infinity can be written in terms of a divergence-free plus a curl-free part, or equivalently as a curl plus a gradient. The Hodge decomposition theorem also implies that harmonic forms (which are closed) cannot be written in the form $d\alpha$ (so they are not exact), which gives Hodge's theorem,

$$\mathrm{Harm}^p(M) \equiv H^p(M)$$

where $H^p(M)$ is a de Rham cohomology group.

4.7 Lagrangian Formulation

We now introduce Lagrangian mechanics on a Lorentzian manifold.

- For a minimally coupled real scalar field $\Phi\colon M \to \mathbb{R}$, the action is

$$S[\Phi] = \int_M d^4x\,\sqrt{-g}\mathcal{L}, \quad \mathcal{L} = -\frac{1}{2}g^{ab}\nabla_a\Phi\nabla_b\Phi - V(\Phi).$$

  In general, we allow the Lagrangian to depend on the fields and their first covariant derivatives; it must be a scalar. Note that one could also introduce non-minimal couplings such as $R\phi^2$.

- The derivation of the Euler-Lagrange equations goes through as before. The divergence theorem is used to integrate by parts, and as usual we throw away surface terms, which live at spatial infinity.

- In general, the Euler-Lagrange equations read

$$\frac{\partial\mathcal{L}}{\partial\Phi} - \nabla_\mu\left(\frac{\partial\mathcal{L}}{\partial(\nabla_\mu\Phi)}\right) = 0$$

  and in the case of the real scalar field we have

$$\nabla^a\nabla_a\Phi - V'(\Phi) = 0$$

  as guessed before by minimal coupling.

- Alternatively, in differential form notation, the kinetic term can be written as

$$\int d\Phi \wedge \star d\Phi$$

  which leads to the differential form version of the wave operator

$$\star d \star d\phi = \frac{1}{\sqrt{-g}}\partial_a(\sqrt{-g}g^{ab}\partial_b\Phi) = \nabla_a\nabla^a\Phi$$

  as shown earlier.

- The Einstein–Hilbert action for general relativity is

$$\boxed{S[g] = \int_M d^4x\,\sqrt{-g}R = \int_M R\epsilon.}$$

  This is a reasonable guess, since R is the simplest scalar we know constructed from the Riemann tensor, and it turns out it is unique given some mild assumptions.

- We cannot use the Euler-Lagrange equations because R cannot be written in terms of the metric and its covariant derivative, which vanishes. Instead, we must vary the integral directly.

We now vary the Einstein–Hilbert action.

- Writing the integrand as $\sqrt{-g}R_{ab}g^{ab}$, there are three separate terms. By the identity

$$\delta g^{ab} = -g^{ac}g^{bd}\delta g_{cd}$$

  which is familiar from linearized theory, the third term is $-\sqrt{-g}R^{ab}\delta g_{ab}$.

- We would like to write the first two terms as something multiplied by δg_{ab}. We begin with the metric determinant. Varying the identity

$$\log\det M = \operatorname{tr}\log M$$

by expanding $\log M$ in a Taylor series and using cyclic permutation under a trace, we have

$$\frac{\delta(\det M)}{\det M} = \operatorname{tr}(M^{-1}\delta M)$$

which implies that

$$\delta\sqrt{-g} = \frac{1}{2}\sqrt{-g}\, g^{ab}\delta g_{ab}.$$

This is promising, as this gives a term like $g^{ab}R\delta g_{ab}$ in the variation, which combines with the third term to yield the Einstein tensor.

- Finally, we need to compute the variation of the Ricci tensor. We begin by varying the Christoffel symbols. Since the difference of two connections is a tensor, $\delta\Gamma^{\mu}_{\nu\rho}$ are the components of a tensor $\delta\Gamma^{a}_{bc}$ and hence can be computed in normal coordinates. At first order,

$$\delta\Gamma^{\mu}_{\nu\rho} = \frac{1}{2}g^{\mu\sigma}(\delta g_{\sigma\nu,\rho} + \delta g_{\sigma\rho,\nu} - \delta g_{\nu\rho,\sigma}) = \frac{1}{2}g^{\mu\sigma}(\delta g_{\sigma\nu;\rho} + \delta g_{\sigma\rho;\nu} - \delta g_{\nu\rho;\sigma}).$$

This is a tensorial expression and hence holds in all coordinate systems.

- Again working in normal components, and to first order in $\delta\Gamma$,

$$\delta R^{\mu}{}_{\nu\rho\sigma} = \partial_\rho\delta\Gamma^{\mu}_{\nu\sigma} - \partial_\sigma\delta\Gamma^{\mu}_{\nu\rho} = \nabla_\rho\delta\Gamma^{\mu}_{\nu\sigma} - \nabla_\sigma\delta\Gamma^{\mu}_{\nu\rho}$$

which again is a tensorial expression that holds in all coordinate systems. Contracting indices, we see δR_{ab} is a total covariant derivative, so the variation of the action reduces to a surface term involving $\delta\Gamma^{a}_{bc}$. Since we assume δg_{ab} has compact support, $\delta\Gamma^{a}_{bc}$ does as well, so the contribution is zero.

- Combining our above results gives $G^{ab} = 0$, the vacuum Einstein equation. We can also reach this conclusion by the Palatini procedure, where we vary the metric and the connection independently, only requiring the connection be torsion-free. Then the variation of the metric gives $G^{ab} = 0$ for an arbitrary connection, while the variation of the connection forces it to be the Levi–Civita connection.

- The metric is symmetric, so δg_{ab} must also be symmetric; its components are not all independent. The price of treating δg_{ab} as arbitrary is that the change of the metric must actually be $\delta g_{(ab)}$. This leads to some surprising conclusions, as

$$\frac{\delta g_{\mu\nu}}{\delta g_{\alpha\beta}} = \frac{1}{2}(\delta^{\alpha}_{\mu}\delta^{\beta}_{\nu} + \delta^{\alpha}_{\nu}\delta^{\beta}_{\mu}), \quad \frac{\delta g_{12}}{\delta g_{12}} = \frac{1}{2}.$$

In practice, we can ignore this issue, and account for it by always symmetrizing in a and b when we vary with respect to g_{ab}.

Next, we add in matter fields.

- The action for general relativity with a cosmological constant and matter fields is

$$S = \frac{1}{16\pi}\int_M d^4x\sqrt{-g}\,(R-2\Lambda) + \int_M d^4x\sqrt{-g}\,\mathcal{L}_{\text{matter}}.$$

To recover the Einstein field equation, we define the energy-momentum tensor to be

$$T^{ab} = \frac{2}{\sqrt{-g}}\frac{\delta S_{\text{matter}}}{\delta g_{ab}}$$

which is automatically symmetric, as discussed above.

- What the above notation means is that under a variation of g_{ab},

$$\delta S_{\text{matter}} = \frac{1}{2}\int_M d^4x\,\sqrt{-g}T^{ab}\delta g_{ab}.$$

That is, as usual, the functional derivative above refers to the total variation (which includes changes due to $\delta(\partial g)$), and we may have to integrate by parts to write it in terms of δg alone.

- In general, applying Noether's theorem to spacetime translations gives a different energy-momentum tensor, which may not even be symmetric, and is ambiguous up to the addition of total derivatives to the Lagrangian. The energy-momentum tensor we have defined above is automatically symmetric and is not affected by total derivative terms, since it depends directly on the action. Thus our definition above can be viewed as a statement of which Noether energy-momentum tensor is physical in general relativity.

- Applying this definition to the real scalar field, we have

$$\delta S_{\text{matter}} = \int_M d^4x\,\sqrt{-g}\left(\frac{1}{2}\nabla^a\Phi\nabla^b\Phi + \frac{1}{2}\left(-\frac{1}{2}g^{cd}\nabla_c\Phi\nabla_d\Phi - V(\Phi)\right)g^{ab}\right)\delta g_{ab}$$

where there is a sign flip from the inverse metric variation. Then we read off

$$T^{ab} = \nabla^a\Phi\nabla^b\Phi - \left(\frac{1}{2}g^{cd}\nabla_c\Phi\nabla_d\Phi + V(\Phi)\right)g^{ab}.$$

4.8 Diffeomorphism and Conformal Invariance

Below we will explore some of the consequences of diffeomorphism invariance.

- The total action should be diffeomorphism invariant, $S[g,\Phi] = S[\phi_*(g),\phi_*(\Phi)]$. Moreover, the Einstein–Hilbert action is diffeomorphism invariant by itself, so the matter action S_{m} must also be diffeomorphism invariant. This is true by construction, since we have taken it to be the integral of a scalar Lagrangian.

- Next, consider the effect of an infinitesimal diffeomorphism. We have

$$\delta g_{ab} = \mathcal{L}_\xi g_{ab} = \nabla_a\xi_b + \nabla_b\xi_a, \quad \delta\Phi = \mathcal{L}_\xi\Phi$$

where, in the case of a scalar field, $\mathcal{L}_\xi\Phi = \xi^a\nabla_a\Phi$. Now suppose S_{m} is the integral of a scalar Lagrangian constructed from g, Φ, and arbitrarily many of their derivatives. Then

$$\delta S_{\text{m}} = \int_M d^4x\left(\frac{\delta S_{\text{m}}}{\delta\Phi}\delta\Phi + \frac{\delta S_{\text{m}}}{\delta g_{ab}}\delta g_{ab}\right) = 0.$$

The first term vanishes on-shell; note that there is no factor of $\sqrt{-g}$ here.

- To handle the second term, we expand the definitions to find

$$\frac{\delta S_{\mathrm{m}}}{\delta g_{ab}} \delta g_{ab} = \frac{\sqrt{-g}}{2} T^{ab} (\nabla_a \xi_b + \nabla_b \xi_a).$$

  The two terms in parentheses are equal by the symmetry of T^{ab}. Moreover, the appearance of the $\sqrt{-g}$ allows us to integrate by parts, moving the covariant derivative onto T^{ab}. Since ξ_a is arbitrary, we conclude $\nabla_a T^{ab} = 0$. An identical procedure applied to the Einstein–Hilbert action gives the contracted Bianchi identity $\nabla_a G^{ab} = 0$.

Now consider conformal transformations, as introduced in the notes on String Theory.

- Conformal transformations are equivalent to a special kind of diffeomorphism, composed with a Weyl transformation. The variation of the matter action under a conformal transformation is

$$\delta S_{\mathrm{m}} = \int_M d^4x \left(\frac{\delta S_{\mathrm{m}}}{\delta \Phi} (\delta^d \Phi + \delta^w \Phi) + \frac{\delta S_{\mathrm{m}}}{\delta g_{ab}} (\delta^d g_{ab} + \delta^w g_{ab}) \right).$$

  We treat these four terms very explicitly, as many books do not display them all.

- We name the terms δS_1 through δS_4. Then the following set of facts hold.
  - By the definition of a Weyl transformation, $\delta^w g_{ab} = \omega(x) g_{ab}$.
  - We always have $\delta S_1 + \delta S_3 = 0$ by diffeomorphism invariance, which should hold in any reasonable theory.
  - If the theory is conformally invariant, then $\delta S_{\mathrm{m}} = 0$ off shell.
  - If the theory is Weyl invariant, then we have $\delta S_2 + \delta S_4 = 0$ off shell. Weyl invariance implies conformal invariance but not vice versa, because $\omega(x)$ is arbitrary for a Weyl transformation but not for a conformal transformation. However, the two usually coincide.
  - If the matter is on shell, then $\delta S_1 = \delta S_2 = 0$.
  - If all matter fields have vanishing conformal dimension, then $\delta S_2 = 0$.
  - For a Weyl invariant or conformally invariant theory, combining these results shows $\delta S_4 = 0$ on shell. For a Weyl invariant theory, this implies the trace of the stress-energy tensor vanishes on shell.
  - Note that the composite of the two transformations yields the final metric $g'_{\mu\nu}(x') = g_{\mu\nu}(x)$. Hence if the metric is Minkowski, it is unchanged, and $\delta S_3 + \delta S_4 = 0$.

Note. A more direct motivation of the definition of T_{ab}. Consider a field theory on a fixed, flat spacetime background with invariance under translations ϵ^a. We may quickly derive the associated conserved quantity by the usual "Noether trick". We promote ϵ^a so that it has spacetime dependence. Using the notation introduced above, the variation of the action has the form

$$\delta S_{\mathrm{m}}^{\mathrm{fixed}} = \delta S_1 \propto \int d^4x \, T^{ab} \partial_a \epsilon_b.$$

Integrating by parts and using the fact that $\epsilon(x)$ is arbitrary shows that $\partial_a T^{ab} = 0$ on shell, so it is the conserved tensor associated with translational invariance.

In this derivation, the metric is treated as a fixed background. Now we consider coupling the theory to a dynamical metric. The theory is diffeomorphism invariant, so as above,

$$\delta S_{\text{m}}^{\text{dynamical}} = \delta S_1 + \delta S_3 = 0.$$

But this implies that

$$\int d^4x\, T^{ab}\partial_a \epsilon_b \propto \int d^4x\, \frac{\delta S_{\text{m}}}{\delta g_{ab}} \delta g_{ab} \propto \int d^4x\, \frac{\delta S_{\text{m}}}{\delta g_{ab}} \partial_a \epsilon_b.$$

Hence Noether's theorem directly tells us to take $T^{ab} \propto \delta S_{\text{m}}/\delta g_{ab}$.

We now turn to a closer investigation of stress-energy tensors and their special properties. For simplicity, we restrict to flat spacetime described by the Minkowski metric.

- We let $T_C^{\mu\nu}$ be the canonical stress-energy tensor, derived by translational symmetry by Noether's theorem. Then under an infinitesimal diffeomorphism,

$$\int_M d^4x\, \frac{\delta S_{\text{m}}}{\delta \Phi} \delta\Phi = \int_M d^4x\, (\partial_\mu T_C^{\mu\nu})\xi_\nu.$$

  Noting that we always have $\delta S_{\text{m}} = 0$, this implies that

$$\int_M d^4x\, \partial_\mu (T_C^{\mu\nu} - T^{\mu\nu})\xi_\nu = 0$$

  and since ξ is arbitrary, the two definitions of the stress-energy tensor have the same divergence; hence one is conserved if and only the other is.

- However, the tensor $T_C^{\mu\nu}$ may not be symmetric. It turns out that it is possible in theories with Lorentz symmetry (or rotational symmetry in Euclidean signature) to modify it to a symmetric tensor $T_B^{\mu\nu}$ called the Belinfante tensor, where $\partial_\mu T_B^{\mu\nu} = \partial_\mu T_C^{\mu\nu}$ always.

- Next, consider a theory with dilation symmetry. Taking a single scalar field for simplicity, the corresponding conserved current is

$$j_D^\mu = T_C^{\mu\nu} x^\nu + \frac{\partial \mathcal{L}}{\partial(\partial_\mu \Phi)} \Delta\Phi$$

  where Δ is the scaling dimension of Φ. In this case, one can show it is possible in *some* theories with $d > 2$ to define yet another tensor so that

$$j_D^\mu = T_D^{\mu\nu} x^\nu, \quad \eta_{\mu\nu} T_D^{\mu\nu} = \partial_\mu j_D^\mu$$

  so that $T_D^{\mu\nu}$ has the same divergence as the other stress-energy tensors, is symmetric like $T_B^{\mu\nu}$, and is traceless on-shell. If all this holds, it can be shown that the theory is conformally invariant.

- We have hence shown that in some theories, dilation symmetry is sufficient to show both conformal invariance and tracelessness of an improved stress-energy tensor on-shell. However, it is not sufficient for all theories; electromagnetism in $d = 3$ is a counterexample, as shown here. But it holds for all theories in $d = 2$ under mild technical assumptions, as shown here.

- If all matter fields have vanishing conformal dimension, then conformal invariance is equivalent to tracelessness of the energy-momentum tensor off-shell. Considering a Weyl transformation, we have $\delta S_2 + \delta S_4 = 0$ and $\delta S_2 = 0$, so $\delta S_4 = 0$. However, δS_4 is simply the integral of $\omega(x) T^a{}_a$, and hence the trace vanishes. Furthermore, it can be shown that if the energy-momentum tensor is traceless off-shell, then all matter fields can be chosen to have vanishing conformal dimension. **(how?)**

5 Linearized Theory

5.1 The Linearized Einstein Equation

In a situation with weak gravity, the Einstein equation is approximately linear. We perform some setup to approach this result.

- We work in "almost inertial" coordinates
$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$$
and assume the spacetime manifold is $M = \mathbb{R}^4$. We work to lowest order in h everywhere. Because of this, we may raise and lower indices on h itself with the flat metric, since the correction is second order, so we may think of h as a tensor field in a flat spacetime background.

- As shown earlier, the inverse metric is
$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu}, \quad h^{\mu\nu} = \eta^{\mu\rho}\eta^{\nu\sigma}h_{\rho\sigma}$$
which follows from the identity $g^{\mu\nu}g_{\nu\rho} = \delta^\mu_\rho$.

- To first order, the Christoffel symbols are
$$\Gamma^\mu_{\nu\rho} = \frac{1}{2}\eta^{\mu\sigma}(h_{\sigma\nu,\rho} + h_{\sigma\rho,\nu} - h_{\nu\rho,\sigma}).$$
That is, the lowest-order term is $O(h)$. Then we may neglect $\Gamma\Gamma$ terms in the Riemann tensor,
$$R_{\mu\nu\rho\sigma} = \eta_{\mu\tau}(\partial_\rho\Gamma^\tau_{\nu\sigma} - \partial_\sigma\Gamma^\tau_{\nu\rho}) = \frac{1}{2}(h_{\mu\sigma,\nu\rho} + h_{\nu\rho,\mu\sigma} - h_{\nu\sigma,\mu\rho} - h_{\mu\rho,\nu\sigma}).$$
Contracting a pair of indices gives
$$R_{\mu\nu} = \partial^\rho\partial_{(\mu}h_{\nu)\rho} - \frac{1}{2}\partial^\rho\partial_\rho h_{\mu\nu} - \frac{1}{2}\partial_\mu\partial_\nu h, \quad h = h^\mu{}_\mu.$$

- Thus, we arrive at the linearized Einstein equation,
$$G_{\mu\nu} = \partial^\rho\partial_{(\mu}h_{\nu)\rho} - \frac{1}{2}\partial^\rho\partial_\rho h_{\mu\nu} - \frac{1}{2}\partial_\mu\partial_\nu h - \frac{1}{2}\eta_{\mu\nu}(\partial^\rho\partial^\sigma h_{\rho\sigma} - \partial^\rho\partial_\rho h) = 8\pi T_{\mu\nu}.$$
All quantities above besides the metric itself are $O(h)$ as they vanish in Minkowski space. Moreover, we see that $T_{\mu\nu}$ must also be small for consistency.

- The equation above can be simplified using the trace-reversed metric perturbation
$$\overline{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}h\eta_{\mu\nu}, \quad h_{\mu\nu} = \overline{h}_{\mu\nu} - \frac{1}{2}\overline{h}\eta_{\mu\nu}$$
where the second equation follows because $\overline{h} = -h$. Then we find
$$G_{\mu\nu} = -\frac{1}{2}\partial^\rho\partial_\rho\overline{h}_{\mu\nu} + \partial^\rho\partial_{(\mu}\overline{h}_{\nu)\rho} - \frac{1}{2}\eta_{\mu\nu}\partial^\rho\partial^\sigma\overline{h}_{\rho\sigma}.$$
That is, the trace terms have canceled out.

Next, we simplify the situation using the gauge symmetry.

- We don't want to apply an arbitrary diffeomorphism, because the resulting metric might be far from the Minkowski metric. Instead we restrict ourselves to infinitesimal diffeomorphisms,

$$(\phi_{-t})_* T = T + t\mathcal{L}_X T + O(t^2) = T + \mathcal{L}_\xi T$$

  where $\xi^a = tX^a$, X^a is the vector field that generates ϕ_t, and T is an arbitrary tensor.

- We expand all quantities to lowest order in ξ and h. Since all our quantities above are already $O(h)$, they are completely unchanged. However, the metric picks up a new term,

$$(\phi_{-t})_*(g) = g + \mathcal{L}_\xi g + \ldots = \eta + h + \mathcal{L}_\xi \eta + \ldots$$

  which yields the shift

$$h_{\mu\nu} \to h_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu.$$

  This is closely analogous to the gauge transformation of A_μ in electromagnetism. In that case, $F^{\mu\nu}$ is gauge invariant; in this case, the Riemann tensor is.

- We choose the harmonic/Lorenz/de Donder gauge

$$\partial^\nu \overline{h}_{\mu\nu} = 0.$$

  To see this is possible, note that under a gauge transformation we have

$$\overline{h}_{\mu\nu} \to \overline{h}_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu - \eta_{\mu\nu} \partial^\rho \xi_\rho$$

  and thus

$$\partial^\nu \overline{h}_{\mu\nu} \to \partial^\nu \overline{h}_{\mu\nu} + \partial^\nu \partial_\nu \xi_\mu$$

  so we simply choose ξ_μ to be $\Box^{-1}(\partial^\nu \overline{h}_{\mu\nu})$, which is possible with a Green's function and appropriate boundary conditions.

- In harmonic gauge, the linearized Einstein equation (with $G = 1$) simplifies to

$$\partial^\rho \partial_\rho \overline{h}_{\mu\nu} = -16\pi T_{\mu\nu}.$$

  That is, every component of $\overline{h}_{\mu\nu}$ satisfies the wave equation, sourced by $T_{\mu\nu}$, so the Einstein equation can be solved by Green's functions.

5.2 Gravitational Waves

We now characterize the gravitational wave solutions.

- We work in vacuum and look for plane wave solutions,

$$\overline{h}_{\mu\nu}(x) = \operatorname{Re} H_{\mu\nu} e^{ik_\rho x^\rho}.$$

  The linearized Einstein equation gives $k_\mu k^\mu = 0$, confirming that gravitational waves move at the speed of light. Here, $H_{\mu\nu}$ is a constant complex symmetric matrix describing the polarization of the wave.

- The harmonic gauge condition yields
$$k^\nu H_{\mu\nu} = 0$$
which tells us the waves are transverse. Now note that there is additional gauge freedom by choosing any ξ_μ with $\partial^\nu \partial_\nu \xi_\mu = 0$. Taking $\xi_\mu = X_\mu e^{ik_\rho x^\rho}$ and using the transformation of $\overline{h}_{\mu\nu}$,
$$H_{\mu\nu} \to H_{\mu\nu} + i(k_\mu X_\nu + k_\nu X_\mu - \eta_{\mu\nu} k^\rho X_\rho).$$
One can show this is enough to impose the longitudinal and trace-free conditions
$$H_{0\mu} = 0, \quad H^\mu{}_\mu = 0.$$
In this gauge, $h_{\mu\nu} = \overline{h}_{\mu\nu}$. The combination of all the conditions above is called transverse traceless gauge.

- Now consider a wave traveling in the z-direction, so $k^\mu = \omega(1, 0, 0, 1)$ and the transverse condition gives $H_{\mu 0} + H_{\mu 3} = 0$. Combining all of our constraints gives two degrees of freedom,
$$H_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & H_+ & H_\times & 0 \\ 0 & H_\times & -H_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$
The H_+ and $H_\times$ solutions cause stretching and shrinking in a "plus" and "cross" configuration.

- To show the gravitational wave has a physical effect, consider a particle at rest in Minkowski space. Then the geodesic equation is
$$\dot{u}^\alpha + \Gamma^\alpha_{00} = 0.$$
Now suppose a gravitational wave passes through, so the metric becomes $g = \eta + h$ and our frame is almost inertial. The transverse traceless condition imposes $\Gamma^\alpha_{00} = 0$, so the particle has no coordinate acceleration. Indeed, one can think of this as the definition of transverse traceless gauge: its coordinates are defined by the positions of freely falling test particles.

- However, the metric is perturbed as
$$ds^2 = -dt^2 + (1 + h_+)dx^2 + (1 - h_+)dy^2 + 2h_\times dx dy + dz^2$$
which means that the proper distance between different particles changes. The change in proper distance can be measured via interferometry, as done in LIGO.

We can also justify this result within an inertial frame.

- We begin by constructing a physical inertial frame. Consider an observer moving along a geodesic. At a point p, they hold three orthogonal measuring rods. This defines an orthonormal basis $\{e_\alpha\}$ at p where
$$e_0^a = u^a, \quad u_a e_i^a = 0, \quad g_{ab} e_i^a e_j^b = \delta_{ij}$$
where u^a is the observer's four-velocity.

- As the rulers fall, their directions are parallel transported,

$$u^b \nabla_b e_i^a = 0$$

which defines an inertial frame at every point along the geodesic. Moreover, by metric compatibility, the orthonormality of the rulers is preserved. Therefore, a change in the coordinate separation of particles in this family of frames corresponds to an increase in proper distance between them.

- Now suppose the observer releases a family of freely-falling test particles. The geodesic deviation equation is

$$u^b \nabla_b u^c \nabla_c S_a = R_{abcd} u^b u^c S^d.$$

Contracting with e_α^a and using the fact that the basis is parallel transported,

$$u^b \nabla_b u^c \nabla_c (e_\alpha^a S_a) = R_{abcd} e_\alpha^a u^b u^c S^d.$$

The quantity $e_\alpha^a S_a = S_\alpha$ is a scalar, giving the coordinate separation along the α direction, so the covariant derivative simplifies to

$$\frac{d^2 S_\alpha}{d\tau^2} = R_{abcd} e_\alpha^a u^b u^c e_\beta^d S^\beta.$$

- Now, while S_α is a property of the observer's frame, we can evaluate the right-hand side in any frame. Thus we revert to almost inertial coordinates, working in the linearized approximation. Since the Riemann tensor is already $O(h)$, everything else on the right-hand side can be evaluated in flat spacetime. We assume the observer is at rest, $u^\mu = (1, 0, 0, 0)$, so

$$\frac{d^2 S_\alpha}{d\tau^2} \approx R_{\mu 00\nu} e_\alpha^\mu e_\beta^\nu S^\beta \approx \frac{1}{2} \frac{\partial^2 h_{\mu\nu}}{\partial t^2} e_\alpha^\mu e_\beta^\nu S^\beta.$$

- To leading order, we have $e_1^\mu = (0, 1, 0, 0)$, $e_2^\mu = (0, 0, 1, 0)$ and $e_3^\mu = (0, 0, 0, 1)$, and the position of the observer is $x^\mu \approx (\tau, 0, 0, 0)$. Then since $h_{0\mu} = h_{3\mu} = 0$,

$$\frac{d^2 S_0}{d\tau^2} = \frac{d^2 S_3}{d\tau^2} = 0$$

so there is no relative acceleration of the test particles in the z-direction.

- For a purely + polarized wave, we have

$$\frac{d^2 S_1}{d\tau^2} = -\frac{1}{2} \omega^2 |H_+| \cos(\omega t - \alpha) S_1, \quad \frac{d^2 S_2}{d\tau^2} = \frac{1}{2} \omega^2 |H_+| \cos(\omega \tau - \alpha) S_2, \quad \alpha = \arg H_+.$$

These differential equations can be solved perturbatively. Choosing the test particles to be initially at rest, at leading order S_1 and S_2 are constant. Then the next-order correction is an oscillation, just as we found heuristically earlier.

- Taking linear combinations in analogy with circularly polarized electromagnetic waves gives gravitational waves which distort a circle of particles into a rotating ellipse. Note that the particles don't rotate around, only the wave does; the individual particles barely move at all.

Note. In Newtonian physics, gravitational waves can be described as oscillating forces that act on matter, and stretch and shrink light. LIGO has widely separated freely suspended mirrors, whose separation oscillates when a gravitational wave passes through. This leads to a common question: how can LIGO use light as a ruler, when the light is stretched too?

For simplicity, consider a "step function" gravitational wave. The point is that the moment the step function turns on, only the light currently inside the detector is stretched. After a time $2L/c$, all the light in the detector has the same wavelength as before, but the detector size remains changed, so the effect is observable. Switching back to a sinusoidal gravitational wave with wavelength λ, the phase lag effect here is negligible as long as $L \ll \lambda$.

5.3 Far-Field Limit

Next, we investigate the far-field limit for a source of gravitational waves.

- The linearized Einstein equation can be solved with the retarded Green's function, giving
$$\overline{h}_{\mu\nu}(t, \mathbf{x}) = 4 \int d\mathbf{x}' \, \frac{T_{\mu\nu}(t - |\mathbf{x} - \mathbf{x}'|, \mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|}$$
where $|\mathbf{x} - \mathbf{x}'|$ is calculated with the Euclidean metric. As usual, $\mathbf{x}$ indicates the field point and $\mathbf{x}'$ indicates the source point.

- Now suppose the matter is confined to a region near the origin of size d. Then for $r = |\mathbf{x}| \gg d$,
$$|\mathbf{x} - \mathbf{x}'|^2 \approx r^2(1 - (2/r)\hat{\mathbf{x}} \cdot \mathbf{x}'), \quad |\mathbf{x} - \mathbf{x}'| \approx r - \hat{\mathbf{x}} \cdot \mathbf{x}'$$
to first order in d/r, which implies
$$T_{\mu\nu}(t - |\mathbf{x} - \mathbf{x}'|, \mathbf{x}') \approx T_{\mu\nu}(t', \mathbf{x}') + \hat{\mathbf{x}} \cdot \mathbf{x}'(\partial_0 T_{\mu\nu}(t', \mathbf{x}')), \quad t' = t - r.$$

- Next, let τ be the timescale on which $T_{\mu\nu}$ varies. Assuming the matter is moving nonrelativistically, we have $d \ll \tau$ and $\partial_0 T_{\mu\nu} \sim T_{\mu\nu}/\tau$, so the second term above is negligible,
$$\overline{h}_{ij}(t, \mathbf{x}) \approx \frac{4}{r} \int d\mathbf{x}' \, T_{ij}(t', \mathbf{x}'), \quad t' = t - r.$$
We only consider the spatial components of $\overline{h}$ here, but the gauge conditions can be used to recover the others.

- Next, we evaluate the integral above. We suppress the t' since it does not depend on the integration variable $\mathbf{x}'$. We drop primes on the coordinates; since the matter is compactly supported we can ignore surface terms. We use energy-momentum conservation, which at this order is $\partial_\nu T^{\mu\nu} = 0$. Finally, we use the trick $\delta^i_j = \partial_j x^i$ to introduce powers of x, because we want two powers of x to get a quadrupole moment.

- Applying these tricks, we have
$$\int d\mathbf{x} \, T^{ij} = \int d\mathbf{x} \, T^{ik} \partial_k x^j = -\int d\mathbf{x} \, (\partial_k T^{ik}) x^j = \int d\mathbf{x} \, \partial_0 (T^{i0}) x^j.$$
Next, symmetrizing on i and j and introducing another power of x gives
$$\int d\mathbf{x} \, T^{ij} = \frac{1}{2} \partial_0 \partial_0 \int d\mathbf{x} \, T_{00} x^i x^j$$
where we used $T^{00} = T_{00}$ to leading order.

- Finally, defining the quadrupole moment of energy as
$$I_{ij}(t) = \int d\mathbf{x}\, T_{00}(t,\mathbf{x}) x^i x^j$$
we have
$$\overline{h}_{ij}(t,\mathbf{x}) \approx \frac{2}{r}\ddot{I}_{ij}(t-r).$$
This result is very similar to the expression for electric dipole radiation.

- To recover the other components, note that
$$\partial_0 \overline{h}_{0i} = \partial_j \overline{h}_{ji}, \quad \partial_0 \overline{h}_{00} = \partial_i \overline{h}_{0i}.$$
Taking the time integral is straightforward; taking the spatial derivative gives two terms, and we use the fact that we in the radiation zone, $r \gg \tau$, to pick the larger term. Then
$$\overline{h}_{0i} \approx -\frac{2\hat{x}_j}{r}\ddot{I}_{ij}(t-r), \quad \overline{h}_{00} \approx \frac{2\hat{x}_i\hat{x}_j}{r}\ddot{I}_{ij}(t-r).$$

- On the other hand, we can also calculate these terms by naively expanding our original equation, giving
$$\overline{h}_{00} \approx \frac{4E}{r}, \quad \overline{h}_{0i} \approx -\frac{4P_i}{r}$$
where E and P_i are the first and second moments of energy,
$$E = \int d\mathbf{x}'\, T_{00}(t', \mathbf{x}'), \quad P_i = -\int d\mathbf{x}'\, T_{0i}(t',\mathbf{x}').$$
These look completely different from our expressions above. This is because they represent the lowest order contributions, while our previous expressions for $\overline{h}_{0i}$ and $\overline{h}_{00}$ are higher order in d/τ. We didn't see these terms in our higher order expressions because there we dropped integration constants.

- Using energy-momentum conservation to lowest order shows that E is constant, even though we expect gravitational waves to carry away energy. This is because the loss of energy is a higher-order effect; the term in $\nabla_\mu T^{\mu\nu} = 0$ we have neglected is second order in h. Thus to treat the energy loss consistently we have to expand everything to second order.

- In the center of momentum frame where $P_i = 0$, the two leading terms together give
$$\overline{h}_{00}(t,\mathbf{x}) \approx \frac{4M}{r} + \frac{2\hat{x}_i\hat{x}_j}{r}\ddot{I}_{ij}(t-r), \quad \overline{h}_{0i}(t,\mathbf{x}) \approx -\frac{2\hat{x}_j}{r}\ddot{I}_{ij}(t-r).$$
In summary, we are working to first order in h, and assuming $r \gg \tau \gg d$ and expanding up to the lowest-order time-dependent term.

Note. Comparison with electromagnetic radiation. Radiation can be sourced by the multipole moments of electric charge. Since the monopole moment (the total charge) is conserved, the lowest-order effect is dipole radiation. Analogously, gravitational waves can be sourced by T^{00}. The monopole term is constant by conservation of energy. For the dipole term, boost symmetry yields $\mathbf{P}/E = \dot{\mathbf{X}}$ where $\mathbf{X}$ is the dipole moment of energy, so $\ddot{\mathbf{X}}$ is zero by conservation of momentum. Hence the lowest-order effect is quadrupole radiation.

Radiation can also be sourced by currents. For electromagnetism, the lowest-order term is from a changing magnetic dipole moment, but this radiation is much weaker than electric dipole radiation if the charges are moving nonrelativistically. For gravitational waves, the analogous radiation is sourced by T^{0i}, and there is no dipole term because of conservation of angular momentum. Then the lowest-order term is the analogue of magnetic quadrupole, which is much weaker than the electric quadrupole radiation we considered above.

Note. Above, we started with the linearized theory and took the far field limit. Alternatively, if we work in a frame where all the matter is moving slowly, with speed u, we could instead perform an expansion in u. The elements of the stress tensor have magnitude $T_{00} \sim u^0$, $T_{0i} \sim u$, and $T_{ij} \sim u^2$. At zeroth order in u, we simply recover the Newtonian limit. At first order, we get the theory of gravitoelectromagnetism. Incorporating the effect of T_{0i} gives the metric

$$ds^2 = -(1-2\Phi)dt^2 + 2(\mathbf{A}\cdot d\mathbf{x})\,dt + (1+2\Phi)\delta_{ij}dx^i dx^j$$

where the elements of the "gravitoelectromagnetic four-potential" are

$$\Phi = \int d\mathbf{x}'\,\frac{T_{00}(t_{\text{ret}},\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}, \quad A_i = \int d\mathbf{x}'\,\frac{T_{0i}(t_{\text{ret}},\mathbf{x}')}{|\mathbf{x}-\mathbf{x}'|}.$$

Gravitoelectromagnetism is similar to actual electromagnetism, with energy and momentum corresponding to charge and current, but it isn't Lorentz invariant since it requires working in a frame where u is small. (Essentially, it is treating $T_{0\mu}$ as if it were a four-vector.) However, it is useful for analyzing the results of precision laboratory tests of gravity, such as Gravity Probe B.

5.4 Energy of Gravitational Waves

To define the energy in gravitational waves, we need to work to second order in the metric perturbation. Setting up the perturbation theory is a bit subtle.

- We expand the metric to second order as

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} + h^{(2)}_{\mu\nu}.$$

  At second order, the Einstein tensor is

$$G_{\mu\nu}[g] = G^{(1)}_{\mu\nu}[h] + G^{(1)}_{\mu\nu}[h^{(2)}] + G^{(2)}_{\mu\nu}[h].$$

  Here, $G^{(n)}_{\mu\nu}[h']$ refers to the part of the expansion of G for $g = \eta + h'$ that is n^{th} order in h'. Then the first term is the first-order term we calculated earlier, and there are two second order terms: $G^{(1)}_{\mu\nu}[h^{(2)}]$, which is linear in the second order term, or $G^{(2)}_{\mu\nu}[h]$, which is quadratic in the first order term.

- The quadratic part can be written as

$$G^{(2)}_{\mu\nu}[h] = R^{(2)}_{\mu\nu}[h] - \frac{1}{2}R^{(1)}[h]h_{\mu\nu} - \frac{1}{2}R^{(2)}[h]\eta_{\mu\nu}$$

  where we introduce a similar notation for the Ricci tensor and scalar; we get two terms from the Ricci scalar $g_{\mu\nu}R^{\mu\nu}$ because both the metric and Ricci tensor can be perturbed.

- Similarly, the quadratic part of the Ricci scalar can be expanded as

$$R^{(2)}[h] = \eta^{\mu\nu} R^{(2)}_{\mu\nu}[h] - h^{\mu\nu} R^{(1)}_{\mu\nu}[h].$$

  The quadratic part of the Ricci tensor is a long expression with eight terms.

- Now suppose that no matter is present. Then at first order, the linearized Einstein equation is $G^{(1)}_{\mu\nu}[h] = 0$, or equivalently $R^{(1)}_{\mu\nu}[h] = 0$. At second order, we have

$$G^{(1)}_{\mu\nu}[h^{(2)}] = 8\pi t_{\mu\nu}[h], \quad t_{\mu\nu}[h] \equiv -\frac{1}{8\pi} G^{(2)}_{\mu\nu}[h].$$

  That is, the first order term h acts as a source for the second order term $h^{(2)}$, and we can solve for $h^{(2)}$ using Green's function methods similar to before.

- Finally, assuming that the linearized Einstein equation holds, we have

$$t_{\mu\nu}[h] = -\frac{1}{8\pi}\left(R^{(2)}_{\mu\nu}[h] - \frac{1}{2}\eta^{\rho\sigma} R^{(2)}_{\rho\sigma}[h]\eta_{\mu\nu} \right).$$

We would like to interpret $t_{\mu\nu}[h]$ as an energy momentum tensor for the gravitational field.

- First, we check conservation. At first order, the contracted Bianchi identity $g^{\mu\rho}\nabla_\rho G_{\mu\nu} = 0$ is

$$\partial^\mu G^{(1)}_{\mu\nu}[h] = 0$$

  for arbitrary h. At second order, we have

$$\partial^\mu (G^{(1)}_{\mu\nu}[h^{(2)}] + G^{(2)}_{\mu\nu}[h]) + \text{“}hG^{(1)}[h]\text{”} = 0$$

  where the third term stands for the first order changes in $g^{\mu\rho}\nabla_\rho$. The first term vanishes by the first-order contracted Bianchi identity with $h = h^{(2)}$, and the third term vanishes by the linearized Einstein equation. The remaining term is simply $\partial^\mu t_{\mu\nu}[h]$, giving the result.

- Therefore, $t_{\mu\nu}$ is a symmetric tensor that is quadratic in h, conserved if h obeys the equations of motion, and appears on the right-hand side of the second-order Einstein equation, all of which are good properties for a stress-energy tensor.

- The problem with $t_{\mu\nu}$ is that it is not gauge-invariant; this is expected, as we do not expect to have a local expression for the gravitational field energy. However, one can show that the integral of t_{00} over a surface of constant time is gauge invariant, as long as $h_{\mu\nu}$ is restricted to decay at infinity, giving a definition for the total energy of the gravitational field.

- One can use the second order Einstein equation to convert the spatial integral, which is quadratic in $h_{\mu\nu}$, into a surface integral at infinity which is linear in $h^{(2)}_{\mu\nu}$. Indeed, this works for any asymptotically flat spacetime, outside the weak-field approximation, and the result is called the ADM energy.

- The ADM energy is conserved, while a related quantity, the Bondi energy, is non-increasing in time. Its rate of change can be interpreted as the rate of energy loss in gravitational waves.

- We will instead follow a less rigorous but more intuitive procedure. For any point p, we consider a region containing it of typical coordinate radius a, and define the average of a tensor $X_{\mu\nu}$ at p by
$$\langle X_{\mu\nu} \rangle = \int_R X_{\mu\nu}(x) W(x) d^4x$$
where $W(x)$ is positive, integrates to one on R, and is zero at ∂R. We imagine this integration as taking place in Minkowski space, since corrections to this are higher-order, so it makes sense to add the integrand $X_{\mu\nu}(x)$ at different points.

- Now, by integrating by parts we have
$$\langle \partial_\mu X_{\nu\rho} \rangle = -\int_R X_{\nu\rho} \partial_\mu W(x) d^4x.$$
If the components $X_{\mu\nu}$ change on the scale λ, we would naively expect the left-hand side to be $O(X/\lambda)$, but instead it is $O(X/a)$. Hence if we choose $a \gg \lambda$, the average of a total derivative is negligible, and hence we can integrate by parts,
$$\langle A \partial B \rangle = \langle \partial(AB) \rangle - \langle (\partial A) B \rangle \approx -\langle (\partial A) B \rangle.$$

- Using this identity, it can be shown using the linearized Einstein field equation that
$$\langle \eta^{\mu\nu} R^{(2)}_{\mu\nu}[h] \rangle = 0.$$
Then our expression for $\langle t_{\mu\nu} \rangle$ reduces to
$$\langle t_{\mu\nu} \rangle = \frac{1}{32\pi} \langle \partial_\mu \bar{h}_{\rho\sigma} \partial_\nu \bar{h}^{\rho\sigma} - \frac{1}{2} \partial_\mu \bar{h} \partial_\nu \bar{h} - 2 \partial_\sigma \bar{h}^{\rho\sigma} \partial_{(\mu} \bar{h}_{\nu)\rho} \rangle$$
and we can explicitly check it is gauge invariant.

- Note that the averaging region can be quite large. For example, LIGO is sensitive to waves with $\lambda \sim 3000\,\text{km}$, so the averaging region is larger than the Earth.

Example. For the plane gravitational wave considered earlier, we have
$$\langle t_{\mu\nu} \rangle = \frac{1}{32\pi} (|H_+|^2 + |H_\times|^2) k_\mu k_\nu$$
where the last two terms in $\langle t_{\mu\nu} \rangle$ drop out by the gauge condition, we pick up a factor of $1/2$ from averaging a squared sinusoid, and we pick up a factor of 2 from $H^{\mu\nu} H_{\mu\nu}$.

In particular, the magnitude of the energy flux is
$$\mathcal{F} \sim \frac{\omega^2 c}{32\pi G} h^2 \sim 0.01\,\text{W/m}^2$$
where we reinstated factors of c and G and used the typical parameters $h \sim 10^{-21}$ and $\omega \sim 100\,\text{Hz}$. This is a large result: the difficulty of detecting gravitational waves comes not from the size of the energy flux but from the weakness of its interaction with matter. Intuitively, one can think of spacetime as being very stiff.

5.5 The Quadrupole Formula

In this section, we derive the famous quadrupole formula for the rate of energy loss due to gravitational radiation.

Note. Isotropic integrals. Consider the three-dimensional Cartesian integral

$$I_{ij} = \int d\Omega\, \hat{x}_i \hat{x}_j.$$

By rotational symmetry, the answer must be proportional to some combination of the Levi–Civita symbol ϵ^{ijk} and Kronecker delta. Then the only possibility is $I_{ij} \propto \delta_{ij}$, and taking the trace gives

$$I_{ij} = \frac{4\pi}{3}\delta_{ij}.$$

Next, consider the integral

$$I_{ijk\ell} = \int d\Omega\, \hat{x}_i \hat{x}_j \hat{x}_k \hat{x}_\ell.$$

The Levi–Civita does not appear as it has three indices; we could use two Levi–Civitas and contract them together, but this just reduces to Kronecker deltas. Then the most general possibility is

$$I_{ijk\ell} = \alpha\delta_{ij}\delta_{k\ell} + \beta\delta_{ik}\delta_{j\ell} + \gamma\delta_{i\ell}\delta_{jk}$$

and total symmetry fixes $\alpha = \beta = \gamma$. Tracing $i = j$ and $k = \ell$ gives $\alpha = 4\pi/15$.

Next, we derive the quadrupole formula.

- The averaged energy flux across a large sphere of radius r is

$$\langle P \rangle = -\int d\Omega\, r^2 \langle t_{0i} \rangle \hat{x}_i$$

 where we picked up a minus sign from lowering indices on t.

- Plugging in our expression for $\langle t_{\mu\nu} \rangle$ in harmonic gauge and explicitly separating spatial and temporal components,

$$\begin{aligned}\langle t_{0i} \rangle &= \frac{1}{32\pi}\langle \partial_0 \overline{h}_{\rho\sigma} \partial_i \overline{h}^{\rho\sigma} - \frac{1}{2}\partial_0 \overline{h} \partial_i \overline{h} \rangle \\ &= \frac{1}{32\pi}\langle \partial_0 \overline{h}_{jk} \partial_i \overline{h}_{jk} - 2\partial_0 \overline{h}_{0j} \partial_i \overline{h}_{0j} + \partial_0 \overline{h}_{00} \partial_i \overline{h}_{00} - \frac{1}{2}\partial_0 \overline{h} \partial_i \overline{h} \rangle\end{aligned}$$

 where a sign flip occurs due to the metric; we work exclusively with lowered indices from this point. The rest of the derivation is simply a matter of plugging in our earlier results.

- First, since $\overline{h}_{jk}(t, \mathbf{x}) = (2/r)\ddot{I}_{jk}(t - r)$, we have

$$\partial_0 \overline{h}_{jk} = \frac{2}{r}\dddot{I}_{jk}(t - r), \quad \partial_i \overline{h}_{jk} = \left(-\frac{2}{r}\dddot{I}_{jk}(t - r) - \frac{2}{r^2}\ddot{I}_{jk}(t - r) \right) \hat{x}_i.$$

 The second term in brackets is negligible since $\tau/r \ll 1$. Then these terms contribute

$$-\frac{1}{32\pi}\int d\Omega\, r^2 \langle \partial_0 \overline{h}_{jk} \partial_i \overline{h}_{jk} \rangle \hat{x}_i = \frac{1}{2}\langle \dddot{I}_{ij} \dddot{I}_{ij} \rangle_{t-r}$$

 to the power, where the average is taken over a time much larger than τ, a distance much greater than λ, and centered on the retarded time $t - r$.

- Next, we have $\overline{h}_{0j} = (-2\hat{x}_k/r)\ddot{I}_{jk}(t-r)$, which implies

$$\partial_0 \overline{h}_{0j} = -\frac{2\hat{x}_k}{r}\dddot{I}_{jk}(t-r), \quad \partial_i \overline{h}_{0j} \approx \frac{2\hat{x}_k}{r}\dddot{I}_{jk}(t-r)\hat{x}_i.$$

  The resulting contribution is $(-1/3)\langle \dddot{I}_{ij}\dddot{I}_{ij}\rangle_{t-r}$ where we performed an isotropic integral.

- For the third term, we have $\overline{h}_{00} = 4M/r + (2\hat{x}_j\hat{x}_k/r)\ddot{I}_{jk}(t-r)$, so

$$\partial_0 \overline{h}_{00} = \frac{2\hat{x}_j\hat{x}_k}{r}\dddot{I}_{jk}(t-r), \quad \partial_i \overline{h}_{00} \approx -\left(\frac{4M}{r^2} - \frac{2\hat{x}_j\hat{x}_k}{r}\dddot{I}_{jk}(t-r)\right)\hat{x}_i.$$

  We can ignore the $4M/r^2$ term, since it leads to a term proportional to $\langle \dddot{I}_{jk}\rangle$, which is small since it is the average of a derivative. Then the remaining term can be evaluated using the isotropic integral above.

- Similarly, for the final term we use $\overline{h} = \overline{h}_{jj} - \overline{h}_{00}$. We find four terms that contribute, though they're all of the same form as previous terms. Combining all terms together gives

$$\langle P\rangle = \frac{1}{5}\langle \dddot{I}_{ij}\dddot{I}_{ij} - \frac{1}{3}\dddot{I}_{ii}\dddot{I}_{jj}\rangle_{t-r}.$$

  Defining the energy quadrupole tensor as the traceless part of I_{ij},

$$Q_{ij} = I_{ij} - \frac{1}{3}I_{kk}\delta_{ij}, \quad \boxed{\langle P\rangle = \frac{1}{5}\langle \dddot{Q}_{ij}\dddot{Q}_{ij}\rangle_{t-r}.}$$

  This is the famous quadrupole formula, valid in the radiation zone of a nonrelativistic source, $r \gg \tau \gg d$.

- As an example, a spherically symmetric body has $Q_{ij} = 0$ and hence radiates no power. This is in agreement with Birkhoff's theorem.

We now estimate the energy emitted by a realistic binary system.

- Consider two stars with mass M and separation d in the nonrelativistic regime $d \gg M$. By Newtonian mechanics, we have $\omega \sim M^{1/2}d^{-3/2}$. The quadrupole tensor has components of typical size Md^2, so the third derivative is of order $Md^2\omega^3$. Applying the quadrupole formula,

$$P \sim (M/d)^5.$$

  This is a strong dependence; it's much easier to see gravitational waves from compact binaries.

- Usually $d \gg M$, but this is compensated by the units of c and G,

$$P \sim (M/d)^5 L_{\text{planck}}, \quad L_{\text{planck}} = \frac{c^5}{G} \approx 4 \times 10^{52}\,\text{W}.$$

  The luminosity of all stars in the universe is about $10^{-5}L_{\text{planck}}$, so a binary system with $M/d \sim 0.1$ would equal their power output.

- A typical star has $M/R \sim 10^{-6}$ and a typical binary star system has $d \gg R$, so they are far from detectable. On the other hand, a Schwarzschild black hole has $M/R \sim 1$, so we expect to able to see the moments before black holes merge, where $d \sim R$. Neutron stars are also very compact, so we can see NS/NS and NS/BH mergers. Of course, our Newtonian result above breaks down close to merger; in this case we must turn to numerical simulation.

- Note that since neutron stars obey the stringent TOV limit, almost all NS/NS mergers will result in black holes.
- The power lost to gravitational waves causes orbits to decay. In 1974, Hulse and Taylor detected a binary pulse, an NS/NS binary where one of the stars is a pulsar. Pulsars emit a beam of radio waves, which align with the Earth periodically; thus the change of the orbital period could be measured, agreeing with the quadrupole formula.
- For direct detection, we refer back to our expression for $\bar{h}_{ij}$, giving
$$\bar{h}_{ij} \sim \frac{Md^2\omega^2}{r} \sim \frac{M^2}{dr}.$$
A signal with $M = 10M_\odot$ and d about ten times the Schwarzschild radius gives $h \sim 10^{-21}$ and $\omega \sim 100\,\text{Hz}$, the general type of signal LIGO is made to detect.
- One can show that, if the frequency and rate of change of frequency of the gravitational waves can be measured, one can infer the chirp mass
$$M_{\text{chirp}} = \frac{(M_1M_2)^{3/5}}{(M_1+M_2)^{1/5}}.$$
The amplitude of the waves can be used to deduce the distance to the source. Finally, the "ringdown" after merger only depends on the final black hole mass. In practice, a complex best-fit protocol is used to estimate these parameters.
- In 2015, LIGO made the first detection of a BH/BH merger, about a billion light years away. In 2017, LIGO detected an NS/NS merger, which was accompanied by the detection of electromagnetic radiation. These dual detections have a variety of applications, such as determining the Hubble constant.

Note. An estimate of LIGO's sensitivity. The fractional changes in length are of order $h \sim 10^{-21}$. We hence need to measure phase differences
$$\Delta\phi \sim k\Delta L \sim \frac{hL}{\lambda} \sim 4 \times 10^{-12}.$$
As a mild oversimplification, each photon has a probability $|\Delta\phi|^2$ of being detected in the other interferometer output, so we need $N \sim 10^{25}$ photons in the apparatus to see this reliably. This corresponds to a laser power $P \sim 100\,\text{kW}$. LIGO makes this easier by using a cavity where the photons bounce back and forth about 100 times, requiring only about 1/100 as much input power.

LIGO also uses a frequency modulation scheme that significantly improves on the naive interferometer scheme we described here, though we still get within a few orders of magnitude because of the amplitude-phase uncertainty principle. There is an upper bound on power because otherwise the radiation pressure, and hence fluctuations in pressure, on the mirrors would wash out the signal.

Note. The history of the theory and detection of gravitational waves is also notable. Little progress was made in the 1920s and 1930s, as most physicists turned to quantum mechanics. Because of the conceptual difficulties of general relativity, the fact that gravitational waves exit and can carry energy was not well-established until the late 1950s, and probably not accepted by Einstein himself. Attempts to detect gravitational waves began with Weber in the 1960s, with a false detection in 1969. LIGO's detection of gravitational waves in 2015 marked the culmination of a half-century of work. For more details, see *Gravitational Waves and the Long Relativity Revolution* and *Sufficiently Advanced Technology for Gravitational Wave Detection*.

6 The Schwarzschild Solution

6.1 The Schwarzschild Metric

Before starting, we define some useful properties of spacetimes.

- A spacetime is symmetric in a variable s if there exist coordinates x^α so that one of the x^α is s, and the metric components don't depend on s.
- A spacetime is stationary if there exist coordinates x^α so that x^0 is timelike at infinity and the metric components don't depend on x^0. Equivalently, the spacetime has a Killing vector that is timelike at infinity. We need the 'at infinity' condition because x^0 may become spacelike in the interior, as it does for the Schwarzschild metric. If the Killing vector is always timelike, the spacetime is strictly stationary.
- A spacetime is static if it is stationary, and there are no cross-terms in the metric, $g_{0i} = 0$. Equivalently, the timelike Killing vector is orthogonal to a family of hypersurfaces.
- Intuitively, a stationary spacetime is 'doing the same thing at all times', since it has time translational symmetry, while a static spacetime is 'doing nothing at all', since it has time reversal symmetry; a cross term in the metric would pick up a sign, $dt dx^i \to -dt dx^i$. For example, a rotating black hole is stationary but not static.

Note. Relating the two definitions of stationary spacetime. Suppose we start with a Killing vector k^a. Then take a hypersurface Σ nowhere tangent to k^a. We define spatial coordinates x^i on Σ. The spacetime point (t, x^i) is reached by starting on Σ and following an integral curve of k^a for parameter distance t. Then the metric components are independent of t.

Note. Relating the two definitions of static spacetime. First, we give a criterion for a Killing vector to be orthogonal to a family of hypersurfaces. We say a one-form n is normal to a hypersurface Σ if $n(t) = 0$ for any tangent vector t to Σ. For example, df is normal to the hypersurface $f = 0$. Any other one-form n normal to Σ must have the form $n = g df + f n'$, and direct calculation gives

$$(n \wedge dn)|_\Sigma = n_{[a} \nabla_b n_{c]}|_\Sigma = 0$$

which is a useful identity. Conversely, a differential form version of Frobenius' theorem says that if n is a nonzero one-form so that $n \wedge dn = 0$ everywhere, then $n = g df$, so n is normal to surfaces of constant f, which foliate the spacetime.

Now suppose a Killing vector k^a satisfies this condition, where we use the metric to convert it to a one-form. Then we construct spatial coordinates on a hypersurface Σ of constant f and construct the t coordinate as above. By orthogonality the cross-terms in the metric vanish on Σ. Since the cross-terms are independent of t, they vanish everywhere.

We can show that the Schwarzschild metric is unique, assuming the spacetime is static and spherically symmetric.

- We take spherical symmetry to mean that we can choose coordinates (t, r, θ, ϕ) where θ and ϕ take the usual angle values, so that the last two coordinates contribute to the metric in the combination $d\Omega^2$, and t is as constructed above.

- More formally, this means that the spacetime has an $SO(3)$ subgroup of isometries whose orbits are 2-spheres; the angular variables θ and ϕ are coordinates on these spheres.
- We also assume there are no cross terms between r and θ and ϕ. Heuristically, this follows from demanding invariance under the 'parity transformation' $\theta \to \pi - \theta$, $\phi \to -\phi$. In this case, the most general possible metric is

  $$ds^2 = -e^{2\alpha(r)}dt^2 + e^{2\beta(r)}dr^2 + e^{2\gamma(r)}r^2 d\Omega^2$$

  where we use exponentials to fix the metric signature.
- We are free to rescale the radial coordinate so that

  $$ds^2 = -e^{2\alpha(r)}dt^2 + e^{2\beta(r)}dr^2 + r^2 d\Omega^2$$

  where we have redefined r and $\beta(r)$. The new coordinate r is defined so that the area of a sphere with coordinate r is $4\pi r^2$.
- Imposing the constraint $R_{\mu\nu}$, we eventually find the constraints

  $$\partial_r \alpha + \partial_r \beta = 0, \quad \partial_r(re^{2\alpha}) = 1.$$

  The solution to the first is $\alpha = -\beta + c$, and we can set $c = 0$ by rescaling the time coordinate. The second equation implies that $e^{2\alpha} = 1 - 2GM/r$, so

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

- Since the Schwarzschild metric reduces to the Minkowski metric for $r \gg 2M$, we can interpret t as the time measured by an observer at spatial infinity. In this limit, the Newtonian limit says that the potential is $-GM/r$, so we interpret M as the mass. A more formal approach is to calculate the ADM mass, which also turns out to be M.
- The Schwarzschild metric is singular at $r = 0$ and $r = 2M$, but the latter is just an artifact of the coordinates; we know this because we can choose coordinates without singularities there. A physical singularity is associated with singularities in curvature scalars. Since

  $$R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma} = \frac{48M^2}{r^6}$$

  we see that $r = 0$ is a physical singularity. To make sure it's physically relevant, we must also check that it can be reached by traveling a finite distance along a curve.

Birkhoff's theorem states that the Schwarzschild metric is the unique vacuum solution to the Einstein equations with spherical symmetry. We'll give a sketch of this proof below.

- First, we need to define spherical symmetry. It's not suitable to define spherical symmetry relative to an origin, because the space $\mathbb{R} \times S^2$ has no identifiable origin, but clearly has spherical symmetry. Instead, we suppose there are three Killing vectors R, S, and T with the $SO(3)$ commutation relations

  $$[R, S] = T, \quad [S, T] = R, \quad [T, R] = S.$$

  By Frobenius' theorem, this provides a foliation of spacetime by 2-spheres. This can only fail at isolated points, such as the origin at $\mathbb{R}^3$, since the Killing vectors vanish there.

- Next, we define angular coordinates (θ, ϕ) on each sphere, and define coordinates (a, b) on the set of spheres so that a sphere has constant a and b. Then the metric on each sphere is $f(a,b)d\Omega^2$. Moreover, no metric coefficients can depend on θ and ϕ, except for the $\sin^2\theta$ in $d\Omega^2$, by spherical symmetry.

- However, there may be cross terms such as $dad\theta$, since we haven't lined up the angular coordinates on the spheres with each other; moving perpendicular to a sphere may change θ and ϕ. Instead, we define the coordinate system by choosing coordinates (θ, ϕ) on one sphere, then considering the set of geodesics perpendicular to this sphere. A geodesic that goes through (θ_0, ϕ_0) on our original sphere defines points with angular coordinates (θ_0, ϕ_0) on nearby spheres, ensuring there are no cross terms.

- Now, the angular part of the metric is $r^2(a,b)d\Omega^2$ and we replace the coordinate b with r for

  $$ds^2 = g_{aa}(a,r)da^2 + g_{ar}(a,r)(dadr + drda) + g_{rr}(a,r)dr^2 + r^2 d\Omega^2.$$

  We claim that we can replace a with $t(a,r)$ so that there is no $drdt$ cross term, giving

  $$ds^2 = m(t,r)dt^2 + n(t,r)dr^2 + r^2 d\Omega^2.$$

  This is generically possible because we start with three degrees of freedom, g_{aa}, g_{ar}, and g_{rr}, and end with three, t, m, and n. To do it explicitly, we use integrating factors.

- We then make the assumption that m is negative and n is positive, giving the parametrization

  $$ds^2 = -e^{2\alpha(t,r)}dt^2 + e^{2\beta(t,r)}dr^2 + r^2 d\Omega^2.$$

  This is the same as what we had before, but α and β may depend on t.

- Einstein's equations show that

  $$\partial_t \beta = 0, \quad \partial_t \partial_r \alpha = 0$$

  which implies that

  $$\beta = \beta(r), \quad \alpha = f(r) + g(t).$$

  We can rescale the time-coordinate so that $g(t) = 0$, getting us back to where we were before.

- Remarkably, we've shown that spherical symmetry implies a unique solution, which must be static. This is in some sense a generalization of the shell theorem. For example, the metric outside a spherically symmetric star must be Schwarzschild no matter how it evolves; for example, this implies it emits no gravitational waves during its collapse.

6.2 Spherical Stars

We now apply our general relativity to spherically symmetric stars. First, we review astrophysics.

- Stars are supported against gravity by the pressure generated by nuclear fusion. When the fuel for these reactions runs out, the star collapses.

- For smaller stars, the final state is a white dwarf, where gravity is balanced by electron degeneracy pressure. A white dwarf cannot have a mass greater than the Chandrasekhar limit $1.4M_\odot$, derived with Newtonian gravity.

- Above this limit, the final state can be a neutron star, supported by neutron degeneracy pressure; the protons are removed by inverse beta decay. Neutron stars are extremely small, with Newtonian potentials $|\Phi| \sim 0.1$, so general relativity is required to describe them.
- Neutron stars cannot exist with a mass above the TOV limit, $3M_\odot$, and we will heuristically derive this result below. The outside of a spherical star is described by the Schwarzschild metric, while we model the inside by a perfect fluid.

Now we turn to the metric inside a spherical star.

- From our earlier work, we know we can set the metric to
$$ds^2 = -e^{2\Phi(r)}dt^2 + e^{2\Psi(r)}dr^2 + r^2 d\Omega^2.$$
The matter is a perfect fluid,
$$T_{ab} = (\rho + p)u_a u_b + p g_{ab}$$
where u^a is the four-velocity of the fluid, and since the situation is static and spherically symmetric the four-velocity must point in the timelike direction, $u^\mu = e^{-\Phi}(\partial_t)^\mu$. Here, p and ρ are functions of r that vanish outside the star, $r > R$.
- The equations of motion for the fluid follow from conservation of T^{ab}, but this follows from the Einstein equation. By symmetry, the only independent components of this equation are the tt, rr, and $\theta\theta$ components.
- We define $m(r)$ by
$$e^{2\Psi(r)} = \left(1 - \frac{2m(r)}{r}\right)^{-1}$$
so that in the Newtonian limit, $m(r)$ corresponds roughly to the mass within radius r. The tt and rr components of the Einstein equation are
$$\frac{dm}{dr} = 4\pi r^2 \rho, \quad \frac{d\Phi}{dr} = \frac{m + 4\pi r^3 p}{r(r-2m)}.$$
Finally, the $\theta\theta$ component is more easily derived as the r component of $\nabla_\mu T^{\mu\nu} = 0$,
$$\frac{dp}{dr} = -(p+\rho)\frac{m + 4\pi r^3 p}{r(r-2m)}.$$
This is essentially force balance. These equations are collectively called the TOV equations.
- An equation of state relates T, p, and ρ. In the zero-temperature limit, this gives a barotropic equation of state $p = p(\rho)$, where we assume $dp/d\rho > 0$, giving four equations for the four unknowns m, p, ρ, and Φ.
- In the case of no matter, we have $p = \rho = 0$ and $m(r) = M$, implying
$$\Phi = \frac{1}{2}\log(1 - 2M/r) + \Phi_0.$$
Rescaling the time to set Φ_0 to zero recovers the Schwarzschild solution. Since the solution outside a star is always Schwarzschild, a star must have radius $R > 2M$.

Next, we integrate these equations.

- Integrating the equation for dm/dr gives
$$m(r) = 4\pi \int_0^4 \rho(r')r'^2 \, dr' + m_*.$$
The integration constant m_* must be zero, because spacetime is locally flat; very small spheres near the origin should have the same area/radius relation as in flat space. Then $\Psi(0) = 0$ which implies $m_* = 0$.

- To match onto the Schwarzschild solution, we must have
$$M = 4\pi \int_0^R \rho(r)r^2 \, dr.$$
This looks deceptively like the Newtonian formula for total mass but it has the wrong volume element. The total mass-energy is actually
$$E = 4\pi \int_0^R \rho e^{\Psi} r^2 \, dr > M.$$
We interpret M as the total energy, and $E - M$ as the gravitational binding energy.

- We can improve our bound on the radius of a star. Since we must have $m(r)/r < 1/2$ for all r, and $dp/dr \leq 0$ which implies $d\rho/dr \leq 0$, it can be shown that
$$\frac{m(r)}{r} \leq \frac{2}{9}\left(1 - 6\pi r^2 p(r) + \sqrt{1 + 6\pi r^2 p(r)}\right)$$
and evaluating at $r = R$ and $p = 0$, we have the Buchdahl inequality $R > 9M/4$.

- In general, to solve the equations numerically, we fix $p(\rho)$ and regard the dm/dr and dp/dr equations as coupled first-order differential equations with initial condition $m(0) = 0$ and $\rho(0) = \rho_c$. We then integrate again to find $\Phi(r)$.

- For a typical equation of state, the result is that $M(\rho_c)$ has a maximum, implying a maximum possible mass. For example, using the equation of state for a white dwarf reproduces the Chandrasekhar bound.

- Remarkably, we can still find an upper bound for an arbitrary equation of state. Define the core of the star to be the region $r < r_0$ where we don't know the equation of state, and let $m_0 = m(r_0)$. Since $d\rho/dr < 0$,
$$m_0 \geq \frac{4}{3}\pi r_0^3 \rho_0.$$
On the other hand, the Buchdahl inequality also holds for the core,
$$\frac{m_0}{r_0} \leq (\text{RHS at } p = p_0) \leq (\text{RHS at } p = 0) = \frac{4}{9}.$$
These two inequalities define a finite region in the (m_0, r_0) plane with bound
$$m_0 < \sqrt{\frac{16}{243\pi\rho_0}}.$$
The mass of the envelope outside the core can be determined with a known equation of state and typically contributes less than 1%.

- We typically set $\rho_0 = 5\times 10^{14}\,\mathrm{g/cm^3}$, the density of nuclear matter. Then numerically optimizing over the combined core and envelope mass, the mass of the star is at most $5M_\odot$. This bound can be strengthened with further physical assumptions. For example, the speed of sound is $\sqrt{dp/d\rho}$, so requiring $dp/d\rho \leq 1$ gives a bound of $3M_\odot$.
- The physical intuition behind these bounds is that the gravity due to mass density can be resisted by pressure, but in general relativity, pressure is also a source of gravity. Thus, at some point collapse will occur no matter what the pressure is.

6.3 Geodesics of Schwarzschild

In this section, we find the geodesics and cover some classic tests of general relativity.

- Rather than use the geodesic equation, we work directly with the Lagrangian,
$$\mathcal{L} = -\left(1 - \frac{2M}{r}\right)\dot{t}^2 + \left(1 - \frac{2M}{r}\right)^{-1}\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\,\dot{\phi}^2$$
where the dot is a derivative with respect to the parameter λ. The θ component of the Euler-Lagrange equation is
$$\ddot{\theta} + 2\frac{\dot{r}\dot{\theta}}{r} - \sin\theta\cos\theta\,\dot{\phi}^2 = 0$$
which shows that if we begin with $\theta = \pi/2$ and $\dot{\theta} = 0$, then $\dot{\theta} = 0$ at all times. Hence we can choose our coordinate system so that $\theta = \pi/2$ without loss of generality.
- There are also conserved quantities due to the cyclic coordinates t and ϕ, with
$$\frac{\partial\mathcal{L}}{\partial\dot{t}} = -2\left(1 - \frac{2M}{r}\right)\dot{t}, \quad \frac{\partial\mathcal{L}}{\partial\dot{\phi}} = 2r^2\sin^2\theta\dot{\phi} = 2r^2\dot{\phi}.$$
There is also the analogue of 'time translation' symmetry, since $\partial\mathcal{L}/\partial\lambda = 0$. This yields conservation of the 'Hamiltonian', but since $\mathcal{L}$ is a homogeneous polynomial in the velocities $\dot{x}^\mu$, it is equivalent to conservation of the Lagrangian, $d\mathcal{L}/d\lambda = 0$.
- This leads to the conserved quantities
$$E = \left(1 - \frac{2M}{r}\right)\dot{t}, \quad L = r^2\dot{\phi}, \quad Q = -\left(1 - \frac{2M}{r}\right)\dot{t}^2 + \left(1 - \frac{2M}{r}\right)^{-1}\dot{r}^2 + r^2\dot{\phi}^2$$
and we scale λ so that $Q = -1$ on timelike paths and $Q = 0$ on null paths. Note that we can't change λ arbitrarily, as the non-square root form of the action we're using is not reparametrization invariant; only affine transformations are allowed.
- There are two other conserved quantities from the other two components of angular momentum; these specify the direction of angular momentum and imply the motion lies in a plane, which we used earlier to set $\theta = \pi/2$.
- Note that the quantity E includes "gravitational potential energy". The "kinetic energy" alone is given by $p^\mu u_\mu$, but this is not conserved. For $r \gg 2M$, the two match up, as $E = \dot{t} = \gamma$. Then E and L can be interpreted as the energy and angular momentum per unit mass.

Now we use our setup to investigate the orbits.

- Plugging everything into our equation for Q gives

$$\frac{1}{2}\dot{r}^2 + V(r) = \frac{1}{2}E^2, \quad V(r) = \frac{1}{2}\left(1 - \frac{2M}{r}\right)\left(\frac{L^2}{r^2} - Q\right).$$

  The E^2 on the right-hand side looks a bit strange, but when we take the Newtonian limit, we will have $E \approx mc^2 + mv^2/2$, and $E^2/2$ has quadratic term proportional to $mv^2/2$ as expected.

- To analyze the geodesics, it is more convenient to write

$$\frac{1}{2}\dot{r}^2 + V_{\text{GR}}(r) = \text{const}, \quad V_{\text{GR}}(r) = \frac{1}{2}\frac{L^2}{r^2} + Q\frac{GM}{r} - \frac{GML^2}{r^3}$$

  where we restored factors of G. By contrast, in the Newtonian case we have

$$V_{\text{N}}(r) = \frac{1}{2}\frac{L^2}{r^2} - \frac{GM}{r}.$$

  That is, the $1/r^3$ term is missing, and $Q = -1$ is fixed.

- First, consider a massless particle. This is a bit ambiguous; if we take the Newtonian limit first, the massless limit does nothing, since the mass cancels out everywhere. On the other hand, if we set $Q = 0$ first in general relativity, then take the Newtonian limit, the potential is just $L^2/2r^2$, that of a free particle.

- For $Q = -1$, in the Newtonian limit, all circular orbits are stable. In the case of general relativity, we find two circular orbits for each value of L,

$$r = \frac{L^2}{2M} \pm \sqrt{\frac{L^4}{4M^2} - 3L^2}$$

  where the outer one is stable and the inner is unstable. For $L^2 = 12M^2$ these orbits merge at $r = 6M$. Then there are no stable orbits for $r < 6GM$. As L is varied, we find unstable orbits for $3GM < r < 6GM$.

- We can also handle $Q = 0$ in general relativity. In this case, there is only one initial condition, the direction of the light ray at infinity, and accordingly E^2 and L^2 combine into one parameter, since reparametrization absorbs the second. We find a single unstable circular orbit at $r = 3M$, called the light ring or photon sphere.

We now study the precession of elliptical orbits in detail. **(todo: Shapiro delay)**

- It is useful to parametrize the orbit as $r(\phi)$. Since an ellipse has the form $r \propto 1/(1 + \epsilon\cos\phi)$, we expect the equation will be simpler if we use the dimensionless variable $x = L^2/Mr$. Switching to x and differentiating with respect to ϕ to remove the constant energy term gives

$$\frac{d^2x}{d\phi^2} - 1 + x = \alpha x^2, \quad \alpha = \frac{3M^2}{L^2}.$$

  The term on the right is the contribution from general relativity.

- We expand in a perturbation series in α, letting $x = x_0 + x_1 + \ldots$. At zeroth order,

$$\frac{d^2 x_0}{d\phi^2} - 1 + x_0 = 0, \quad x_0 = 1 + e\cos\phi$$

which recovers the elliptical orbit. The first-order equation is

$$\frac{d^2 x_1}{d\phi^2} + x_1 = \alpha x_0^2 = \alpha\left[\left(1 + \frac{e^2}{2}\right) + 2e\cos\phi + \frac{e^2}{2}\cos 2\phi\right].$$

- Now, $x_1(\phi)$ satisfies the same equation as a driven harmonic oscillator. The constant driving term just gives a constant shift, and the $\cos 2\phi$ term yields an oscillation proportional to $\cos 2\phi$. However, the $\cos\phi$ term is resonant, so the solution is $x_1 \sim \phi\cos\phi$. Since we are interested in perihelion shift, we keep only this term, for

$$x \approx 1 + e\cos\phi + \alpha e\phi\sin\phi \approx 1 + e\cos[(1-\alpha)\phi].$$

The last step drops $O(\alpha^2)$ terms, which is fine since we're working to leading order in α.

- Therefore, we find that per orbit, the perihelion precesses by

$$\Delta\phi = 2\pi\alpha = \frac{6\pi M^2}{L^2}.$$

Using the Newtonian equation $L^2 = GM(1-e^2)a$, correct to leading order in α, we have

$$\Delta\phi = \frac{6\pi GM}{c^2(1-e^2)a}$$

where we restored units. For Mercury, this is 43 arcseconds per century, matching the experimental result. Note that this required sophisticated developments in celestial mechanics, because the true precession is over 10 times larger, with most of it accounted for by the influences of the other planets; the relativistic perihelion shift is just the remainder.

As a second application, we consider the deflection of light.

- In the Newtonian picture where light does deflect, it obeys the equation

$$\frac{dx^2}{d\phi^2} + x = 1.$$

A general solution has the form

$$x = 1 + a\sin\phi, \quad a = \frac{L^2}{Mb} = \frac{b}{M}$$

where b is the distance of closest approach. Note that since L is the angular momentum per unit mass, we have $L = bc = b$ since we've set $c = 1$.

- As a tends to infinity, the path becomes a straight line, so that the velocities at infinity are related by an angle π. Generally, the angle is the difference of the zeroes of $x(\phi)$. For finite a, the angle is shifted by $\Delta\phi = 2/a = 2M/b$.

- In general relativity, the equation of motion is instead

$$\frac{dx^2}{d\phi^2} + x = \frac{3M^2}{L^2}x^2 = \frac{3x^2}{a^2}.$$

  We expand order by order in $1/a$, where

$$x_0 = a\sin\phi, \quad \frac{dx_1^2}{d\phi^2} + x_1 = 3\sin^2\phi.$$

  Then the solution for x_1 is

$$x_1 = \frac{3}{2} + \frac{1}{2}\cos(2\phi) + \text{homogeneous solution}.$$

- It would be best to choose the homogeneous solution so that $x_0 + x_1$ has the same distance of closest of approach as x_0, but it doesn't matter since it only results in higher-order corrections. Instead, we simply set it to zero, and compute the angle shift

$$\Delta\phi = \frac{4}{a} = \frac{4M}{b}$$

  where b is the distance of closest approach for x_0 alone; adjusting b gives a second-order correction. We find an angle shift that is twice as large as in the Newtonian case, as famously observed by Eddington during the solar eclipse of 1919. In some sense, this is because the Newtonian case only accounts for the shift to g_{tt}, and not the equal shift to g_{rr}.

Example. Gravitational redshift. Suppose that observers A and B at $r = r_A$ and $r = r_B$ exchange signals. If A sends two photons separated by a coordinate time Δt, then they arrive at B separated by a coordinate time Δt since $\partial/\partial t$ is a Killing vector. Then the proper times are related by

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

Now for light waves, the period is equal to the wavelength, so the wavelength redshifts by

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

in the case where B is at infinity. This diverges when $r_A \to 2M$. In addition, by the Buchdahl inequality, the maximum redshift observable from the surface of a spherical star is $z = 2$.

As stated earlier, for weak fields, the gravitational redshift is by $\Delta\phi/c^2$. The gravitational redshift has been tested and confirmed in various circumstances, such as in Earth orbit by Gravity Probe-A ($\Delta\phi/c^2 \sim 0.0002$), and with stars approaching the supermassive black hole at the center of the galaxy ($\Delta\phi/c^2 \sim 0.1$).

Note. Detection methods for black holes. Due to the Chandrasekhar and TOV limits, a very small and very massive object must be a black hole. For example, stars are observed to be orbiting about the galactic center, which we can infer has a mass of $4 \times 10^6 M_\odot$. The stars also get close enough to bound the radius, ruling out anything besides a black hole. Many galaxies are believed to contain a supermassive black hole at their centers, i.e. a black hole with mass over $10^6 M_\odot$. Usually, these black holes contain about 0.1% of the mass of the galaxy; it is unclear how they form.

Another way to detect black holes is by their accretion disks. As a particle orbits a black hole, it slowly loses energy, decreasing its orbit radius, until it hits $r = 6M$, at which point it quickly falls in. It can be shown that this process releases $1 - \sqrt{8/9} \approx 0.06$ of the rest mass as energy, typically as X-rays. Such a signal has a characteristic cutoff corresponding to $r = 6M$. The accretion disks around supermassive black holes power quasars; accretion disks can also form around smaller black holes in binary systems, by stripping mass off their companion by tidal forces.

6.4 Schwarzschild Black Holes

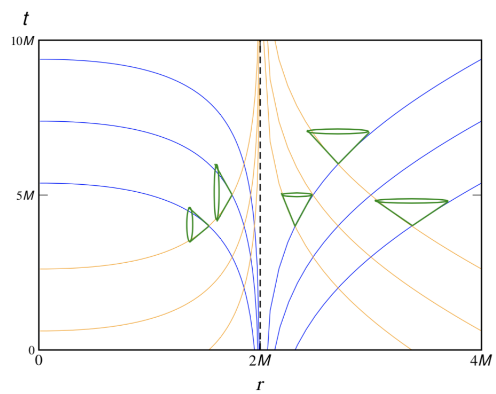
We now investigate the event horizon of the Schwarzschild solution.

- We focus on radially moving light, where setting $ds^2 = 0$ gives

$$\frac{dt}{dr} = \pm \frac{r}{r - 2M}, \qquad t(r) = \pm(r + 2M \log|r - 2M|) + k$$

  where k is a real constant. We take the absolute value in the logarithm to keep everything real, though we can still get a valid solution without it by letting k be complex.

- When $r > 2M$, the $+$ sign gives outgoing geodesics and the $-$ sign gives ingoing geodesics. The ingoing geodesics take an infinite amount of coordinate time to hit $r = 2M$.



  When $r < 2M$, the situation is more subtle. Now r is timelike, and it is ambiguous whether the light cone points towards decreasing or increasing r. We can't resolve this by continuity across $r = 2M$ since the metric is singular there; more rigorously we really shouldn't consider the $r < 2M$ region to be covered by the coordinates at all because of the coordinate singularity. It turns out there are *two* $r < 2M$ regions, one where the light cone points in and one where it points out, as we'll show later.

- Next, we consider the perspective of an infalling observer. Energy conservation gives

$$\left(1 - \frac{2M}{r}\right)\dot{t} = E$$

  and we parametrize by proper time so that

$$-\left(1 - \frac{2M}{r}\right)\dot{t}^2 + \left(1 - \frac{2M}{r}\right)^{-1}\dot{r}^2 = -1.$$

For simplicity we suppose the observer is at rest at infinity, so $E = 1$. Rearranging,

$$\dot{r}^2 = \frac{2M}{r}, \quad \Delta\tau = -\frac{2}{3\sqrt{2M}}\Delta(r^{3/2}).$$

Then a falling observer takes a finite proper time to fall through the event horizon, with nothing special happening when they cross it.

- On the other hand, we can parametrize the geodesic by t for

$$\frac{dt}{dr} = \frac{\dot{t}}{\dot{r}} = -\sqrt{\frac{r}{2M}}\left(1 - \frac{2M}{r}\right)^{-1}.$$

The solution is complicated, but it takes infinite coordinate time to fall through $r = 2M$. That is, a distance observer only sees the infalling one slowly redshift more and more as they approach the horizon, never quite crossing it. However, an observer falling into a black hole doesn't 'see the end of the universe'. A detailed calculation shows that they can only receive a finite number of evenly-spaced light signals from outside.

To better understand the behavior near the event horizon, we switch to an improved coordinate system adapted to null geodesics.

- In the incoming Eddington-Finkelstein (EF) coordinates, we define

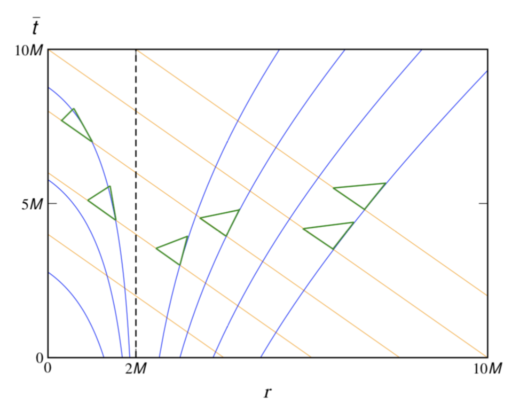
$$\bar{t} = t + 2M \log|r - 2M|, \quad d\bar{t} = dt + \frac{2M}{r - 2M}dr$$

so that the Schwarzschild metric becomes

$$ds^2 = -\left(1 - \frac{2M}{r}\right)d\bar{t}^2 + \frac{4M}{r}d\bar{t}dr + \left(1 + \frac{2M}{r}\right)dr^2 + r^2 d\Omega^2.$$

These coordinates are chosen so that ingoing radial null geodesics are simple; they are

$$\bar{t} = \begin{cases} -r + k & \text{incoming} \\ r + 4M\log|r - 2M| + k & \text{outgoing.} \end{cases}$$



- The light cones now vary smoothly across $r = 2M$, and 'tip over' at the event horizon, so that particles can only fall inward. It's now clear that the event horizon itself is a null surface; a photon could travel on it forever.

- Formally, we define an event horizon to be the outermost boundary of a region of spacetime from which no null geodesics, and hence no timelike curves can escape to infinity. Israel's theorem states that the Schwarzschild spacetime is the unique static, asymptotically flat spacetime with a regular horizon.

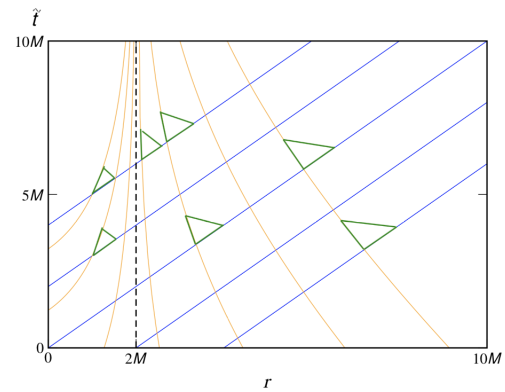
- Mathematically, we have extended the original Schwarzschild solution, i.e. found a larger spacetime with metric which contains the $r > 2M$ region of the Schwarzschild solution as a subset. The Einstein field equation still holds everywhere, because it holds for $r > 2M$ and the metric is real analytic.

- Similarly, we can adapt our coordinate system to outgoing null geodesics, defining the outgoing EF coordinates, with metric

$$ds^2 = -\left(1 - \frac{2M}{r}\right) d\tilde{t}^2 - \frac{4M}{r} d\tilde{t} dr + \left(1 + \frac{2M}{r}\right) dr^2 + r^2 d\Omega^2.$$

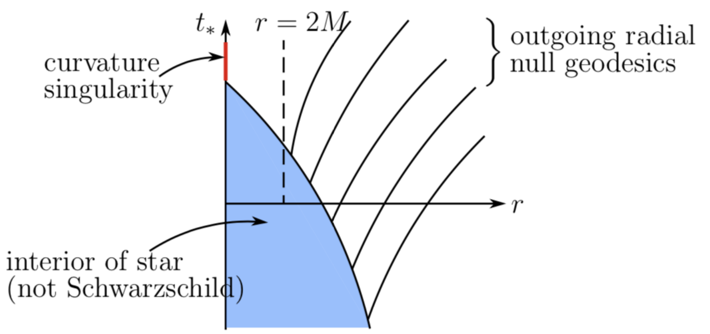
  The radial null geodesics are

$$\tilde{t} = t - 2M \log |r - 2M|, \quad \tilde{t} = \begin{cases} -r - 4M \log |r - 2M| + k & \text{incoming} \\ r + k & \text{outgoing.} \end{cases}$$



- In this case, the physical picture is exactly the opposite: geodesics can only ever come out of $r = 2M$. This is still a valid extension of the Schwarzschild spacetime, but it's not the same one as the incoming EF coordinates.

Note. The (incoming) Finkelstein diagram for a collapsing star is shown below.



The metric everywhere outside the star is Schwarzschild, so to an external observer it is exactly Schwarzschild once the outside passes $r = 2M$, which occurs in finite time. It can be shown that a particle on the outside must then hit $r = 0$ within proper time πM, i.e. the singularity forms in finite proper time. In the original Schwarzschild coordinates, the star never finishes collapsing; instead it makes an increasingly thin and redshifted shell at $r = 2M$ that quickly becomes invisible. Of course, this makes no difference to an observer actually falling in.

6.5 Kruskal Coordinates

Now we switch to the Kruskal–Szekeres coordinates, which yield the maximal extension of the Schwarzschild spacetime.

- For the incoming EF coordinates, we transform to the null coordinate

$$v = \bar{t} + r, \quad ds^2 = -\left(1 - \frac{2M}{r}\right) dv^2 + 2drdv + r^2 d\Omega^2.$$

  Similarly, for the outgoing EF coordinates, we define

$$u = \tilde{t} - r, \quad ds^2 = -\left(1 - \frac{2M}{r}\right) du^2 - 2dudr + r^2 d\Omega^2.$$

- Next, we switch to the variables u and v, which obey

$$\frac{1}{2}(v + u) = t, \quad \frac{1}{2}(v - u) = r + 2M \log \frac{r - 2M}{r_*}$$

  where we've absorbed the integration constant into r_*. The Schwarzschild metric becomes

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

  where r is implicitly defined in terms of u and v.

- Next, we change variables to an exponential version of u and v,

$$V = e^{v/4M}, \quad U = -e^{-u/4M}$$

  so that

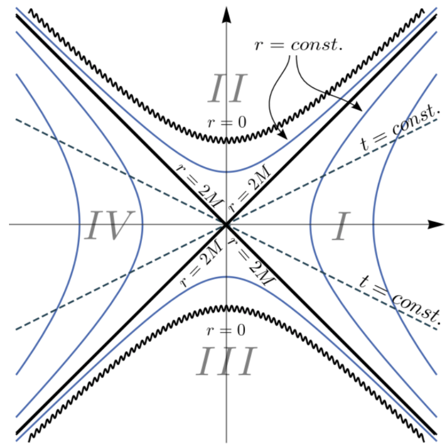
$$ds^2 = \frac{16M^2}{UV}\left(1 - \frac{2M}{r}\right) dUdV + r^2 d\Omega^2.$$

  To simplify, we note that

$$UV = -e^{v-u}4M = -\frac{r - 2M}{r_*} e^{r/2M}, \quad ds^2 = -\frac{16M^2}{r/r_*} e^{-r/2M} dUdV + r^2 d\Omega^2.$$

  The original spacetime only contained $U \geq 0$ and $V \leq 0$, but we can now extend to all U and V, with $r(U, V)$ defined the same way.

- Now, U and V are null coordinates, so we can switch back to spatial and radial coordinates by
$$\hat{t} = \frac{1}{2}(V+U), \quad \hat{r} = \frac{1}{2}(V-U).$$
Then we have $dV\,dU = d\hat{t}^2 - d\hat{r}^2$, so
$$ds^2 = \frac{16M^2}{r} e^{-r/2M}(-d\hat{t}^2 + d\hat{r}^2) + r^2 d\Omega^2$$
where we set $r_* = 1$ by rescaling, though sometimes $r_* = 2M$ is also chosen.



- The metric is now manifestly regular at $r = 2M$, and radial null geodesics have the simple form $\hat{t} = \pm\hat{r} + k$. To relate $\hat{r}$ and $\hat{t}$ back to r, note that
$$\hat{t}^2 - \hat{r}^2 = -(r-2M)e^{r/2M}.$$
Then $r = 2M$ is given by $\hat{t} = \pm\hat{r}$, while $r = 0$ is given by $\hat{t} = \pm\sqrt{\hat{r}^2 + 2M}$. There are no restrictions on our coordinates besides $r > 0$.

- We can now understand some of the puzzles we ran into earlier. The Schwarzschild coordinates only work in region I, running into singularities at $r = 2M$. The incoming geodesics take an infinite amount of time to fall into $r = 2M$, and the outgoing geodesics take an infinite amount of time to come out.

- The incoming EF coordinates work in regions I and II, extending through $t \to \infty$, so that they contain the entirety of incoming geodesics. The outgoing EF coordinates instead extend through $t \to -\infty$, so they contain the entirety of outgoing geodesics.

- The point is that it's perfectly possible to come out of the surface $r = 2M$, though we can never observe it since it takes an infinite coordinate time. But this isn't that strange, because we can't observe anything falling in either.

- We can think of region III as the image of region II under time reversal; in this region there is a 'white hole' from which things can only exit. One might think that a black hole must map to a black hole under time reversal, since the Schwarzschild metric is static, but that's not right because t is spacelike inside the hole. The $r = 0$ singularity of the black hole is a *time*, not a place, lying in the infinite future; the $r = 0$ singularity of the white hole lies in the infinite past. Both describe gravitationally attractive objects of mass M. They behave somewhat analogously to a 'Big Bang' and 'Big Crunch'.

- Finally, region IV is a mirror image of region I. We can think of it as being briefly connected to region I by a wormhole, as can be shown by taking slices of constant T, which closes too fast for any timelike observer to pass through.

Note. Why don't we see white holes in reality? All of this discussion only applies to the 'eternal' black hole of the Schwarzschild metric. In a real black hole formed by star collapse, there is no region III, so the white hole isn't physical. A deeper reason is from thermodynamics. We expect that a black hole is stable, i.e. that small perturbations decay. Then small perturbations of a white hole grow, so it is thermodynamically impossible to create them.

We give a more formal definition of 'black hole' and 'event horizon'.

- A vector is causal if it is timelike or null, where we stipulate that a null vector must be nonzero. A curve is causal if its tangent vector is everywhere causal. Note that a causal curve traveled backwards is also causal.
- A spacetime is time-orientable if it admits a time-orientation, i.e. a global causal vector field T^a. Another causal vector X^a is future-directed if it lies on the same light cone as T^a and past-directed otherwise. Because of the $(-+++)$ metric convention, if T^a and X^a have negative inner product, they are in the same light cone.
- It is most convenient to use the null incoming EF coordinates (v, r, Ω) defined above, as ∂_r is null everywhere. Note that ∂_r in the original EF coordinates is not null, because ∂_r is defined as an element of the dual basis of the dx^μ, where the x^μ are all the coordinates, so changing t to v changes ∂_r.
- At infinity, we choose positive time to point along ∂_t, and this is parallel to ∂_v. Then $\partial_v \cdot \partial_r$ is positive at infinity, which means our time-orientation is $-\partial_r$.
- We can now use this setup to show rigorously that if $x^\mu(\lambda)$ is a future-directed causal curve and $r(\lambda_0) \leq 2M$, then $r(\lambda) \leq 2M$ for $\lambda \geq \lambda_0$. The tangent vector $V^\mu = dx^\mu/d\lambda$ satisfies
$$0 \geq -\partial_r \cdot V = -g_{r\mu} V^\mu = -V^v = -\frac{dv}{d\lambda}.$$
Next, rearranging the norm of V^2 gives
$$-2\frac{dv}{d\lambda}\frac{dr}{d\lambda} = -V^2 + \left(\frac{2M}{r} - 1\right)\left(\frac{dv}{dr}\right)^2 + r^2\left(\frac{d\Omega}{d\lambda}\right)^2$$
where the last term stands in for the angular parts. For $r \leq 2M$, we have $(dv/d\lambda)(dr/d\lambda) \leq 0$, which essentially gives the result.
- There are a few more annoying cases. For example, we could have $dr/d\lambda > 0$ if $dv/d\lambda = 0$. In that case, we need $V^2 = d\Omega/d\lambda = 0$. But then only V^r is nonzero, and V^r is a negative multiple of $-\partial_r$, so it is not in the same light cone. There's a similarly annoying case when $r = 2M$ exactly.
- This establishes that it is impossible to send a signal from $r \leq 2M$ to infinity. We define a region of spacetime where this is true to be a black hole, and the boundary of a black hole to be an event horizon.

- As another application, the incoming and outgoing EF coordinates only differ by the sign of the time orientation; this is formally the statement that they are time reverses of each other.

Now we give a few more details about the Kruskal coordinates.

- The time translation vector field is
$$k = \frac{1}{4M}\left(V\frac{\partial}{\partial V} - U\frac{\partial}{\partial U}\right), \quad k^2 = -\left(1 - \frac{2M}{r}\right)$$
and it is timelike in regions I and IV, and spacelike in regions II and III.

- There is a 'wormhole' between regions I and IV. We define the isotropic coordinate ρ by
$$r = \rho + M + \frac{M^2}{4\rho}$$
so that for a fixed $r > 2M$, there are two solutions for ρ. We choose $\rho > M/2$ for region I and $0 < \rho < M/2$ for region IV. Then the metric in coordinates (t, ρ, θ, ϕ) is
$$ds^2 = -\frac{(1 - M/2\rho)^2}{(1 + M/2\rho)^2}dt^2 + \left(1 + \frac{M}{2\rho}\right)^4 (d\rho^2 + \rho^2 d\Omega^2).$$
The resulting spacetime is symmetric between regions I and IV by the isometry $\rho \to M^2/4\rho$. The metric is singular at $\rho = M/2$, but this is just a coordinate singularity.

- These coordinates are called isotropic coordinates because for fixed t, the metric is Euclidean up to a local scale factor. The metric
$$ds^2 = \left(1 + \frac{M}{2\rho}\right)^4 (d\rho^2 + \rho^2 d\Omega^2)$$
describes a Riemannian 3-manifold with topology $\mathbb{R} \times S^2$ called an Einstein–Rosen bridge.

- A wormhole connects the asymptotically flat regions $\rho \to \infty$ and $\rho \to 0$ by a 'throat' of minimum radius $2M$ at $\rho = M/2$. We may visualize the wormhole by embedding it in Euclidean $\mathbb{R}^4$, straightforward in isotropic coordinates, and suppressing an angular coordinate. The wormhole closes too fast to be traversed, as seen from the Kruskal diagram.

Finally, we take a careful look at singularities.

- A spacetime is extendable if it is isometric to a proper subset of another spacetime; we have seen that the Schwarzschild spacetime is extendable to the Kruskal spacetime, which is not extendable.

- We have defined physical singularities as points where a curvature scalar diverges, but this is not general enough. For example, consider the conical space
$$ds^2 = dr^2 + \lambda^2 r^2 d\phi^2$$
where $\lambda > 0$ is not equal to one. Then the curvature vanishes everywhere, but the point $r = 0$ is not locally isomorphic to Euclidean space; circles don't have the right circumferences, no matter how small we make them. This is called a conical singularity.

- Mathematically, we don't want to define singularities as points where curvature scalars diverge because we are working with smooth manifolds with smooth metrics; the singularities aren't regarded as points in the manifold at all. Instead we detect singularities through geodesics; they are places where geodesics end in finite time. To rule out cases where we just haven't made the parameter space large enough, we define inextendability.
- We say $p \in M$ is a future endpoint of a future-directed causal curve $\gamma\colon (a,b) \to M$ if, for any neighborhood O of p, there exists t_0 so that $\gamma(t) \in O$ for all $t > t_0$. We say γ is future-inextendable if it has no future endpoint.
- For example, in Minkowski spacetime, consider $\gamma\colon (-\infty, 0) \to M$ where $\gamma(t) = (t, 0, 0, 0)$. Then γ has a future endpoint, the origin, so it is not future-inextendable. However, if the origin is deleted, then γ is future-inextendable.
- A geodesic is complete if an affine parameter for the geodesic extends to $\pm\infty$, and a spacetime is geodesically complete if all inextendable causal geodesics are complete.
- For example, the Schwarzschild spacetime is not geodesically complete, because of geodesics that go through the horizon; here the incompleteness arises because we are not considering the entire spacetime. We define a spacetime to be singular if it is geodesically incomplete and inextendable, so the Kruskal spacetime is singular.

7 The Penrose Singularity Theorem

7.1 The Initial Value Problem

In this section, we outline the proof of the Penrose singularity theorem, which states that singularities are 'generic' in general relativity. To begin, we describe the initial value problem in general relativity.

- Let (M, g) be a time-orientable spacetime. A partial Cauchy surface Σ is a hypersurface for which no two points are connected by a causal curve. The future domain of dependence of Σ, denoted $D^+(\Sigma)$, is the set of points so that every past-inextendible causal curve through p intersects Σ, and the past domain of dependence $D^-(\Sigma)$ is defined similarly. Their union is the domain of dependence $D(\Sigma)$. The boundaries of $D^\pm(\Sigma)$, if they exist, are called future/past Cauchy horizons.

- The domain of dependence is the region of spacetime where one can determine what happens from data specified on Σ. For example, any causal geodesic in $D(\Sigma)$ must intersect Σ at a unique point; then the geodesic is specified by its velocity at that point.

- More generally, a hyperbolic PDE is one of the form

  $$g^{ef}\nabla_e\nabla_f T^{(i)ab\ldots}{}_{cd\ldots} = \text{linear in } T^{(i)} \text{ and its first derivatives}$$

  for a set of tensor fields $T^{(i)}$, and the right-hand side can depend on the metric and its derivatives in an arbitrary way. The Klein–Gordan/wave equation takes this form, as well as the Maxwell equations in Lorenz gauge. Then one can show that specifying initial data $T^{(i)}$, $\partial_t T^{(i)}$ on Σ specifies the $T^{(i)}$ on all of $D(\Sigma)$.

- A spacetime (M, g) is globally hyperbolic if it admits a Cauchy surface, i.e. a partial Cauchy surface Σ so that $M = D(\Sigma)$. Then a globally hyperbolic spacetime is one where one can predict what happens everywhere from data on Σ.

- **Theorem.** (Geroch) If (M, g) is globally hyperbolic, then there exists a global time function, a map $t\colon M \to \mathbb{R}$ so that $-(dt)^a$ is future-directed, timelike surfaces of constant t are Cauchy surfaces with the same topology Σ, and the topology of M is $\mathbb{R} \times \Sigma$. This rules out pathologies such as closed timelike curves.

- As a result, in a globally hyperbolic spacetime we can perform an 'ADM decomposition' of spacetime. Let t be the time function and choose coordinates x^i on the Cauchy surface $t = 0$. Then we can define the x^i globally by following the integral curves of ∂_t, giving coordinates (t, x^i). It is conventional to write the metric as

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

  where $N(t, x)$ is called the lapse function and $N^i(t, x)$ is called the shift vector. We used a similar construction in the Schwarzschild spacetime.

Example. We give some basic examples of these definitions.

- Let Σ be the positive x-axis in two-dimensional Minkowski space M. Then the boundary of $D(\Sigma)$ is bounded by the null rays $t = \pm x$. If Σ' is the entire x-axis then $D(\Sigma') = M$, so M is globally hyperbolic.

- If we delete the origin from M, the resulting spacetime is not globally hyperbolic because geodesics can end at the origin.
- The Kruskal spacetime is globally hyperbolic, with global time function $\hat{t} = U + V$. The surface $U + V = 0$ is a Cauchy surface, and it is an Einstein–Rosen bridge with topology $\mathbb{R} \times S^2$. Then the spacetime has topology $\mathbb{R}^2 \times S^2$.

Next, we describe the initial value problem for general relativity.

- The initial data for Einstein's equation is a triple (Σ, h_{ab}, K_{ab}) where (Σ, h_{ab}) is a Riemannian 3-manifold and K_{ab} is a symmetric tensor. Intuitively, Σ is a spacelike hypersurface in spacetime, h_{ab} is the pullback of the metric, and K_{ab} is the extrinsic curvature tensor of Σ.
- Let n^a denote the unit vector normal to Σ. Then the Einstein equation imposes constraints on the initial data. Contracting it with $n^a n^b$ gives the Hamiltonian constraint

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

  where R' is the Ricci tensor of h_{ab} and all indices are raised and lowered with h^{ab}, and $\rho = T_{ab} n^a n^b$ is the energy density measured by an observer with velocity n^a.
- Contracting the Einstein equation with n^a and projecting orthogonally to it gives the momentum constraint

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

  where D_a uses the Levi–Civita connection of h_{ab}, and the right-hand side is the momentum density measured by an observer with velocity n^a.
- **Theorem.** Consider initial data as defined above satisfying the constraints in vacuum. Then there exists a unique spacetime (M, g_{ab}), up to diffeomorphism, called the maximal Cauchy development of the initial data, so that

  1. (M, g_{ab}) satisfies the vacuum Einstein equation,
  2. (M, g_{ab}) is globally hyperbolic with Cauchy surface Σ,
  3. the induced metric and extrinsic curvature of Σ are h_{ab} and K_{ab},
  4. any other spacetime satisfying these conditions is isometric to a subset of (M, g_{ab}).

  Analogous theorems also exist for suitable matter obeying hyperbolic PDEs.
- Note that the maximal Cauchy development could be extendible to (M', g'_{ab}), but then Σ would necessarily not be a Cauchy surface for (M', g'_{ab}). Then we cannot predict the metric in all of M' using only data on Σ, i.e. the extension can't be unique.

Example. Now we give some examples of this result.

- Consider initial data on a surface $\Sigma = \{(x, y, z) \colon x > 0\}$ with flat metric and vanishing extrinsic curvature. The maximal development is the region $|t| < x$ of Minkowski spacetime, which can be extended. However, the extension is far from unique: it could be the entirety of Minkowski spacetime or it could be curved.

- Consider the Schwarzschild solution with $M < 0$,

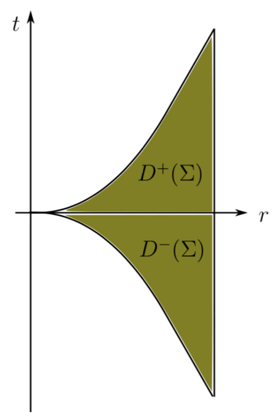
$$ds^2 = -\left(1 + \frac{2|M|}{r}\right) dt^2 + \left(1 + \frac{2|M|}{r}\right)^{-1} dr^2 + r^2 d\Omega^2.$$

  There is a curvature singularity at $r = 0$ but no event horizon. We take initial data on the surface Σ given by $t = 0$, with pullback metric h_{ab}. As a Riemannian manifold, (Σ, h_{ab}) is inextendible, but it is not geodesically complete since some geodesics hit $r = 0$ in finite affine parameter. We say the initial data is singular.

- The resulting maximal development is not the entire Schwarzschild spacetime. Consider an outgoing radial null geodesic; it obeys

$$\frac{dt}{dr} \approx \frac{r}{2|M|}$$

  at small r. Then the domain of dependence of Σ is shown below.



  Points outside of $D(\Sigma)$ have causal geodesics that do not pass through Σ because they instead end at $r = 0$. The boundary of $D(\Sigma)$ is given by the null geodesics that emerge from $r = 0$ at time $t = 0$.

- Therefore, the initial data do not predict the metric outside of $D(\Sigma)$. There could be other extensions besides the $M < 0$ Schwarzschild solution. By Birkhoff's theorem, they are necessarily not spherically symmetric.

- So far we've seen examples where the initial data is extendible, in which case it's clear we should be 'missing information', and cases where the initial data is singular, in which case one 'can't predict what comes out of a singularity'. Thus we restrict to initial data which is geodesically complete and hence also inextendible.

- Let Σ be the hyperboloid $t^2 - \mathbf{x}^2 = -1$ in Minkowski space for $t < 0$, with h_{ab} and K_{ab} defined as in Minkowski space. This initial data is geodesically complete, but its maximal Cauchy development is the past light cone of the origin, and hence is extendible. The intuitive reason is that Σ is 'asymptotically null' which allows information to 'arrive from infinity'.

Now that we've seen the possible problems, we can define new restrictions to avoid them.

- An initial data set (Σ, h_{ab}, K_{ab}) is an asymptotically flat end if

1. Σ is diffeomorphic to $\mathbb{R}^3/B$ where B is a closed ball centered on the origin,
2. if we pullback the $\mathbb{R}^3$ coordinates to define coordinates x^i on Σ then
$$h_{ij} = \delta_{ij} + O(1/r), \quad K_{ij} = O(1/r^2)$$
for large $r = \sqrt{x^i x^i}$,
3. derivatives of the latter expressions also hold, e.g. $h_{ij,k} = O(1/r^2)$.

An initial data set is asymptotically flat with N ends if it is the union of a compact set with N asymptotically flat ends. If matter fields are also present, they should also decay appropriately. We sometimes say Σ itself is asymptotically flat.

- Intuitively, in an asymptotically flat end, Σ looks like a surface of constant t in Minkowski spacetime for large r. If Σ is asymptotically flat with N ends, it looks like the union of N such surfaces.
- For example, in the $M > 0$ Schwarzschild spacetime, one can show that the initial data on the surface $\Sigma = \{t = \text{const}, r > 2M\}$ is an asymptotically flat end. However, it is not geodesically complete, since it stops at $r = 2M$. Now, in the Kruskal spacetime Σ is part of an Einstein–Rosen bridge, which is asymptotically flat with two ends, because it is the union of the compact sphere $U = V = 0$ with two copies of the asymptotically flat end defined above.
- Penrose's strong cosmic censorship conjecture states that generically, a geodesically complete, asymptotically flat initial data set has an inextendible maximal Cauchy development. It is related to the weak cosmic censorship conjecture, which informally states that 'every singularity is hidden behind an event horizon', because both are about the predictability of GR.
- For small perturbations of Minkowski spacetime, it is known that the spacetime 'settles down to Minkowski spacetime at late time', which implies strong cosmic censorship holds.
- As we'll see, strong cosmic censorship fails for charged and rotating black holes, which exhibit Cauchy horizons. However, the conjecture does hold for any small perturbation of these initial conditions, i.e. it fails only on a set of 'measure zero'.
- A maximal Cauchy development cannot contain a region with closed timelike curves, since by definition such a region contains causal curves that don't intersect Σ. This represents another counterexample, but it is again not generic.
- The conjecture can be extended to include matter, but we must impose the dominant energy condition, which essentially requires matter with positive energy density that doesn't travel faster than light. We'll discuss these energy conditions in more detail later.

7.2 Geodesic Congruences

Next, we need to establish some basic definitions.

- A null hypersurface $\mathcal{N}$ is a hypersurface whose normal N_a is everywhere null. That is, $N_a X^a = 0$ for any X^a tangent to $\mathcal{N}$, so X^a is spacelike or parallel to N^a. In particular, N^a itself is tangent to $\mathcal{N}$, so its integral curves, called the generators of $\mathcal{N}$, lie within $\mathcal{N}$.

- We claim that the generators of $\mathcal{N}$ are null geodesics.

  The generators are null by definition. Now let $\mathcal{N}$ be defined by $f = \text{const}$ where $df \neq 0$ on $\mathcal{N}$. Then we have $N = h df$ for some function h, and we can rescale so that $N = df$, since this just reparametrizes the geodesics. Since $N^a N_a = 0$, its gradient is normal to $\mathcal{N}$, so

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

  for some function α. We also have $\nabla_a N_b = \nabla_a \nabla_b f = \nabla_b \nabla_a f = \nabla_b N_a$, giving

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

  which is simply the geodesic equation for a non-affine parameter.

Example. In the Kruskal spacetime, let $N = dU$. Since $g^{UU} = 0$, N is null everywhere and hence normal to a family of null hypersurfaces, each with constant U. In particular, we have $\nabla_a (N^b N_b) = 0$, so the right-hand side of the above equations is zero. Then the generators are affinely parametrized null geodesics. Raising an index gives

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

For $U = 0$, we have $r = 2M$, so N^a is just a constant multiple of $\partial/\partial V$. Then V is an affine parameter for the generators of the surface $U = 0$.

Next, we review geodesic deviation and introduce geodesic congruences.

- We recall that a one-parameter family of geodesics gives a surface with coordinates (s, λ) where $U = \partial/\partial\lambda$ is the geodesic velocity and $S = \partial/\partial s$ is the deviation vector, and $[U, S] = 0$. The geodesic equation is $\nabla_U U = 0$ and the geodesic deviation equation is

  $$\nabla_U \nabla_U S^a = R^a{}_{bcd} U^b U^c S^d$$

  and given an affinely parametrized geodesic γ with tangent U^a, a solution S^a of this equation along γ is called a Jacobi field.

- A geodesic congruence in an open set $\mathcal{U} \subset M$ is a family of geodesics so that exactly one geodesic passes through each point in $\mathcal{U}$. We will consider the case where all the geodesics in a congruence are null/spacelike/timelike, normalizing the tangent vector U^a to $U^2 = 0/1/-1$.

- Consider a one-parameter family of geodesics belonging to a congruence, so that

  $$\nabla_U S^a = B^a{}_b S^b, \quad B^a{}_b = \nabla_b U^a.$$

  Then we have the identities

  $$B^a{}_b U^b = 0, \quad U_a B^a{}_b = \frac{1}{2} \nabla_b (U^2) = 0.$$

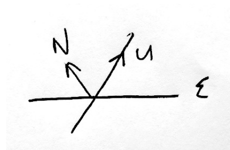
  by the geodesic equation, so

  $$\nabla_U (U \cdot S) = (\nabla_U U^a) S_a + U^a \nabla_U S_a = 0$$

  where we again used the geodesic equation. Therefore, $U^a S_a$ is constant along geodesics.

- Now, a one-parameter family of geodesics has plenty of freedom in the coordinates, since we may redefine the parameter on each geodesic, $\lambda' = \lambda - a(s)$, inducing the change
$$S'^a = S^a + \frac{da}{ds} U^a.$$
Intuitively, it's nice to make the separation normal to the velocity, and
$$U \cdot S' = U \cdot S + \frac{da}{ds} U^2.$$
Then in the spacelike and timelike case, we can straightforwardly set $U^a S'_a = 0$ everywhere.

- In the null case, U^2 vanishes and we have to work more carefully.



We choose a spacelike hypersurface Σ which intersects each geodesic once. Let N^a be a vector field on Σ so that $N^2 = 0$ and $N^a U_a = -1$. We extend N^a off Σ by parallel transport along the geodesics, $\nabla_U N^a = 0$. Then
$$N^2 = 0, \quad N \cdot U = -1, \quad \nabla_U N^a = 0$$
everywhere.

- Therefore, we can decompose any deviation vector uniquely as
$$S^a = \alpha U^a + \beta N^a + \hat{S}^a, \quad U \cdot \hat{S} = N \cdot \hat{S} = 0$$
where $\hat{S}^a$ points 'into the page' in the above diagram. Note that $U \cdot S = -\beta$, so β is constant along each geodesic. In the case where we consider a subset of the generators of a null hypersurface, we automatically have $\beta = 0$ by definition.

- We can project onto $\hat{S}^a$ by
$$\hat{S}^a = P^a_b S^b, \quad P^a_b = \delta^a_b + N^a U_b + U^a N_b$$
where P^a_b is a projection operator, $P^a_b P^b_c = P^a_c$, onto the subset $T_\perp$ of the tangent space at p containing vectors orthogonal to U^a and N^a. We also have
$$\nabla_U P^a_b = 0$$
because it is built out of N^a and U^b, which are both parallel transported.

- We claim that if $U \cdot S = 0$, then
$$\nabla_U \hat{S}^a = \hat{B}^a{}_b \hat{S}^b, \quad \hat{B}^a{}_b = P^a_c B^c{}_d P^d_b.$$
This is intuitively reasonable, as it's just one of our earlier results with projectors applied everywhere. Explicitly, we have
$$\nabla_U \hat{S}^a = \nabla_U (P^a_c S^c) = P^a_c \nabla_U S^c = P^a_c B^c{}_d S^d = P^a_c B^c{}_d P^d_e S^e$$
where the final step works because $U \cdot S = 0$ and $B^c{}_d U^d = 0$. Finally, using $P^2 = P$ gives the desired result.

We now examine $\hat{B}^a{}_b$ in more detail.

- We can think of $\hat{B}^a{}_b$ as a matrix that acts on the two-dimensional space $T_\perp$. To understand it geometrically, we divide it into the expansion, shear, and rotation defined as

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

  Then we have

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

- By plugging in the definitions, we have

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

  which implies that

  $$\theta = g^{ab} B_{ab} = \nabla_a U^a$$

  so it can be interpreted as the divergence of the geodesics; this shows that θ is independent of the choice of N^a. Similarly, scalar invariants of the rotation and shear are independent of N^a.

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

  To see this, start with our expression for $\hat{B}^b{}_c$ and note that

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

  where the extra terms drop out of the antisymmetrization. Using the definition of B_{ab},

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

  As we've shown earlier, this vanishes if U is normal to $\mathcal{N}$, so on $\mathcal{N}$,

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

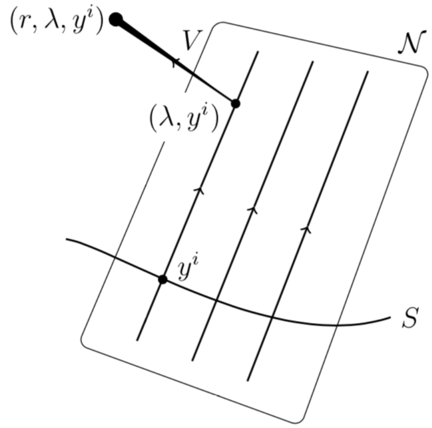
  Contracting this with N^a and using $\hat{\omega} \cdot N = 0$ gives the result. The reasoning can be run backwards by Frobenius' theorem.

- Therefore, in the case of a null hypersurface we only have to deal with expansion and shear. Intuitively, expansion increases the cross-sectional area of a family of geodesics, while shear compresses in one direction and stretches in the other, keeping the area constant.

To understand the expansion more quantitatively, we use Gaussian null coordinates.

- We pick a two-dimensional spacelike surface S within $\mathcal{N}$ and let y^i be coordinates on this surface; we then define coordinates (λ, y^i) on $\mathcal{N}$ by following the point with coordinates y^i for parameter distance λ along the generator through it, so $U = \partial/\partial\lambda$.

- Next, let V^a be a null vector field on $\mathcal{N}$ so that $V \cdot \partial/\partial y^i = 0$ and $V \cdot U$, similarly to how we defined N^a, and define the r coordinate by following its geodesics so $V = \partial/\partial r$ and $\mathcal{N}$ is the surface $r = 0$. The coordinates (r, λ, y^i) are Gaussian null coordinates, shown below.



- Now we consider the form of the metric. Since V^a is null, $g_{rr} = 0$, and the geodesic equation for V^a gives $\partial_r g_{r\mu} = 0$. On the surface, we have $g_{r\lambda} = 1$ and $g_{ri} = 0$, which then hold for all r. We also know that $g_{\lambda\lambda} = g_{\lambda i} = 0$ at $r = 0$, so

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

where F and h_i are smooth functions.

- The metric restricted to $\mathcal{N}$ is

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

so since $U^\mu = (0, 1, 0, 0)$ on $\mathcal{N}$, we have $U_\mu = (1, 0, 0, 0)$ on $\mathcal{N}$. Then since $U \cdot B = B \cdot U = 0$, we have $B^r{}_\mu = B^\mu{}_\lambda = 0$. Therefore on $\mathcal{N}$ we have

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

Using the form of the metric, we have

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

where $h = \det h_{ij}$, h^{ij} is the matrix inverse, and we used the identity $\delta(\det A) = (\det A)\,\mathrm{tr}(A^{-1}\delta A)$.

- Therefore, we have

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

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

7.3 Raychaudhuri's Equation

Next, we define trapped surfaces.

- Let S be a two-dimensional spacelike surface. Then for any point $p \in S$ there are two future-directed null vectors U_1 and U_2 orthogonal to S, up to scaling. If S is orientable, then U_1 and U_2 can be desired continuously over S. This defines two families of null geodesics which start on S and are orthogonal to S, forming the null hypersurfaces $\mathcal{N}_1$ and $\mathcal{N}_2$.

- For example, in Minkowski space, the two-sphere $r = \text{const}$, $t = \text{const}$ has U_1 and U_2 pointing radially inward and outward. This is a bit tricky to visualize or draw, since it requires all four spacetime dimensions.

- In the Kruskal spacetime, let S be the two-sphere $U = U_0$, $V = V_0$. Then the generators of the $\mathcal{N}_i$ are radial null geodesics as in Minkowski space. We know that dU and dV correspond to affine parametrization, so raising an index gives
$$U_1 = re^{r/2M}\frac{\partial}{\partial V}, \quad U_2 = re^{r/2M}\frac{\partial}{\partial U}$$
where the signs are chosen so that U_1 and U_2 are future-directed.

- Using the divergence formula we can compute the expansion
$$\theta_1 = \nabla_a U_1^a = \frac{1}{\sqrt{-g}}\partial_\mu(\sqrt{-g}U_a^\mu) = r^{-1}e^{r/2M}\partial_V(re^{-r/2M}re^{r/2M}) = -\frac{8M^2}{r}U$$
and similarly
$$\theta_2 = -\frac{8M^2}{r}V.$$
Then in region I, the outgoing null geodesics on S are expanding and the ingoing null geodesics are converging, as expected under normal conditions.

- A compact, orientable, spacelike two-dimensional surface is trapped if both families of null geodesics orthogonal to S have negative expansion everywhere on S, and marginally trapped if both families have non-positive expansion. Then two-spheres in region II are trapped, and two-spheres on the event horizon are marginally trapped.

Next, we derive Raychaudhuri's equation, which describes the evolution of the expansion along the geodesics of a null geodesic congruence. Applying it will requires discussing energy conditions.

- Raychaudhuri's equation states that
$$\frac{d\theta}{d\lambda} = -\frac{1}{2}\theta^2 - \hat{\sigma}^{ab}\hat{\sigma}_{ab} + \hat{\omega}^{ab}\hat{\omega}_{ab} - R_{ab}U^aU^b.$$
To see this, note that by definition we have
$$\frac{d\theta}{d\lambda} = \nabla_U(B^a{}_b P_a^b) = P_a^b \nabla_U B^a{}_b = P_a^b U^c \nabla_c \nabla_b U^a.$$
Next, we commute the covariant derivatives to get a factor of the Riemann tensor,
$$\frac{d\theta}{d\lambda} = P_a^b U^c(\nabla_b\nabla_c U^a + R^a{}_{dcb}U^d) = -P_a^b(\nabla_b U^c)(\nabla_c U^a) + \delta_a^b R^a{}_{dcb}U^cU^b$$
where we used the geodesic equation and the antisymmetry of the Riemann tensor. The first term is $-B^c{}_b P_a^b B^a{}_c$, and inserting projectors and identities turns it into $-\hat{B}^c{}_a\hat{B}^a{}_c$, which expands to give the first three terms.

- To control the last term in Raychaudhuri's equation, we impose energy conditions. The most important condition is the dominant energy condition (DEC), which requires $-T^a{}_b V^b$ to be a future-directed causal vector, or zero, for all future-directed timelike vectors V^a.
- The motivation for the DEC is that the energy momentum current measured by an observer with four-velocity u^a is $j^a = -T^a{}_b u^b$. Heuristically, the dominant energy condition restricts matter to not move faster than light. For instance, one can show that if T_{ab} is zero in a closed region S of some spacelike hypersurface Σ and obeys the DEC, then T_{ab} is zero within $D^+(S)$.
- The weak energy condition (WEC) only requires $T_{ab}V^aV^b \geq 0$ for any causal vector V^a, which corresponds to all observers measuring nonnegative energy density.
- The null energy condition (NEC) is even weaker, specializing to null V^a. It implies $|w| \leq 1$ for perfect fluids obeying $p = w\rho$.
- The strong energy condition (SEC) is stronger than the NEC but independent of the WEC and DEC. It requires
$$\left(T_{ab} - \frac{1}{2}g_{ab}T^c{}_c\right)V^aV^b \geq 0$$
for all causal vectors V^a. By the Einstein equation, this is equivalent to $R_{ab}V^aV^b \geq 0$, which implies that gravity is attractive. However, while the DEC appears to be satisfied in our universe, the SEC is not, since the cosmological constant is positive.
- If the NEC applies, then the generators of a null hypersurface satisfy
$$\frac{d\theta}{d\lambda} \leq -\frac{1}{2}\theta^2.$$
To see this, note that $\hat{\omega}$ is zero in Raychaudhuri's equation. The metric restricted to $T_\perp$ is positive-definite, so $\hat{\sigma}^{ab}\hat{\sigma}_{ab}$ is positive. Since U^a is null, Einstein's equation gives $R_{ab}U^aU^b = 8\pi T_{ab}U^aU^b$, so $R_{ab}U^aU^b$ is positive.
- Therefore, if $\theta = \theta_0 < 0$ at a point p on a generator γ of a null hypersurface, then $\theta \to -\infty$ along γ within an affine parameter distance $2/|\theta_0|$, provided γ extends this far.

Example. Consider a massless scalar field, with
$$T_{ab} = \partial_a\Phi\partial_b\Phi - \frac{1}{2}g_{ab}(\partial\Phi)^2.$$
Then we have
$$j^a = -T^a{}_b V^b = -(V^b\partial_b\Phi)\partial^a\Phi + \frac{1}{2}V^a(\partial\Phi)^2, \quad j^2 = \frac{1}{4}V^2((\partial\Phi)^2)^2.$$
For timelike V^a, this implies j^a is causal or zero. To check its orientation, note that
$$V^a j_a = (-V\cdot\partial\Phi)^2 + \frac{1}{2}V^2(\partial\Phi)^2 = -\frac{1}{2}(V^a\partial_a\Phi)^2 + \frac{1}{2}V^2\left(\partial^a\Phi - \frac{V^b\partial_b\Phi}{V^2}V^a\right)^2.$$
The final expression in brackets is orthogonal to V^a, so its norm is non-negative. Therefore, $V\cdot j \leq 0$ so j^a is future-directed or zero, establishing the DEC.

Finally, we define conjugate points.

- Two points p and q are conjugate along a geodesic γ if γ passes through p and q, and there exists a Jacobi field along γ that vanishes at p and q but is not identically zero. Intuitively, this means that a group of geodesics infinitesimally close to γ converge at both p and q.
- **Theorem.** Consider a null geodesic congruence including all of the null geodesics through p. If $\theta \to -\infty$ at a point q on a null geodesic γ through p, then p is conjugate to q along γ.
- **Theorem.** Let γ be a causal curve containing p and q. Then there does not exist a smooth, one-parameter family of causal curves γ_s connecting p and q with $\gamma_0 = \gamma$ and γ_s timelike for $s > 0$ (i.e. γ cannot be deformed smoothly to a timelike surface) if and only if γ is a null geodesic with no point conjugate to p along γ between p and q.
- Now suppose we have a two-dimensional spacelike surface S, and consider one of the geodesics γ that generate $\mathcal{N}_1$ or $\mathcal{N}_2$. We say a point p along γ is conjugate to S if there exists a Jacobi field along γ that vanishes at p, and is tangent to S on S. That is, null geodesics emitted from S converge at p. The analogue of the above theorem is that p is conjugate to S if $\theta \to -\infty$ at p.

Example. Consider geodesics on $\mathbb{R} \times S^2$ with metric $ds^2 = -dt^2 + d\Omega^2$. The geodesics travel on great circles of S^2. Then the North and South poles are conjugate points. Now consider a null geodesic from the North pole to the equator, which passes through the South pole; such a path wraps around the sphere 3/4 of the way. Then it's possible to deform the path to be shorter, making it timelike.

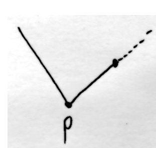
7.4 Causal Structure

We now make some formal definitions regarding causal structure.

- Let (M, g) be a time-orientable manifold with $U \subset M$. The chronological future $I^+(U)$ of U is the set of points of M which can be reached by a future-directed timelike curve starting on U. The causal future $J^+(U)$ of U is the union of U with the set of points of M that can be reached by a future-directed causal curve starting on U. We define the chronological past $I^-(U)$ and the causal past $J^-(U)$ similarly.
- For a point p in Minkowski space, $J^+(p)$ is the set of points on or inside the future light cone including p itself, while $I^+(p)$ is the interior of $J^+(p)$.
- It can be shown that in general, we have

  $$I^+(U) = \mathrm{int}(J^+(U)), \quad J^+(U) \subset \overline{I^+(U)}.$$

  In Minkowski space, the latter is an equality. However, consider two-dimensional Minkowski space with the origin deleted, as shown below.

  Then the dotted line is in $\overline{I^+(p)}$ but not in $J^+(p)$, since a light ray would have to pass through the origin to reach it.

- We write the boundary of $U \subset M$ as $\dot{U} = \overline{U}/\mathrm{int}(U)$. Then we have

$$\overline{J^+(U)} = \overline{I^+(U)}, \quad \dot{J}^+(U) = \dot{I}^+(U).$$

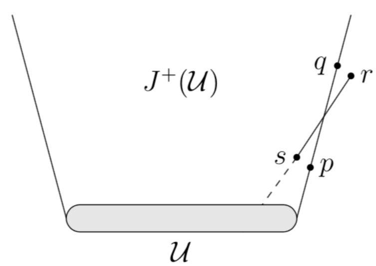
In Minkowski space, $\dot{I}^+(p)$ is the set of points along future-directed null geodesics starting from p. In general, this statement holds locally, as shown by the following theorem.

- **Theorem.** Given $p \in M$ there exists a convex normal neighborhood U of p, where for any $q, r \in U$ there exists a unique geodesic connecting q and r that stays in U. Then $\dot{I}^+(p)$ in the spacetime (U, g) is the set of all points in U along future-directed null geodesics in U that start at p.

- **Corollary.** If $q \in J^+(p) \setminus I^+(p)$, there exists a null geodesic from p to q.

Proof: given a causal curve connecting p and q with parameter in $[0, 1]$, we can cover it with finitely many convex normal neighborhoods since $[0, 1]$ is compact. Then we use the above theorem in each neighborhood.

- **Theorem.** A set $S \subset M$ is achronal if no two points in S are connected by a timelike curve. Then $\dot{J}^+(U)$ is an achronal three-dimensional submanifold of M.



Proof: consider $p, q \in \dot{J}^+(U)$ and suppose $q \in I^+(p)$. Since $I^+(p)$ is open, there exists r near q in $I^+(p)$ but not $J^+(U)$. Similarly, since $I^-(r)$ is open, there exists s near p in $I^-(r)$ and $J^+(U)$. Then there exists a causal curve from U to s to r, so $r \in J^+(U)$, a contradiction.

- As an example, consider $M = \mathbb{R} \times S^1$ with the flat metric

$$ds^2 = -dt^2 + d\phi^2$$

which is a two-dimensional Einstein static universe. Then $\dot{J}^+(p)$ is a pair of null geodesic segments that start at p and end where they meet at q. The geodesics have future endpoint q and past endpoint p. In our example above, if p is on the dotted line then the geodesic is past-inextendible as it hits the origin. This is general, as shown by the following theorem.

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

- In a globally hyperbolic spacetime, the above theorem can be strengthened to rule out the former case, as shown by the following theorem.

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

- Finally, we formally define the future Cauchy horizon of a partial Cauchy surface Σ as $H^+(\Sigma) = \overline{D^+(\Sigma)}/I^-(D^+(\Sigma))$.

  Note that we don't define $H^+(\Sigma)$ as $\dot{D}^+(\Sigma)$ because this includes Σ itself. However, one can show that $\dot{D}(\Sigma) = H^+(\Sigma) \cup H^-(\Sigma)$ and that the $H^\pm$ are null hypersurfaces.

Finally, we're ready to state the Penrose singularity theorem.

- **Theorem.** Let (M, g) be globally hyperbolic with a noncompact Cauchy surface Σ. Assume the Einstein equation and the NEC are satisfied and M contains a trapped surface T. Let $\theta_0 < 0$ be the maximum value of θ on T for both sets of null geodesics orthogonal to T. Then at least one of these geodesics is future-inextendible and has affine length no greater than $2/|\theta_0|$.
- We give a very basic proof sketch. Assume the opposite for the sake of contradiction. Then by our previous results, any future-inextendible null geodesic orthogonal to T contains a point conjugate to T within affine parameter $2/|\theta_0|$.
- Next, let $p \in \dot{J}^+(T)$ with $p \notin T$. Then by our previous theorem, p lies on a future-directed null geodesic γ starting from T which is orthogonal to T and has no point conjugate to T between T and p. Then p cannot lie beyond the point conjugate to T.
- Therefore, $\dot{J}^+(T)$ is a subset of the compact set consisting of the set of points along the null geodesics orthogonal to T with affine parameter less than or equal to $2/|\theta_0|$. Since $\dot{J}^+(T)$ is closed, $\dot{J}^+(T)$ is thus compact. On the other hand, $\dot{J}^+(T)$ is a manifold, and hence can't have a boundary.
- This is a contradiction, unless Σ is compact, because the 'ingoing' and 'outgoing' congruences orthogonal to T can 'join up', as we saw in the Einstein static universe. Assuming Σ is noncompact gives the desired contradiction.
- The Penrose singularity theorem assumes the existence of a trapped surface, and it can be shown that trapped surface are generic. There is plenty of numerical evidence for this, as well as some mathematical evidence.
- For example, the Einstein equations possess the property of Cauchy stability, which implies that the solution in a compact region of spacetime depends continuously on the initial data. Now consider a sphere in region II of the Kruskal diagram, which contains a trapped surface. Then by Cauchy stability we would also have a trapped surface if the initial data were perturbed, so trapped surfaces occur generically in gravitational collapse.
- A theorem due to Schoen and Yau shows that asymptotically flat initial data will contain a trapped surface if the energy density of matter is sufficiently large. Christodoulou has shown that trapped surfaces can be formed even in the absence of matter, by gravitational waves.
- The Penrose singularity theorem states that if the maximal development of asymptotically flat initial data contains a trapped surface, then the maximal development is not geodesically complete. This could be because the maximal development is extendible, but this is not generic if the strong cosmic censorship conjecture holds. Then generically the result is a singularity.
- A different singularity theorem due to Hawking and Penrose relaxes the assumption that spacetime is globally hyperbolic and adds the SEC and a mild genericity assumption, and arrives at the same result.

- Thus, we have very good reasons to believe that gravitational collapse leads to the formation of a singularity. Note that this need not be a curvature singularity.

8 Asymptotic Flatness

8.1 Conformal Compactification

In this section, we rigorously define a black hole. We begin by studying conformal compactification, a useful tool for visualizing spacetimes.

- Given a spacetime (M, g), we can define a new, 'unphysical' metric $\overline{g} = \Omega^2 g$ where Ω is smooth and positive. We say $\overline{g}$ is obtained from g by a conformal transformation. (Note that in conformal field theory, such an operation is instead called a Weyl transformation.)
- Conformal transformations preserve timelike, spacelike, and null directions; in particular they preserve light cones, and hence the causal structure.
- In conformal compactification, we choose Ω so that the 'points at infinity' with respect to g are at finite distance with respect to $\overline{g}$, which requires $\Omega \to 0$ at infinity.
- The resulting spacetime $(M, \overline{g})$ is extendible to a larger spacetime $(\overline{M}, \overline{g})$, and we identify M as a subset of $\overline{M}$ where $\Omega = 0$ on ∂M.

Example. Minkowski spacetime. The metric is

$$g = -dt^2 + dr^2 + r^2 d\omega^2$$

where we changed the angular measure to avoid confusion with the conformal factor Ω. We then define the null coordinates

$$u = t - r, \quad v = t - r, \quad -\infty < r \leq v < \infty, \quad g = -du dv + \frac{1}{4}(u - v)^2 d\omega^2.$$

Next, we define the coordinates (p, q) by

$$u = \tan p, \quad v = \tan q, \quad -\pi/2 < p \leq q < \pi/2, \quad g = (2 \cos p \cos q)^{-2}(-4 dp dq + \sin^2(q - p) d\omega^2).$$

The original 'infinity' corresponds to $|t| \to \infty$ or $r \to \infty$, and now corresponds to $|p| \to \pi/2$ or $|q| \to \pi/2$. To perform conformal compactification, we define

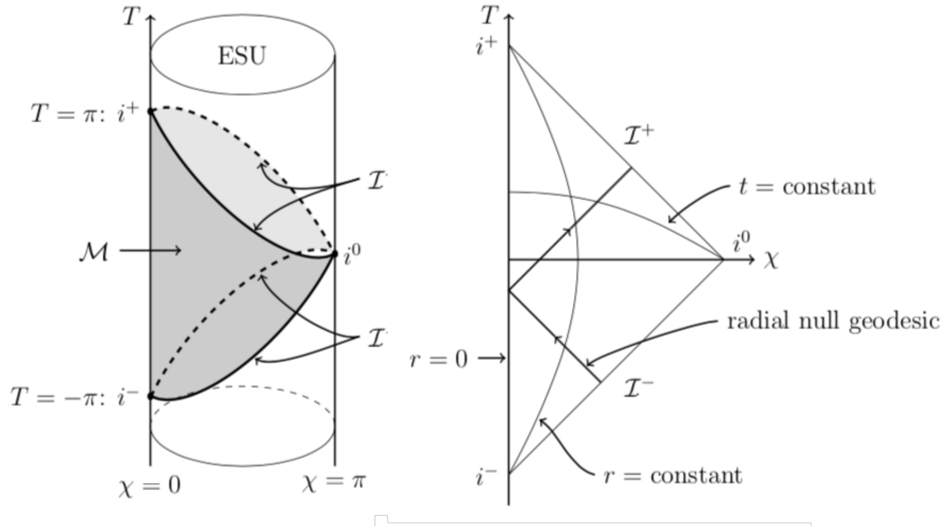
$$\Omega = 2 \cos p \cos q, \quad \overline{G} = -4 dp dq + \sin^2(p - q) d\omega^2.$$

Finally, we switch back to timelike and spacelike coordinates by

$$T = q + p \in (-\pi, \pi), \quad \chi = q - p \in [0, \pi), \quad \overline{g} = -dT^2 + d\chi^2 + \sin^2 \chi d\omega^2.$$

By extending the range of T to $(-\infty, \infty)$ and the range of χ to $[0, \pi]$, we arrive at the Einstein static universe $\mathbb{R} \times S^3$. To visualize our spacetime, we show only T and χ. Then Minkowski space can be depicted as a slice of the Einstein static universe. Alternatively, we can project the slice to get a Penrose diagram. Formally, a Penrose diagram is a bounded subset of $\mathbb{R}^2$ endowed with a flat Lorentzian metric; every point on the internet corresponds to a sphere S^2. Points on the boundary can represent either points at infinity, or points of symmetry, such as $r = 0$.

There are several regions of interest on the boundary. Radial null geodesics come from the null hypersurface $\mathcal{I}^-$, called past null infinity, and end at $\mathcal{I}^+$, called future null infinity. Similarly, radial timelike geodesics start at past timelike infinity i^- and end at future timelike infinity i^+, while radial spacelike geodesics start and end at spatial infinity i^0.



We can also consider non-radial geodesics. From the point of view of this diagram, non-radial motion is simply a reduction in spatial velocity, so a non-radial null geodesic looks like a radial timelike geodesic. Also note that a timelike curve that is not a geodesic can end up at $\mathcal{I}^+$, provided it is 'asymptotically null'.

Note. The fact that i^0 and $i^\pm$ are single points is a bit misleading. Timelike geodesics do not actually converge; they merely approach regions that are increasingly shrunk by the conformal transformation. The real lesson here is that the past light cones of any two events will intersect.

Note. The behavior of geodesics has an analogue for fields. Consider a massless scalar field ψ in Minkowski spacetime, which satisfies the wave equation $\nabla^a \nabla_a \psi = 0$. Spherically symmetric solutions take the form

$$\psi(t,r) = \frac{1}{r}(f(t-r) + g(t+r)).$$

This is singular unless $g(x) = -f(x)$, giving

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

Now, on $\mathcal{I}^-$, we have $p = -\pi/2$, and the solution is

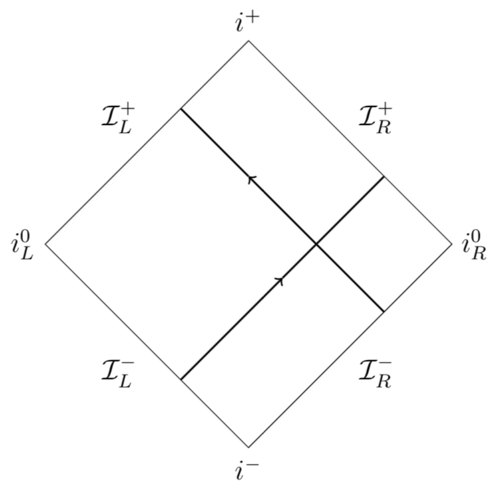
$$\frac{1}{r}F_0(q) = \frac{1}{r}(F(-\pi/2) - F(q)).$$

Then the solution everywhere can be written in terms of $F_0(q)$,

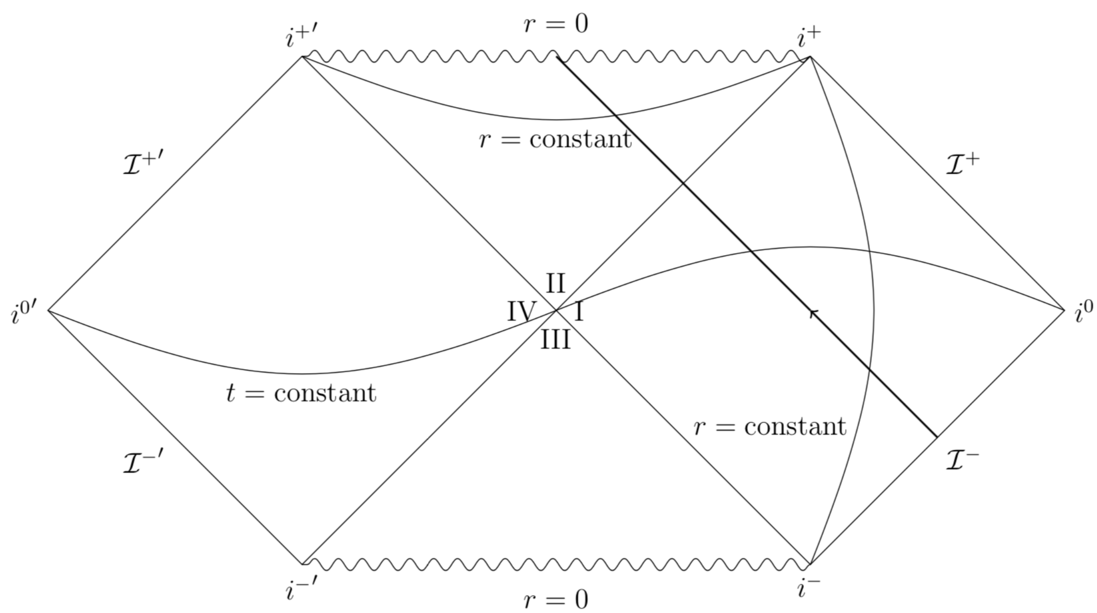
$$\psi(t,r) = \frac{1}{r}(F_0(q) - F_0(p))$$

so it is determined by the solution on $\mathcal{I}^-$. Similarly, it is determined by the solution on $\mathcal{I}^+$.

Example. Two-dimensional Minkowski spacetime, $g = -dt^2 + dr^2$. Everything proceeds as before, except that now $r \in (-\infty, \infty)$. Then the Penrose diagram has two disconnected spatial infinities and null infinities. In the three dimensional case, spatial infinity is instead a sphere S^2 and is hence connected.

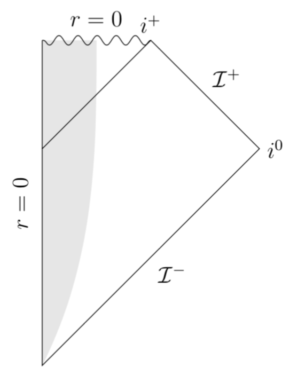


Example. The Kruskal spacetime. We know the spacetime has two asymptotically flat regions, so the 'infinity' in each of these regions should be like that of four-dimensional Minkowski spacetime. The coordinates U and V are already null, so to construct the Penrose diagram, we would have to define coordinates $P = P(U)$ and $Q = Q(V)$ so that the range of P and Q is finite, and the unphysical metric $\overline{g}$ has a smooth extension.

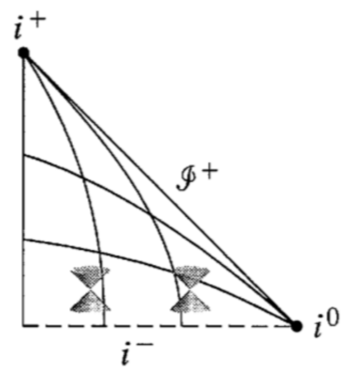


Performing this explicitly takes a lot of work, but we can guess the answer using the Kruskal diagram. We leave everything unchanged, except we use the conformal freedom to turn the curvature singularity at $r = 0$ into a horizontal line. Note that timelike infinity is singular, since lines of constant r meet there, including the curvature singularity $r = 0$. The timelike infinities are single points, but as in the Minkowski case, this doesn't mean that every timelike geodesic converges; there are plenty of such geodesics that don't hit the singularity.

Example. Consider spherically symmetric gravitational collapse. Since the region outside the star is asymptotically flat, we simply have part of the Minkowski space diagram. The metric everywhere outside of the shaded region is Schwarzschild, so the upper part is taken from the Kruskal diagram. Again, past spatial infinity i^- is a single point, from which the matter arrives.



Example. The Penrose diagram for a Robertson-Walker universe with $a(t) \propto t^q$ and $q \in (0, 1)$. There is a singularity at $T = 0$.



Here, the singularity forms the past spatial infinity, which is no longer a single point. Accordingly, there are causally disconnected regions in the early universe.

Next, we give some more facts about conformal transformations. To avoid confusion, we will switch here to calling such a transformation a Weyl transformation, in agreement with conventions outside of relativity.

- We denote the Levi–Civita connection of the unphysical metric $\overline{g}$ by $\overline{\nabla}$, and define

$$\overline{\nabla}_b Y^a = \nabla_b Y^a + C^a{}_{bc} Y^c.$$

  The difference of two connections is tensorial, so we have a tensor

$$C(X, Y) = \overline{\nabla}_X Y - \nabla_X Y$$

  whose components are $C^a{}_{bc}$, and T is symmetric since the torsion vanishes.

- Using the formula for the Christoffel symbols, we can compute

$$C^a{}_{bc} = \frac{1}{\Omega}(\delta^a_b \nabla_c \Omega + \delta^a_c \nabla_b \Omega - g_{bc} g^{ad} \nabla_d \Omega).$$

  Plugging this into the Ricci identity, we get the transformation of the Ricci tensor,

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

- As we've already seen, Weyl transformations preserve null curves. They also preserve null geodesics; consider a null geodesic with affine parameter V such that $\nabla_V V = 0$. When we compute $\overline{\nabla}_V V$, we get extra terms from $V^b V^c C^a{}_{bc}$ which are simply proportional to V, so we have a geodesic with a non-affine parameter.

- In two dimensions, every metric is conformally equivalent to a flat metric. To see this, note that we can always switch to coordinates u and v which are everywhere null, so the metric is proportional to $du\,dv$. By a Weyl transformation, we can set $\overline{g} = 2du\,dv$, which is flat.

- The simplest action for a non-minimally coupled scalar field is

$$S = \int dx \sqrt{-g}\left(\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi) - \frac{\xi}{2}R\phi^2\right)$$

  and in four dimensions, a lengthy calculation shows that it is Weyl invariant when $V = 0$ and $\xi = 1/6$. More generally, for $d \neq 2$ we require $V = 0$ and $\xi = (d-2)/4(d-1)$, while for $d = 2$, ξ can be arbitrary.

- The Maxwell action is also Weyl invariant,

$$S = -\frac{1}{16\pi}\int dx\sqrt{-g}\, g^{\alpha\beta}g^{\mu\nu}F_{\alpha\mu}F_{\beta\nu}.$$

  The easiest way to show this is to write

$$S \propto \int (dA) \wedge \star(dA)$$

  where it's manifest that only the Hodge star is affected by the conformal transformation; its rescaling factor precisely cancels that of A. Alternatively one can see that the theory is scale invariant and the stress-energy tensor is traceless. As an application, we can find how the electromagnetic field behaves in a Friedmann universe by conformal mapping to a flat universe, then mapping back.

- Scale invariance is the case of constant Ω. We see that under a scale transformation,

$$R_{ab} = \overline{R}_{ab}, \quad R = \Omega^2\overline{R}, \quad G_{ab} = \overline{G}_{ab}.$$

  As a result, the vacuum Einstein equation is scale invariant if the cosmological constant is zero.

8.2 Asymptotic Flatness

In this section, we formally define asymptotic flatness.

- Intuitively, an asymptotically flat spacetime is one that looks like Minkowski spacetime at infinity. We would like to regard the Kruskal spacetime as flat, but $i^\pm$ and i^0 are singular. Instead, we base our definition around $\mathcal{I}^\pm$.

- A time-orientable spacetime (M, g) is asymptotically flat at null infinity if there exists a spacetime $(\overline{M}, \overline{g})$ so that

  1. There exists a positive function Ω on M so that $(\overline{M}, \overline{g})$ is an extension of $(M, \Omega^2 g)$.

2. Within $\overline{M}$, M can be extended to obtain a manifold with boundary, $M \cup \partial M$.
3. Ω can be extended to a function on $\overline{M}$ so that $\Omega = 0$ and $d\Omega \neq 0$ on ∂M.
4. ∂M is the disjoint union of two components $\mathcal{I}^+$ and $\mathcal{I}^-$, each diffeomorphic to $\mathbb{R} \times S^2$.
5. No past/future directed causal curve starting in M intersects $\mathcal{I}^+/\mathcal{I}^-$.
6. The $\mathcal{I}^\pm$ are 'complete', as defined below.

The first three conditions just require the existence of an appropriate conformal compactification. The requirement $d\Omega \neq 0$ ensures that the spacetime metric approaches the Minkowski metric at an appropriate rate near $\mathcal{I}^\pm$, and the last three conditions ensure that the infinity has the same structure as in Minkowski spacetime.

- For example, consider the Schwarzschild solution in outgoing EF coordinates (u, r, θ, ϕ), and define $r = 1/x$. Then

$$g = -(1 - 2Mx)du^2 + 2\frac{du dx}{x^2} + \frac{1}{x^2}(d\theta^2 + \sin^2\theta d\phi^2)$$

so choosing a conformal factor $\Omega = x$ gives the unphysical metric

$$\overline{g} = -x^2(1 - 2Mx)du^2 + 2du dx + d\theta^2 + \sin^2\theta d\phi^2$$

which can be smoothly extended across $x = 0$.

- In this case, $\mathcal{I}^+$ corresponds to $r \to \infty$ with finite u, so here it is the surface $x = 0$. It is parametrized by (u, θ, ϕ) and hence diffeomorphic to $\mathbb{R} \times S^2$. Similarly one can do the same for $\mathcal{I}^-$ with the same conformal factor, but in incoming EF coordinates. Thus the Schwarzschild solution is asymptotically flat at null infinity.

- We have not required that $\mathcal{I}^\pm$ be null hypersurfaces; instead we can derive it. We multiply our transformation for R_{ab} by Ω for

$$0 = \Omega R_{ab} = \Omega \overline{R}_{ab} + 2\overline{\nabla}_a \overline{\nabla}_b \Omega + \overline{g}_{ab}\overline{g}^{cd}(\overline{\nabla}_c \overline{\nabla}_d \Omega - 3\Omega^{-1}\partial_c\Omega\partial_d\Omega)$$

where we assume a vacuum solution for simplicity. The first three terms are regular, so the last one must be as well. Then $\overline{g}^{cd}\partial_c\Omega\partial_d\Omega$ vanishes on $\mathcal{I}^\pm$ so $d\Omega$ is null on $\mathcal{I}^\pm$ and normal to it. Then the $\mathcal{I}^\pm$ are null hypersurfaces.

- There is substantial freedom in choosing the coordinates (u, θ, ϕ). One can show by some lengthy arguments that we can write

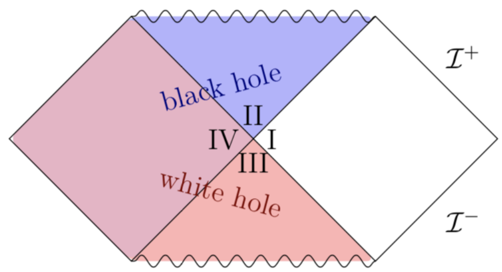
$$\overline{g}|_{\Omega=0} = 2du d\Omega + d\theta^2 + \sin^2 d\phi^2$$

on $\mathcal{I}^+$, with the same for $\mathcal{I}^-$. Finally, one can convert this to inertial frame coordinates (t, x, y, z) so the leading order metric is the Minkowski metric, and corrections are $O(1/r)$. The 'completeness' condition above is that the range of u is $(-\infty, \infty)$.

8.3 Event Horizons and Killing Horizons

Next, we formally define a black hole.

- Let (M, g) be a spacetime that is asymptotically flat at null infinity. The black hole region is $\mathcal{B} = M \setminus (M \cap J^-(\mathcal{I}^+))$ where $J^-(\mathcal{I}^+)$ is defined using the unphysical spacetime $(\overline{M}, \overline{g})$. The future event horizon is the boundary, $\mathcal{H}^+ = \dot{\mathcal{B}} = M \cap \dot{J}^-(\mathcal{I}^+)$. Similarly we define the white hole region as $\mathcal{W} = M \setminus (M \cap J^+(\mathcal{I}^-))$ and the past event horizon as $\mathcal{H}^- = \dot{W} = M \cap \dot{J}^+(\mathcal{I}^-)$.
- Intuitively, the black hole region is the region that cannot send signals to $\mathcal{I}^+$, while the white hole region cannot receive signals from $\mathcal{I}^-$.
- One can easily construct spacetimes with nonempty black hole and white hole regions by deleting points from Minkowski spacetime. To rule out such cases, we focus on spacetimes that are the maximal development of geodesically complete, asymptotically flat initial data.
- The Kruskal spacetime is not asymptotically flat according to our definition, but if we ignore the other null infinities $\mathcal{I}'^{\pm}$, then the spacetime contains a black hole and white hole region.



- It can be shown using our earlier theorems that the event horizons $\mathcal{H}^{\pm}$ are null hypersurfaces. The generators of $\mathcal{H}^+$ cannot have future endpoints, but they can have past endpoints, as seen in spherically symmetric gravitational collapse. The same goes for $\mathcal{H}^-$ in reverse.

Next, we consider some general properties of black holes.

- Unlike bodies in Newtonian gravity, black holes are typically described by a small number of parameters, as stated by no-hair theorems. For example, for gravity and electromagnetism, all stationary, asymptotically flat black hole solutions are fully characterized by mass, electric and magnetic charge, and spin.
- The precise statement of a no-hair theorem depends on the matter content; the theorem above also holds for the Standard Model since electromagnetism is the only long-range field.
- The weak cosmic censorship conjecture roughly states that all singularities are hidden behind event horizons, given suitable generic initial conditions and some energy conditions. Combining this with the singularity theorems implies that event horizons are generic. Note that weak cosmic censorship is not weaker than strong cosmic censorship; the two are logically independent.
- From a field theorist's point of view, general relativity is just an effective field theory. Then cosmic censorship just means we should trust the effective field theory only in its domain of validity: if we had a naked singularity we shouldn't be using general relativity in the first place, as the singularity can produce very heavy particles.

- Our definition of an event horizon is nonlocal, which makes it difficult to use numerically. Alternatively, one can show that trapped surfaces are in $\mathcal{B}$, given suitable assumptions; then $\mathcal{H}^+$ is approximated by taking the boundary of the region where trapped surfaces exist.
- Hawking's area theorem states that, assuming the WEC, weak cosmic censorship, and asymptotic flatness, the area of $\mathcal{H}^+$ is non-decreasing. This is trivial for the Schwarzschild black hole but gives a useful constraint for a rotating black hole, where the area depends on both the mass and the spin. It is violated by Hawking radiation because quantum fields violate the WEC.

Next, we turn to Killing horizons.

- Consider the time translation Killing vector field k^a. We interpret an observer to be 'stationary' if it moves along orbits of k^a, or at least 'stationary according to an observer at infinity'.
- Therefore, the set of points where k^a is null, called the stationary limit surface, bounds regions where it is impossible for an observer to be stationary.
- A related useful concept is a Killing horizon, i.e. a null hypersurface $\mathcal{N}$ where a Killing vector ξ^a is normal to $\mathcal{N}$, which implies that ξ^a is null on $\mathcal{N}$.
- It can be shown under suitable conditions that all event horizons are Killing horizons. For a static spacetime $\xi^a = k^a$, while for a stationary spacetime ξ^a is a combination of k^a and m^a where m^a generates axial rotations.
- Note that the converse is far from true. For example, in Minkowski space the boost generator $x\partial_t + t\partial_x$ is a Killing vector, yielding the Killing horizons $x = \pm t$, even though nothing special happens at these points. By taking linear combinations with translations, every point lies on a Killing horizon.
- The above is an example of a bifurcate Killing horizon, i.e. the intersection of two Killing horizons. At their intersection, the Killing vector must vanish. A bifurcate Killing horizon also appears in the Kruskal spacetime with interaction $U = V = 0$.
- Since $\xi^a \xi_a = 0$ on $\mathcal{N}$, its gradient is normal to $\mathcal{N}$ and hence proportional to ξ_a, so
$$\nabla_a(\xi^b \xi_b)|_{\mathcal{N}} = -2\kappa \xi_a$$
as we've shown earlier. Then, as earlier, we have
$$\xi^b \nabla_b \xi^a|_{\mathcal{N}} = \kappa \xi^a$$
where the proof in this case is shorter since we can use Killing's equation. To fix the normalization, we set $k^a k_a = -1$ at spatial infinity.
- The function κ is called the surface gravity, and it measures the failure of integral curves of ξ^a to be affinely parametrized geodesics. Alternatively, letting $\xi^a = f n^a$ where n^a generates affinely parametrized geodesics, we have
$$\kappa = \xi^a \partial_a \log |f|.$$
Another nice formula that can be shown is
$$\kappa^2 = -\frac{1}{2}(\nabla_a \xi_b)(\nabla^a \xi^b)$$
where we use $\xi_{[a} \nabla_b \xi_{c]} = 0$ since ξ is normal to Σ.

- In the case of a static spacetime, we can physically interpret the surface gravity as the acceleration needed to keep an observer static as measured by a static observer at infinity.
- To see this, we define
$$\xi^a = V u^a$$
where u^a is the four-velocity of a static observer, so $V = \sqrt{-\xi_a \xi^a}$. The four-acceleration is
$$a^a = u^b \nabla_b u^a = \nabla^a \log V$$
where we use $u^a \nabla_b u_a = 0$ since $u^a u_a = -1$, and $\nabla_\xi (\xi^2) = 0$ by Killing's equation.
- Therefore, the magnitude of the acceleration is
$$a = \sqrt{a_a a^a} = V^{-1} \sqrt{\nabla_a V \nabla^a V}$$
and this measures the force felt by the static observer; it diverges as the observer approaches the horizon. This also suggests why Killing horizons and event horizons are related.
- Consider an observer at infinity holding the static observer in place by a long straight rope. The conserved energy of a photon is $E = -p_a \xi^a$, but the measured energy is $E = -p_a u^a$. Thus if the static observer emits a photon, it arrives at infinity redshifted by a factor of V. Now suppose the static observer pulls in the rope by one meter with a force of F. The observer at infinity also sees the rope pulled in by one meter, since both are measuring proper rope length, so by energy conservation the force at infinity must also be redshifted by V.
- Therefore, an observer at infinity pulls with a force per mass
$$V a = \sqrt{\nabla_a V \nabla^a V}$$
and one can show this is equal to κ. The computation is delicate, since we are dealing with products of vanishing and infinite quantities.

9 General Black Holes

9.1 The Reissner–Nordstrom Solution

In this section, we discuss the Reissner–Nordstrom (RN) solution, which describes a charged, spherically symmetric black hole. Such solutions are not very important astrophysically, since most real black holes are neutral, but they are useful theoretical tools.

- The action for the Einstein-Maxwell equation is

$$S = \frac{1}{16\pi} \int dx \sqrt{-g}(R - F^{ab} F_{ab}), \quad F = dA$$

  where the normalization of F differs from particle physics. The Einstein equation is

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

  and the Maxwell equations are the same as usual,

$$\nabla^b F_{ab} = 0, \quad dF = 0.$$

- A generalization of Birkhoff's theorem with a similar proof states that the unique spherically symmetric solution of the Einstein-Maxwell equations with a non-constant area radius function r is the RN solution,

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

  with the potential

$$A = -\frac{Q}{r} dt - P \cos\theta d\phi, \quad e = \sqrt{Q^2 + P^2}.$$

  As we'll show later, M represents the mass, Q represents the electric charge, and P represents the magnetic charge.

- Having a magnetic charge is perfectly acceptable since there is a singularity at $r = 0$. To see that P does represent magnetic charge, note that taking the Hodge dual of F essentially exchanges Q and P, with the factor $d(\cos\theta) = \sin\theta \, d\theta$ accounting for the metric determinant.

- The RN solution is static, with timelike Killing vector $k^a = (\partial/\partial t)^a$, and asymptotically flat at null infinity just like the Schwarzschild solution.

- To simplify the metric we define

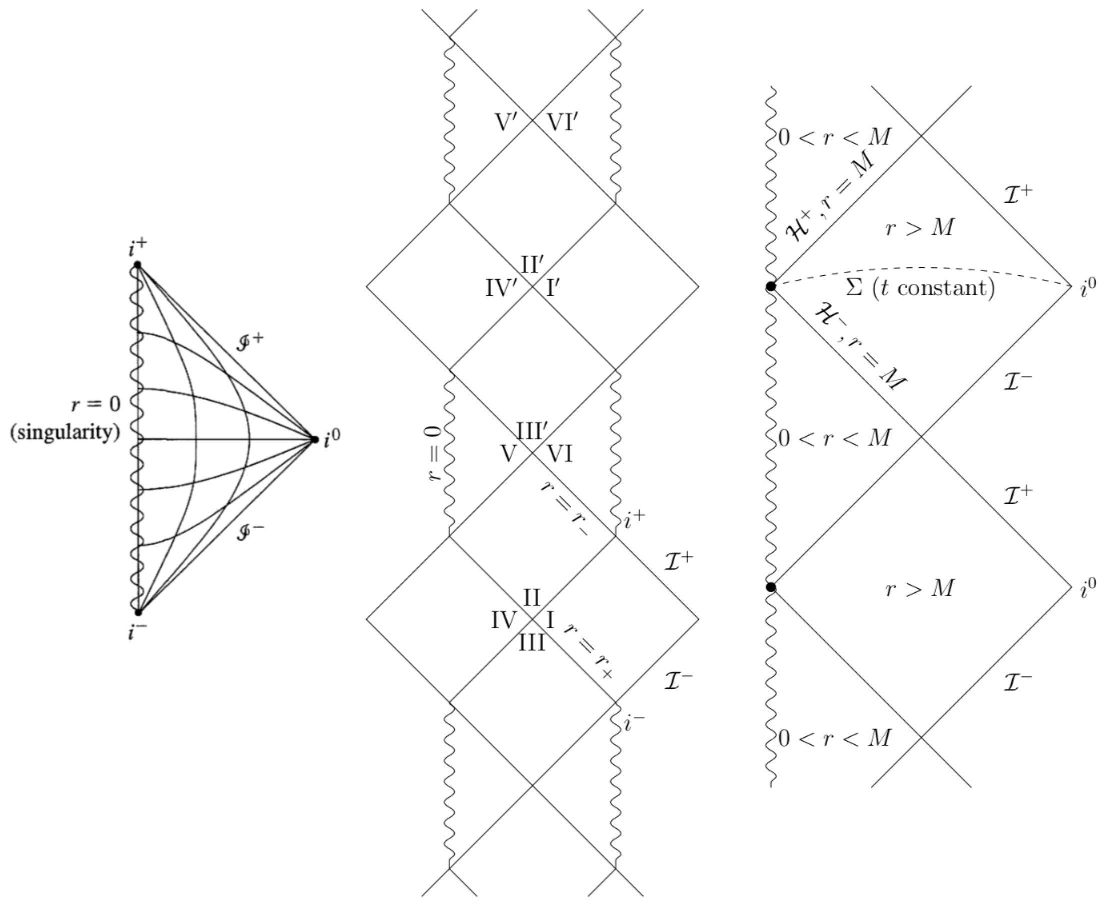
$$\Delta = r^2 - 2Mr + e^2 = (r - r_+)(r - r_-), \quad r_\pm = M \pm \sqrt{M^2 - e^2}$$

  so that

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

  There is a curvature singularity at $r = 0$, as can be checked by computing curvature invariants.

There are three qualitatively different cases: $M < e$, $M > e$, and $M = e$. The Penrose diagrams for these cases are shown below.



- In the case $M < e$, Δ is positive for all $r > 0$ and the metric is smooth down to $r = 0$. This is thus a naked singularity, and dynamical formulation of such a singularity is excluded by cosmic censorship. Unlike in the Schwarzschild solution, the singularity is timelike, not spacelike, so observers don't have to fall into it. In fact, one can check that timelike geodesics *cannot* fall into it, though lightlike geodesics and timelike non-geodesics can.

- Next, we consider the case $M > e$. Then Δ has zeroes at $r = r_\pm > 0$, but there are merely coordinate singularities. To get past them, we use the usual EF trick. In this case we change the radial coordinate, defining

$$dr_* = \frac{r^2}{\Delta}dr, \quad r_* = r + \frac{1}{2\kappa_+}\log\left|\frac{r - r_+}{r_+}\right| + \frac{1}{2\kappa_-}\log\left|\frac{r - r_-}{r_-}\right| + \text{const}, \quad \kappa_\pm = \frac{r_\pm - r_\mp}{2r_\pm^2}.$$

  Switching to null coordinates, we have

$$u = t - r_*, \quad v = t + r_*, \quad ds^2 = -\frac{\Delta}{r^2}dv^2 + 2dvdr + r^2 d\Omega^2$$

  yielding the ingoing null EF coordinates. The metric is smooth for $r > 0$ with a smooth inverse, and we can analytically continue down to $0 < r < r_+$, yielding regions I, II, and V.

- Note that a surface of constant r has normal $n = dr$ and is hence null when $g^{rr} = \Delta/r^2 = 0$. Thus the surfaces $r = r_\pm$ are null hypersurfaces.
- By the same logic as the Schwarzschild case, r decreases along any future-directed causal curve in the region $r_- < r < r_+$. Then no point in the region $r < r_+$ can send a signal to $\mathcal{I}^+$, so there is a black hole region for $r \leq r_+$ and the future event horizon is $r = r_+$. Similarly, using outgoing EF coordinates one finds a white hole.
- Unlike the Schwarzschild case, there is no singularity in regions II or III. Instead, we may start in region II and use ingoing EF coordinates, yielding regions V and VI, where there is a timelike singularity. By switching to Kruskal coordinates, we find region III', which is isometric to region III. Repeating the procedure, we get the infinite conformal diagram shown.

Note. Consider a Cauchy surface Σ that goes through regions I and IV. Then only $D(\Sigma)$, consisting of regions I, II, III, and IV, is determined by the Cauchy data; the Cauchy horizon is $r = r_-$. The spacetime outside $D(\Sigma)$ was determined by analyticity, which is a logically independent assumption that is not necessarily physically relevant. Note that we didn't run into this subtlety with the Schwarzschild spacetime.

Note. The fact that $D(\Sigma)$ is extendible seems to contradict strong cosmic censorship. However, the initial data is not generic. Consider a observer in region I that lives forever, sending signals to an observer that crosses into region II towards region VI. This observer receives an infinite number of signals in a finite proper time, before reaching region VI; this infinite blueshift indicates that a small perturbation in region I becomes large in region II, modifying the singularity structure.

Finally, we consider the extreme RN solution $M = e$.

- In this case, the metric is
$$ds^2 = -\left(1 - \frac{M}{r}\right)^2 dt^2 + \left(1 - \frac{M}{r}\right)^{-2} dr^2 + r^2 d\Omega^2.$$
Using ingoing and outgoing EF coordinates, we can continue to the black hole or white hole region, which contain timelike singularities. Reversing the direction of the EF coordinates allows us to build an infinite Penrose diagram, though we should again take this result with a grain of salt.
- Unlike the previous cases, there is no Einstein–Rosen bridge; instead a surface of constant t is a wormhole of infinite proper length.
- Defining the shifted radial coordinate $\rho = r - M$ and setting $P = 0$, we have
$$ds^2 = -H^{-2}dt^2 + H^2(d\rho^2 + \rho^2 d\Omega^2), \quad H = 1 + \frac{M}{\rho}.$$
In fact, in general we have the Majumdar–Papapetrou solution
$$ds^2 = -H(\mathbf{x})^{-2}dt^2 + H(\mathbf{x})^2(dx^2 + dy^2 + dz^2), \quad A = H^{-1}dt$$
where H is any function obeying $\nabla^2 H = 0$.

- Choosing
$$H = 1 + \sum_{i=1}^{N} \frac{M_i}{|\mathbf{x} - \mathbf{x}_i|}$$
gives a static solution containing N extreme RN black holes; note that these black holes are still spheres, not points, since we're using a shifted radial coordinate. Evidently, the gravitational attraction is perfectly balanced by the electrostatic repulsion.

9.2 The Kerr Solution

First, we describe some uniqueness theorems.

- A spacetime asymptotically flat at null infinity is stationary and axisymmetric if it is stationary with timelike Killing vector k^a, admits a Killing vector m^a that is spacelike near $\mathcal{I}^\pm$ so that $[k, m] = 0$, and m^a generates a one-parameter group of isometries isomorphic to $U(1)$.
- Axisymmetry is a weakening of spherical symmetry. Given these conditions, we may choose coordinates so that $k = \partial/\partial t$ and $m = \partial/\partial\phi$ with $\phi \sim \phi + 2\pi$.
- It can be shown that if (M, g) is stationary, non-static, asymptotically flat, analytic, and suitably regular, then (M, g) is axisymmetric. This is a weaker analogue of Birkhoff's theorem; however, note that it requires analyticity, which is unphysical.
- In any case, if we assume axisymmetry in addition to the other hypotheses above, and assume we are in vacuum, then (M, g) is a member of the Kerr family of solutions parametrized by mass M and angular momentum J. This is an example of a no-hair theorem. Adding charges yields the more general Kerr–Newman solution.
- Note that while the spacetime outside a collapsing spherically symmetric star is Schwarzschild, the spacetime outside a general collapsing star is not Kerr–Newman, because it is not stationary.

Next, we consider the Kerr solution in detail.

- The Kerr metric in Boyer–Lindquist coordinates is
$$\begin{aligned} ds^2 = -\left(1 - \frac{2Mr}{\rho^2}\right) dt^2 - \frac{2Mar\sin^2\theta}{\rho^2}(dtd\phi + d\phi dt) \\ + \frac{\rho^2}{\Delta}dr^2 + \rho^2 d\theta^2 + \frac{\sin^2\theta}{\rho^2}\left((r^2+a^2)^2 - a^2\Delta\sin^2\theta\right)d\phi^2 \end{aligned}$$
where
$$\Delta(r) = r^2 - 2Mr + a^2, \quad \rho^2 = r^2 + a^2\cos^2\theta, \quad a = J/M.$$
The two Killing vectors are ∂_t and ∂_ϕ. There are indeed $dtd\phi$ cross terms, indicating the spacetime is not static.
- Boyer–Lindquist coordinates reduce to Schwarzschild coordinates for $a = 0$, but in general they are not ordinary spherical coordinates. If we take $M \to 0$ with a fixed, then
$$ds^2 = -dt^2 + \frac{(r^2 + a^2\cos^2\theta)^2}{r^2 + a^2}dr^2 + (r^2 + a^2\cos^2\theta)d\theta^2 + (r^2 + a^2)\sin^2\theta d\phi^2$$

which is flat spacetime in ellipsoidal coordinates, related to Cartesian coordinates by

$$x = \sqrt{r^2 + a^2} \sin\theta \cos\phi, \quad y = \sqrt{r^2 + a^2} \sin\theta \sin\phi, \quad z = r\cos\theta.$$

That is, surfaces of constant r are ellipses, not spheres; the surface $r = 0$ is a disc, and $\theta = \pi/2$ is the ring at the boundary of the disc; one can show a curvature singularity occurs here.

- Similarly to the RN solution, we have

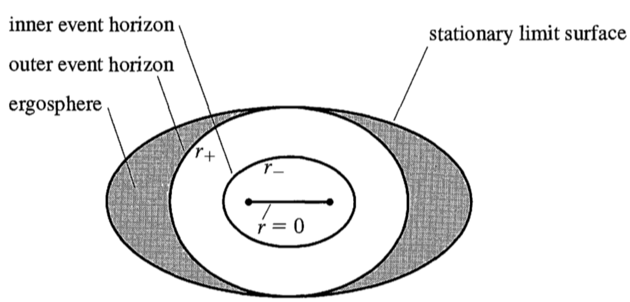
$$\Delta = (r - r_+)(r - r_-), \quad r_\pm = M \pm \sqrt{M^2 - a^2}$$

giving the cases $M > a$, $M = a$, and $M < a$. The case $M < a$ is similar to the $M < e$ case of the RN solution, with a naked singularity, while $M = a$ is unstable. Thus we focus on the case $M > a$.

- The surfaces $r = r_\pm$ are both null hypersurfaces and event horizons. We have coordinate singularities at both event horizons, but we can define analogues of incoming and outgoing EF coordinates to extend through them.

- The maximal analytic extension is similar to RN, with the exception that we can go through the ring singularity at $r = 0$ to reach a region described by the Kerr metric with $r < 0$, where closed timelike curves exist. As argued earlier, these regions are unphysical as they are unstable with respect to perturbations outside the black hole. Note that we can't draw a standard Penrose diagram since we lack spherical symmetry, but we can restrict to $\theta = 0$ and draw a Penrose diagram for that two-dimensional spacetime.

- Next, we turn to the Killing horizons. The Killing horizon for $k = \partial_t$ is the stationary limit surface, which is not an event horizon.



Instead, inside the ergosphere, an observer must move with the rotation of the black hole; this is an extreme example of frame dragging.

- The inner and outer event horizons are Killing horizons for

$$k + \Omega_\pm m, \quad \Omega_\pm = \frac{a}{2Mr_\pm} = \frac{a}{r_\pm^2 + a^2}.$$

More generally, every point inside the ergosphere is on some Killing horizon; the Killing horizon for $k + \Omega m$ marks the boundary where it is impossible to rotate with angular velocity less than Ω. We interpret $\Omega_+ = \Omega_H$ as the angular velocity of the black hole. More explicitly, it is the angular velocity of a photon at $r = r_+$ moving directly against the rotation of the black hole.

9.3 Mass, Charge, and Spin

So far, we haven't defined the mass, charge, or angular momentum of our black hole solutions. For an asymptotically flat end we define the electric and magnetic charges as

$$Q = \frac{1}{4\pi} \lim_{r \to \infty} \int_{S_r^2} \star F, \quad P = \frac{1}{4\pi} \lim_{r \to \infty} \int_{S_r^2} F$$

where S_r^2 is a sphere of radius r. This is in agreement with our earlier definition of electric charge.

Example. The Coulomb potential is $A = -(q/r)dt$, so

$$F = -\frac{q}{r^2} dt \wedge dr.$$

Taking the Hodge dual gives

$$(\star F)_{\theta\phi} = r^2 \sin\theta \, F^{tr} = q \sin\theta$$

where we used $g = r^2 \sin\theta$, and the charge is

$$Q = \frac{1}{4\pi} \int d\theta d\phi \, q \sin\theta = q$$

as expected. By similar reasoning, the Kerr solution has electric charge Q and magnetic charge P.

Next, we define the Komar mass.

- For a stationary spacetime, we can define a conserved energy-momentum current

  $$J_a = -T_{ab} k^b, \quad d \star J = 0$$

  where k is a timelike Killing vector, and the conservation of J follows from the conservation of the stress-energy tensor and Killing's equation.

- Then we can define the total energy on a spacelike hypersurface Σ as

  $$E[\Sigma] = -\int_\Sigma \star J.$$

  This is conserved, i.e. if Σ and Σ' bound a spacetime region R then

  $$E[\Sigma'] - E[\Sigma] = -\int_{\partial R} \star J = -\int_R d \star J = 0.$$

  If we had $\star J = dX$ for some two-form X, then we could write $E[\Sigma]$ as an integral over $\partial\Sigma$ as we did for charge and evaluate it at infinity; however, this is not possible.

- On the other hand, we have

  $$(\star d \star k)_a = -\nabla^b (dk)_{ab} = -\nabla^b \nabla_a k_b + \nabla^b \nabla_b k_a = 2\nabla^b \nabla_b k_a = -2R_{ab} k^b = 8\pi J'_a$$

  where we used Killing's equation and $\nabla_a \nabla_b k^c = R^c{}_{bad} k^d$, and by Einstein's equation

  $$J'_a = -2\left(T_{ab} - \frac{1}{2} T g_{ab}\right) k^b.$$

- The current J' is similar to J, but we now have

$$d \star dk = 8\pi \star J'$$

so $\star J'$ is exact; also note this implies J' is conserved. We thus define the Komar mass

$$M_{\mathrm{Komar}} = -\frac{1}{8\pi} \lim_{r\to\infty} \int_{S_r^2} \star dk.$$

That is, the Killing vector k itself serves as the analogue of the potential A. Physically, the Komar mass measures the total energy of the spacetime, including both the matter and the gravitational field, while our earlier naive definition measured the energy of the matter alone.

- Why is the Komar mass 'really' the energy? We can verify it works as expected in the Newtonian limit and for the Schwarzschild spacetime, and it is conserved during gravitational collapse; thus it should be the right expression for the energy of a black hole.

- Note that we used nothing in the definition of the Komar mass besides the Killing property; thus for an axisymmetric spacetime, we can define the angular momentum by

$$J_{\mathrm{Komar}} = \frac{1}{16\pi} \lim_{r\to\infty} \int_{S_r^2} \star dm$$

where m is the Killing vector that generates rotations about the axis of symmetry, and the proportionality constant is fixed by the Newtonian limit.

Example. For the Schwarzschild solution, the Killing vector is ∂_t, so

$$k = -\left(1 - \frac{2M}{r}\right) dt, \quad dk = \frac{2M}{r^2} dt \wedge dr, \quad \star dk = -2M \sin\theta \, d\theta \wedge d\phi.$$

Integrating, the Komar mass is M as expected.

Note. How can the Kerr solution have nonzero charge if the charge density is zero everywhere? A surface of constant t is asymptotically flat with two ends; the charges on each end are opposite. We simply have a given flux going through a wormhole. One could also take a spacelike slice with one end, which instead includes the singularity; in that case the charge density would be singular at the singularity; in either case we can get a nonzero result.

The same story goes for the Komar mass, since the Ricci tensor is zero everywhere. For instance, in the Schwarzschild solution, the two ends of a surface of constant t have opposite Komar masses, because the Killing vector in region IV points the opposite way as it does in region I.

The Komar mass is only defined for stationary spacetimes. We may instead define the energy as the value of the Hamiltonian in a Hamiltonian formulation of GR, and hence define the ADM mass.

- For simplicity, we work in vacuum and set $16\pi G = 1$. We perform a $3+1$ decomposition of spacetime with lapse function N and shift vector N^i,

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

In terms of these variables, the Einstein-Hilbert action is

$$S = \int dt d^3x \, \sqrt{h} N \left(R^{(3)} + K_{ij}K^{ij} - K^2\right)$$

where $R^{(3)}$ is the Ricci scalar of h_{ij} and K^{ij} is the extrinsic curvature of a constant t surface,

$$K_{ij} = \frac{1}{2N}\left(\dot{h}_{ij} - D_i N_j - D_j N_i\right)$$

and the dot denotes a t-derivative.

- We then switch to the Hamiltonian formulation in the usual way, by identifying canonical momenta and performing a Legendre transformation. The conjugate moment of N and N^i vanish, indicating that they are not dynamical, while

$$\pi^{ij} = \frac{\delta S}{\delta \dot{h}_{ij}} = \sqrt{h}(K^{ij} - K h^{ij}).$$

The Hamiltonian is defined as

$$H = \int d^3x\, \pi^{ij}\dot{h}_{ij} - \mathcal{L}.$$

If we naively integrate by parts, we find that the Hamiltonian vanishes identically on-shell.

- The problem is that we neglected boundary terms. In a closed universe, there is no boundary and the total energy of the universe is indeed zero. But in general we must add a surface term, and it is the ADM energy

$$E_{\mathrm{ADM}} = \frac{1}{16\pi}\lim_{r\to\infty}\int_{S_r^2} dA\, n_i(\partial_j h_{ij} - \partial_i h_{jj})$$

where n^i is the unit outward normal and we restored $G = 1$.

- In general, there is a separate ADM energy for each asymptotic end, just like for the Komar mass. It can be shown that if the surfaces of constant t are orthogonal to the timelike Killing vector as $r \to \infty$, the ADM energy and Komar mass agree.

- We may also define the ADM 3-momentum

$$P_i = \frac{1}{8\pi}\lim_{r\to\infty}\int_{S_r^2} dA\,(K_{ij}n_j - K n_i).$$

We then define the ADM mass by

$$M_{\mathrm{ADM}} = \sqrt{E_{\mathrm{ADM}}^2 - P_i P_i}.$$

- In 1979, Schoen and Yau proved the 'positive energy theorem': for any geodesically complete asymptotically flat initial data obeying the DEC, $E_{\mathrm{ADM}} \geq \sqrt{P_i P_i}$ with equality only for Minkowski space.

- In the Schwarzschild spacetime, $E_{\mathrm{ADM}} = M$, so the ADM energy is negative for $M < 0$. However, in this case a surface of constant t is singular, i.e. not geodesically complete.

9.4 Black Hole Mechanics

We begin with the example of the Penrose process.

- Consider a particle approaching a Kerr black hole along a geodesic, so that $E = -k \cdot p$ is conserved. If the particle decays at a point, p is conserved, so E is conserved. Similarly, $L = m \cdot p$ is conserved.

- Inside the ergosphere, E can be negative. Hence it is possible for a particle to emit a negative energy particle within the ergosphere. That particle falls into the black hole, reducing its energy and angular momentum, while the original particle leaves with more energy than it entered with; this is the Penrose process. In the context of photons, it is called superradiance.

- To understand the constraints on the Penrose process, note that since k is a future-directed causal vector outside the ergosphere, we must have $E \geq 0$ for any particle there to be 'going forward in time'. Similarly, the most restrictive constraint on a particle just outside the outer event horizon is from $\xi = k + \Omega_H m$, which gives

$$E \geq \Omega_H L.$$

  That is, the energy of a particle can't be too negative, or else it can't fall in.

- Next, define the irreducible mass of the black hole,

$$M_{\mathrm{irr}}^2 = \frac{1}{2}(M^2 + \sqrt{M^4 - J^2}).$$

  Then it is straightforward to check that

$$\delta M_{\mathrm{irr}} \propto \Omega_H^{-1} \delta M - \delta J \geq 0.$$

  Therefore, we can use the Penrose process to reduce a black hole's mass to M_{irr}, reducing its angular momentum to zero in the process. Here we are assuming that the black hole settles back down to a Kerr solution.

- This result is simple in terms of the area of the horizon $r = r_+$. Pulling back the metric to the horizon, i.e. setting $\Delta = dt = dr = 0$,

$$\gamma_{ij} dx^i dx^j = (r_+^2 + a^2 \cos^2\theta) d\theta^2 + \frac{(r_+^2 + a^2)^2 \sin^2\theta}{r_+^2 + a^2\cos^2\theta} d\phi^2$$

  and the area is

$$A = \int \sqrt{|\gamma|} d\theta d\phi = \int (r_+^2 + a^2) \sin\theta \, d\theta d\phi = 4\pi (r_+^2 + a^2) = 16\pi M_{\mathrm{irr}}^2 = 8\pi (M^2 + \sqrt{M^4 - J^2}).$$

  Therefore, we have shown that $\delta A \geq 0$.

- A similar procedure can be carried out for RN, decreasing the mass and charge. For a charged particle, $p \cdot \partial_t$ is not conserved because of the electromagnetic force, but the total energy $(p - qA) \cdot \partial_t$ is. Using this, we find an ergosphere outside the event horizon where negatively charged particles can have negative energy.

Example. The surface gravity of the Schwarzschild black hole. From the metric, we read off

$$K^\mu = \partial_t, \quad u^\mu = \left(1 - \frac{2M}{r}\right)^{-1/2}, \quad V = \sqrt{1 - 2M/r}$$

so the redshift factor indeed diverges at the horizon. The four-acceleration is

$$a_\mu = \frac{M}{r^2(1-2M/r)} \nabla_\mu r, \quad a = \frac{M}{r^2\sqrt{1-2M/r}}.$$

The surface gravity is thus

$$\kappa = Va|_{r=2M} = \frac{1}{4M}.$$

Note that it becomes smaller as the black hole gets larger. More generally, in the Kruskal spacetime we have a future event horizon $\mathcal{H}^+$ at $U = 0$ and a past event horizon $\mathcal{H}^-$ at $V = 0$, with surface gravity $1/4M$ and $-1/4M$ respectively. This is an example of a bifurcate Killing horizon, as they intersect on the two-sphere $U = V = 0$ where the Killing vector vanishes.

Next, we state the laws of black hole mechanics.

- By a similar computation to above, the Kerr black hole has surface gravity

$$\kappa = \frac{\sqrt{M^2 - a^2}}{2M(M + \sqrt{M^2 - a^2})}$$

  so the change in the mass is simply

$$\delta M = \frac{\kappa}{8\pi} \delta A + \Omega_H \delta J.$$

  Note that we proved this by assuming the perturbed black hole settled back down to a Kerr solution. In fact, this assumption is not necessary; this result holds for any small perturbation of the Kerr metric, as proven by Sudasky and Wald in 1992, where δM and δJ are the energy and angular momentum of the matter crossing the event horizon.

- We notice that this looks similar to the first law of thermodynamics,

$$dE = TdS + \mu dJ$$

  where μ is a chemical potential for angular momentum. This motivates the identifications

$$E \leftrightarrow M, \quad T \leftrightarrow \kappa/2\pi, \quad S \leftrightarrow A/4, \quad \mu \leftrightarrow \Omega_H.$$

  Here, the normalization of T is set because we know a Schwarzschild black hole indeed radiates at temperature $\kappa/2\pi$, as shown below.

- In the case of a charged black hole, we pick up the term $-\Phi_H \delta Q$, where Φ_H is the electrostatic potential difference between the event horizon and infinity.

- The zeroth law of thermodynamics states that in thermal equilibrium, T is constant throughout a system. Similarly, the zeroth law of black hole mechanics states that the future event horizon of a stationary black hole obeying the DEC has constant κ.

- The second law of black hole mechanics is $\delta A \geq 0$, as shown by Hawking's area theorem. When we account for black hole evaporation, A can decrease while the entropy of matter increases; hence we interpret A as a genuine entropy with the generalized second law $\delta(S_{\text{matter}} + A/4) \geq 0$.

Example. Consider two distant Schwarzschild black holes of masses M_1 and M_2. Then if they merge to a black hole which eventually settles down to a Schwarzschild black hole of mass M, then the second law gives

$$M \geq \sqrt{M_1^2 + M_2^2}$$

placing a limit on the amount of energy that can be carried away by gravitational waves.

Example. Consider an asymptotically flat initial data set with ADM energy E_i and apparent horizon area A_{app}, which settles into a Kerr black hole with parameters M and J. Then we have

$$A_{\text{app}} \leq A_i \leq 8\pi \left(M_f^2 + \sqrt{M_f^4 - J_f^2} \right) \leq 16\pi M_f^2$$

where A_i is the initial horizon area. Since gravitational waves can only carry away energy, $M_f \leq E_i$. Then we have the Penrose inequality $A_{\text{app}} \leq 16\pi E_i^2$, containing only quantities which can be directly computed from the initial data. It has been proven for time-symmetric initial data and serves as a test of weak cosmic censorship.

Note. Suppose we surround a Kerr black hole with a mirror. Then we expect a photon to continually fall through, getting more and more energy with every pass, creating a "black hole bomb". It is difficult to realize this in reality, because there isn't anything strong enough to act as the mirror.

However, the situation changes when we apply quantum mechanics. Massive fields such as the axion can have bound states around black holes, which look like hydrogen bound states at large radii. The black hole can be unstable to decay into such states with angular momentum, spinning down the black hole and building up a condensate of particles outside.

One prediction this makes is that we don't expect to see any black holes with spin above a certain value depending on the mass, which can be tested by LIGO, though measuring black hole spin is difficult. (The black hole doesn't spin down all the way by successively higher angular momentum modes, as they are exponentially suppressed near the black hole by the angular momentum barrier, so the rates are slow. There are also constraints from the field not coupling too strongly to the accretion disk about the black hole, though these wouldn't be a problem for something coupling purely gravitationally since the accretion disk mass isn't too big.) Another prediction is continuous gravitation wave emission at a constant frequency by the rotation of the condensate, which is possible to see since the radius is larger. Both of these should be probed by LIGO in the next decade, putting constraints on light (10^{-12} eV) axion-like particles.

10 Quantum Field Theory in Curved Spacetime

10.1 Flat Spacetime

First, we review some features of quantum fields.

- Consider a simple harmonic oscillator with unit mass,
$$\ddot{q} + \omega^2 q = 0.$$
We define a ground state to be a lowest energy state; classically it is $q(t) = 0$.
- This is impossible in quantum mechanics because the canonical commutators
$$[\hat{q}(t), \hat{p}(t)] = [\hat{q}(t), \hat{\dot{q}}(t)] = i$$
could not hold, where the operators are in Heisenberg picture. Instead we have
$$\psi(q) \propto e^{-\omega q^2/2}$$
and $\delta q \sim 1/\sqrt{\omega}$.
- For a free scalar field, the situation is similar. For a field in a box of volume V,
$$[\hat{\phi}_{\mathbf{k}}(t), \hat{\pi}_{\mathbf{k}'}(t)] = i\delta_{\mathbf{k},-\mathbf{k}'}, \quad \phi(\mathbf{x}, t) = \frac{1}{\sqrt{V}} \sum_{\mathbf{k}} \phi_{\mathbf{k}}(t) e^{i\mathbf{k}\mathbf{x}}$$
along with Hamiltonian and equation of motion
$$H = \frac{1}{2} \sum_{\mathbf{k}} |\dot{\phi}_{\mathbf{k}}|^2 + \omega_k^2 |\phi_{\mathbf{k}}|^2, \quad \omega_k = \sqrt{k^2 + m^2}, \quad \ddot{\phi}_{\mathbf{k}} + (k^2 + m^2)\phi_{\mathbf{k}} = 0.$$
The classical vacuum has field value zero, and the quantum vacuum has the wavefunctional
$$\Psi[\phi] = \exp\left(-\frac{1}{2} \sum_{\mathbf{k}} \omega_k |\phi_{\mathbf{k}}|^2\right).$$
- When we take the infinite volume limit, we have
$$\sum_{\mathbf{k}} \to V \int đ\mathbf{k}, \quad \phi_{\mathbf{k}} \to \sqrt{\frac{(2\pi)^3}{V}} \phi_{\mathbf{k}}$$
so the vacuum wavefunctional becomes
$$\Psi[\phi] = \exp\left(-\frac{1}{2} \int đ\mathbf{k}\, |\phi_{\mathbf{k}}|^2 \omega_k\right), \quad \phi(\mathbf{x}) = \int đ\mathbf{k}\, \phi_{\mathbf{k}} e^{i\mathbf{k}\mathbf{x}}.$$
- As an application, consider the typical values of the field averaged over a box R of volume L^3,
$$\phi_L = \frac{1}{L^3} \int_R d\mathbf{x}\, \phi(\mathbf{x}) \sim \int đ\mathbf{k}\, \phi_{\mathbf{k}} \operatorname{sinc}(k_x L) \operatorname{sinc}(k_y L) \operatorname{sinc}(k_z L).$$
All of the $\phi_{\mathbf{k}}$ fluctuate independently, and the dominant part of the integral comes from the region $kL \lesssim 1$, so
$$\delta\phi_L \sim \sqrt{(\delta\phi_k)^2 k^3}, \quad k = 1/L.$$
In particular, this quantity is divergent for $L \to 0$.

Next, we cover some general philosophy.

- We will consider only free quantum fields; their only coupling is to classical background fields such as an electric or gravitational field, which can induce particle creation.
- Naively, this is sufficient at energy scales much lower than the Planck scale, but since gravity couples to everything, including gravitational energy, any situation where a gravitational field induces, e.g. photon emission will also induce graviton emission, which will be just important.
- Thus, the next level of approximation is to consider linearized gravity, where we also include a quantum gravitational field as a perturbation on a background classical field. This must break down at the Planck scale because of the nonrenormalizability of gravity, but far below the Planck scale we can treat it as an effective field theory, truncated at, say, the one-loop level. Since a loop expansion is also an expansion in $\hbar$, this is "semiclassical" gravity.
- A static electric field can create electron-positron pairs in the Schwinger effect. Heuristically, a virtual particle pair is produced. If the particles move a distance ℓ apart, they harvest energy ℓeE from the electric field, and if $\ell eE \geq 2m_e$ the particles can become real. The probability of separation at distance ℓ is $P \sim \exp(-m_e \ell)$ since $1/m_e$ is the Compton wavelength, so

  $$P \sim \exp\left(-\frac{m_e^2}{eE}\right).$$

  This effect may be soon observable in some experiments.
- More rigorously, we see this is a quantum tunneling effect and apply the WKB approximation; the probability we found above is a typical example of a WKB result. Note that it is nonperturbative in the field strength.
- We can apply the same heuristic argument to gravity, where $\mathbf{g}$ takes the role of $\mathbf{E}$. However, we don't expect pair creation in a constant gravitational field, because both particles in the pair fall the same way and cannot separate; this is in accordance with the equivalence principle. But we can have pair creation in a nonuniform field, and in situations where there is an event horizon, as one particle falls in and the other escapes.

Example. A second functional derivative. Given

$$S[q] = \int dt\, \frac{1}{2}(\dot{q}^2 - \omega^2 q^2)$$

we have

$$\frac{\delta S}{\delta q(t_1)} = -\ddot{q}(t_1) - \omega^2 q(t_1) = -\int dt\, (\ddot{q}(t) + \omega^2 q(t))\delta(t - t_1).$$

Therefore, the second functional derivative is

$$\frac{\delta^2 S}{\delta q(t_2)\delta q(t_1)} = -\delta''(t_2 - t_1) - \omega^2 \delta(t_2 - t_1).$$

As a simple example, we'll consider a driven harmonic oscillator.

- We take a Hamiltonian with driving force for $t \in [0, T]$,

$$H(p,q) = \frac{p^2}{2} + \frac{\omega^2 q^2}{2} - J(t)q, \quad \dot{q} = p, \quad \dot{p} = -\omega^2 q + J(t).$$

Upon canonical quantization in Heisenberg picture, $\hat{q}$ and $\hat{p}$ satisfy the same equations, and we drop the hats.

- We define the instantaneous creation and annihilation operators

$$a^-(t) = \sqrt{\frac{\omega}{2}}\left(q(t) + \frac{i}{\omega}p(t)\right), \quad a^+(t) = \sqrt{\frac{\omega}{2}}\left(q(t) - \frac{i}{\omega}p(t)\right)$$

which are Hermitian conjugates, and obey the commutation relations

$$[a^-(t), a^+(t)] = 1.$$

- The resulting equation of motion for a^- is

$$\dot{a}^- = -i\omega a^- + \frac{i}{\sqrt{2\omega}}J(t), \quad a^-(t) = \left(a^-_{\text{in}}(t) + \frac{i}{\sqrt{2\omega}}\int_0^t e^{i\omega t'} J(t')\, dt'\right) e^{-i\omega t}$$

with conjugate equations for a^+. Changing variables in the Hamiltonian,

$$H = \frac{\omega}{2}(2a^+a^- + 1) - \frac{a^+ + a^-}{\sqrt{2\omega}}J(t).$$

- We define the 'in' and 'out' operators so that

$$a^-(t) = \begin{cases} a^-_{\text{in}} e^{-i\omega t} & t < 0, \\ a^-_{\text{out}} e^{-i\omega t} & t > T, \end{cases} \qquad H = \begin{cases} \omega(a^+_{\text{in}} a^-_{\text{in}} + 1/2) & t < 0, \\ \omega(a^+_{\text{out}} a^-_{\text{out}} + 1/2) & t > T. \end{cases}$$

with similar expressions for the creation operators. They are related by

$$a^-_{\text{out}} = a^-_{\text{in}} + J_0, \quad a^+_{\text{out}} = a^+_{\text{in}} + J_0^*, \quad J_0 = \frac{i}{\sqrt{2\omega}}\int_0^T e^{i\omega t'} J(t')\, dt'.$$

- Next, we may construct 'in' states

$$a^-_{\text{in}}|0_{\text{in}}\rangle = 0, \quad |n_{\text{in}}\rangle = \frac{1}{\sqrt{n!}}(a^+_{\text{in}})^n |0_{\text{in}}\rangle$$

with similar expressions for the 'out' states. Physically, the 'in' vacuum is the state of lowest energy before the driving starts, while the 'out' vacuum is the state of lowest energy after the driving ends. Thus, for example, the amplitude $\langle 2_{\text{out}}|1_{\text{in}}\rangle$ is the amplitude to go from one particle before the driving to two particles after the driving. This is a bit confusing since it's said that in Heisenberg picture the states are time-independent; a better picture is that every state extends *through* time.

- Note in particular that

$$a^-_{\text{out}}|0_{\text{in}}\rangle = J_0|0_{\text{in}}\rangle$$

where J_0 is a number. Therefore, if we start in the ground state, after the driving we have a coherent state with mean occupancy J_0,

$$|0_{\text{in}}\rangle = e^{-|J_0|^2/2}\sum_{n=0}^{\infty}\frac{J_0^n}{\sqrt{n!}}|n_{\text{out}}\rangle.$$

- We can also compute the final energy of the 'in' vacuum. For $t > T$,

$$\langle 0_{\text{in}}|H(t)|0_{\text{in}}\rangle = \langle 0_{\text{in}}|\omega\left(\frac{1}{2} + a^+_{\text{out}}a^-_{\text{out}}\right)|0_{\text{in}}\rangle = \left(\frac{1}{2} + |J_0|^2\right)\omega.$$

  As for the position, for $t > T$,

$$\langle 0_{\text{in}}|q(t)|0_{\text{in}}\rangle = \frac{1}{\sqrt{2\omega}}(J_0 e^{-i\omega t} + J_0^* e^{i\omega t})$$

  which we can write in terms of the retarded Green's function

$$q(t) = \int dt'\, J(t')G_{\text{ret}}(t,t'), \quad G_{\text{ret}}(t,t') = \frac{\sin\omega(t-t')}{\omega}\theta(t-t').$$

  Similarly we may define the advanced Green's function, in the out vacuum.

- The Feynman Green's function goes from the in vacuum to the out vacuum,

$$\frac{\langle 0_{\text{out}}|q(t)|0_{\text{in}}\rangle}{\langle 0_{\text{out}}|0_{\text{in}}\rangle} = \int dt'\, G_F(t,t')J(t'), \quad G_F(t,t') = \frac{i}{2\omega}e^{-i\omega|t-t'|}$$

  and is the Green's function that satisfies the boundary conditions

$$G_F(t,t') \to e^{-i\omega t} \text{ for } t \to \infty, \quad G_F(t,t') \to e^{i\omega t} \text{ for } t \to -\infty.$$

  Generally this Green's function is useful for computing vacuum-to-vacuum transition functions. It is symmetric in its arguments.

- Finally, the Euclidean Green's function appears in Euclidean time, with the boundary conditions $\lim_{\tau\to\pm\infty} G_E(\tau,\tau') = 0$. Then we have

$$G_E(\tau,\tau') = \frac{1}{2\omega}\, e^{-\omega|\tau-\tau'|}$$

  and the Euclidean and Feynman Green's functions are related by analytic continuation; the Feynman boundary conditions turn into exponential decay on both ends. Since path integrals are 'really' in Euclidean space, this shows why the Feynman Green's function is so ubiquitous.

Note. This example was simpler than the examples we will see below. More generally, we will be dealing with fields, and hence have to consider how the field modes depend on space. There are additional subtleties even for a single degree of freedom. Above, we defined instantaneous creation and annihilation operators, and hence could have defined instantaneous vacua and particles. But more generally, this might not be possible. In other situations, there might not be clear "in" and "out" regions.

Next, we turn to the mode expansion of a free real scalar field.

- We know from basic quantum field theory that the Heisenberg picture field is

$$\phi(x) = \int d\!\!\!{}^-\mathbf{k}\frac{1}{\sqrt{2\omega_k}}\left(e^{-i\omega_k t+i\mathbf{k}\cdot\mathbf{x}}a^-_{\mathbf{k}} + e^{i\omega_k t-i\mathbf{k}\cdot\mathbf{x}}a^+_{\mathbf{k}}\right), \quad \omega_k^2 = k^2 + m^2.$$

  where the creation and annihilation operators are time-independent.

- However, let us instead postulate a more general expansion, in terms of mode functions $v_{\mathbf{k}}(t)$,

$$\phi(x) = \int đ\mathbf{k} \frac{1}{\sqrt{2}} \left(v_{\mathbf{k}}^*(t) e^{i\mathbf{k}\cdot\mathbf{x}} a_{\mathbf{k}}^- + v_{\mathbf{k}}(t) e^{-i\mathbf{k}\cdot\mathbf{x}} a_{\mathbf{k}}^+ \right).$$

  We want to know which functions $v_{\mathbf{k}}(t)$ are allowed, which preserve the commutation relations

$$[a_{\mathbf{k}}^-, a_{\mathbf{k}'}^+] = \delta(\mathbf{k} - \mathbf{k}')$$

  with all other commutators zero. In this case, the associated operators $a_{\mathbf{k}}^{\pm}$ can be interpreted as creation and annihilation operators for physical particles.

- The equation of motion for the field is the Klein–Gordan equation, which implies

$$\ddot{v}_{\mathbf{k}} + \omega_k^2 v_{\mathbf{k}} = 0, \quad v_{\mathbf{k}}(t) = \frac{1}{\sqrt{\omega_k}} (\alpha_{\mathbf{k}} e^{i\omega_k t} + \beta_{\mathbf{k}} e^{-i\omega_k t}).$$

  In other words, the mode functions obey the same equations of motion as classical field modes. Note that expanding a time-dependent field operator in terms of time-independent creation and annihilation operators times mode functions is a nontrivial requirement: it only works if the equations of motion are linear, because then $\phi(t)$ and $\pi(t)$ remain in the subspace spanned by the creation and annihilation operators.

- Next, the canonical momentum takes the form

$$\pi(y) = \frac{\partial \phi}{\partial t} = \int đ\mathbf{k} \frac{1}{\sqrt{2}} \left(\dot{v}_{\mathbf{k}}^*(t) e^{i\mathbf{k}\cdot\mathbf{x}} a_{\mathbf{k}}^- + \dot{v}_{\mathbf{k}}(t) e^{-i\mathbf{k}\cdot\mathbf{x}} a_{\mathbf{k}}^+ \right).$$

  This implies that the creation and annihilation operators' commutation relations are compatible with the canonical commutation relations precisely when

$$\dot{v}_{\mathbf{k}}(t) v_{\mathbf{k}}^*(t) - v_{\mathbf{k}}(t) \dot{v}_{\mathbf{k}}^*(t) = 2i.$$

  Note that this is simply the Wronskian of $v_{\mathbf{k}}$ and $v_{\mathbf{k}}^*$, and hence is automatically time-independent. The resulting constraint is

$$|\alpha_{\mathbf{k}}|^2 - |\beta_{\mathbf{k}}|^2 = 1$$

  which is not sufficient to determine the coefficients alone.

- Next we consider the Hamiltonian, which turns out to be

$$H = \int đ\mathbf{k}\, \omega_k \left(\alpha_{\mathbf{k}}^* \beta_{\mathbf{k}}^* a_{\mathbf{k}}^- a_{-\mathbf{k}}^- + \alpha_{\mathbf{k}} \beta_{\mathbf{k}} a_{\mathbf{k}}^+ a_{-\mathbf{k}}^- + (|\alpha_{\mathbf{k}}|^2 + |\beta_{\mathbf{k}}|^2) a_{\mathbf{k}}^+ a_{\mathbf{k}}^- \right)$$

  where we removed an infinite constant, and used $\alpha_{\mathbf{k}} = \alpha_{-\mathbf{k}}$ and $\beta_{\mathbf{k}} = \beta_{-\mathbf{k}}$ since the field is real. But then the vacuum itself, defined by $a_{\mathbf{k}}^- |0\rangle = 0$, is only an eigenvector of H if $\alpha_{\mathbf{k}} \beta_{\mathbf{k}} = 0$.

- Thus we must have

$$\alpha_{\mathbf{k}} = e^{i\delta_{\mathbf{k}}}, \quad \beta_{\mathbf{k}} = 0$$

  and we can set the phases to zero by suitable redefinitions. Thus

$$v_k(t) = \frac{1}{\sqrt{\omega_k}} e^{i\omega_k t}$$

  and we recover our usual result; there is no freedom to redefine particles in flat spacetime.

- It's useful to impose a little more mathematical structure. We define the inner product

$$(\phi_1, \phi_2) = -i \int_{\Sigma_t} (\phi_1 \partial_t \phi_2^* - \phi_2^* \partial_t \phi_1) \, d^{n-1}x$$

  in n-dimensional Minkowski space, where Σ_t is a constant-time hypersurface. Note that we are considering *complex* solutions; if we quantize a real scalar field we impose reality at the operator level, not on the modes themselves. The Wronskian we used above is just the same thing, but specialized for modes with no space dependence.

- In general, we define a creation/annihilation operator associated with a mode f by

$$a(f) = (f, \phi), \quad a(f)^\dagger = -(\overline{f}, \phi)$$

  and as a result, we have

$$[a(f), a(g)^\dagger] = (f, g)$$

  with all other commutators zero.

- Then our quantum field has the form

$$\phi(x) = \int d\!\!{}^-\mathbf{k} \, (a_{\mathbf{k}}^- f_{\mathbf{k}}(x) + a_{\mathbf{k}}^+ f_{\mathbf{k}}^*(x))$$

  where the modes $f_{\mathbf{k}}$ are orthonormal under the inner product,

$$(f_{\mathbf{k}_1}, f_{\mathbf{k}_2}) = \delta(\mathbf{k}_1 - \mathbf{k}_2).$$

  The complex conjugate modes $f_{\mathbf{k}}^*$ are orthogonal to the $f_{\mathbf{k}}$ modes and are orthonormal to each other with a negative norm, and the creation and annihilation operators satisfy the standard commutation relations. We are hence forced to interpret the creation and annihilation operators as creating and annihilating particles, respectively, to avoid negative-norm states.

- We say the $f_{\mathbf{k}}$ modes are positive frequency because they are proportional to $e^{-i\omega t}$, while the $f_{\mathbf{k}}^*$ modes have negative frequency; they span the space of solutions and define particles and antiparticles respectively. If we boost into another inertial frame, the frequencies will change by the Doppler shift, but the signs will remain the same. Thus all inertial observers will agree on the number of particles, and thus on the vacuum state.

- Formally, we were able to define positive and negative-frequency solutions in flat spacetime because of the existence of the timelike Killing vector ∂_t. In general, this would be defined by the Lie derivative, so that positive frequency modes obey

$$\mathcal{L}_K f_\omega = -i\omega f_\omega, \quad \omega > 0$$

  for Killing vector K.

- The fact that all inertial observers agree on the positive/negative frequency decomposition is because all other timelike Killing vectors are related to ∂_t by Lorentz transformations. Alternatively, it's because the notion of a particle in quantum field theory in flat spacetime is defined to be Poincare invariant.

10.2 Curved Spacetime

Next, we turn to a simple example of a curved spacetime.

- We consider the spatially flat Friedmann universe,

$$ds^2 = dt^2 - a(t)^2 \delta_{ik} dx^i dx^k.$$

  It is convenient to introduce the conformal time,

$$\eta(t) = \int^t \frac{dt}{a(t)}, \quad ds^2 = a(\eta)^2 \eta_{\mu\nu} dx^\mu dx^\nu$$

  which makes it clear the metric is conformally flat, a useful simplification.

- The action for a minimally coupled real scalar field is

$$S = \frac{1}{2} \int \sqrt{-g}\, d^4x \left(g^{\alpha\beta} \partial_\alpha \phi \partial_\beta \phi - m^2 \phi^2 \right) = \frac{1}{2} \int d\mathbf{x} d\eta\, a^2 (\phi'^2 - (\nabla\phi)^2 - m^2 a^2 \phi^2)$$

  where the prime indicates a derivative with respect to η. The action is not time-translation invariant, so the field can absorb energy from the gravitational field, creating particles. Note that if the scalar field were massless, the whole system would be conformally equivalent to a standard scalar field in flat spacetime, so there would be no particle creation.

- Changing variables to $\chi = a\phi$ and integrating by parts,

$$S = \frac{1}{2} \int d\mathbf{x} d\eta \left(\chi'^2 + (\nabla\chi)^2 - \left(m^2 a^2 - \frac{a''}{a} \right) \chi^2 \right).$$

  Then the equation of motion is just that of a real scalar field in flat spacetime with a time-dependent mass,

$$\chi'' - \nabla^2 \chi + m_{\text{eff}}^2 \chi = 0, \quad m_{\text{eff}}^2(\eta) = m^2 a^2 - \frac{a''}{a}.$$

  This makes it relatively straightforward to quantize the field χ.

- We first perform the mode expansion for the classical field,

$$\chi(\mathbf{x}, \eta) = \int đ\mathbf{k}\, \chi_{\mathbf{k}}(\eta) e^{i\mathbf{k}\cdot\mathbf{x}}$$

  so that

$$\chi_{\mathbf{k}}'' + \omega_k^2(\eta) \chi_{\mathbf{k}} = 0, \quad \omega_k^2(\eta) = k^2 + m_{\text{eff}}^2(\eta).$$

- Letting v_k and v_k^* be the independent solutions to this equation,

$$\chi_{\mathbf{k}}(\eta) = \frac{1}{\sqrt{2}} \left(a_{\mathbf{k}}^- v_k^*(\eta) + a_{-\mathbf{k}}^\dagger v_k(\eta) \right)$$

  where the $a_{\mathbf{k}}^\pm$ are integration constants and $a_{\mathbf{k}}^+ = (a_{\mathbf{k}}^-)^*$ since the field is real. We can write v_k because the solution depends only on $k = |\mathbf{k}|$, by rotational symmetry.

- We normalize v_k so that $\mathrm{Im}(v_k' v_k^*) = 1$, and this normalization is time-independent since it fixes the Wronskian of v_k and v_k^*,

$$W(v_k, v_k^*) = v_k' v_k^* - v_k v_k^{*\prime} = 2i\,\mathrm{Im}(v' v^*)$$

  and the Wronskian is time-independent.

- Finally, the field takes the form

$$\chi(\mathbf{x}, \eta) = \frac{1}{\sqrt{2}} \int d\!\!{}^-\mathbf{k}\, (a_{\mathbf{k}}^- v_k^*(\eta) e^{i\mathbf{k}\cdot\mathbf{x}} + a_{\mathbf{k}}^+ v_k(\eta) e^{-i\mathbf{k}\cdot\mathbf{x}})$$

  where we changed variables $\mathbf{k} \to -\mathbf{k}$ in the second term.

Next, we turn to canonical quantization.

- For a general field ϕ, we would define the conjugate momentum

$$\pi = \frac{\partial \mathcal{L}}{\partial(\nabla_0 \phi)}$$

  and impose the canonical commutation relations

$$[\phi(t, \mathbf{x}), \pi(t, \mathbf{x})] = \frac{i}{\sqrt{-g}} \delta(\mathbf{x} - \mathbf{x}').$$

  In this case χ is effectively on a flat background, so $\pi = \chi'$ and $\sqrt{-g} = 1$.

- We find that the operators $a_{\mathbf{k}}^{\pm}$ obey the commutation relations

$$[a_{\mathbf{k}}^-, a_{\mathbf{k}'}^+] = \delta(\mathbf{k} - \mathbf{k}')$$

  with all other commutators zero, using the normalization $\mathrm{Im}(v_k' v_k^*) = 1$.

- Note that if we instead had a complex scalar field, the mode expansion would be

$$\chi(\mathbf{x}, \eta) = \frac{1}{\sqrt{2}} \int d\!\!{}^-\mathbf{k}\, (a_{\mathbf{k}}^- v_k^*(\eta) e^{i\mathbf{k}\cdot\mathbf{x}} + b_{\mathbf{k}}^+ v_k(\eta) e^{-i\mathbf{k}\cdot\mathbf{x}})$$

  giving two independent sets of creation and annihilation operators.

- The mode functions $v_k(\eta)$ are not unique. We may define new mode functions by a Bogoliubov transformation,

$$u_k(\eta) = \alpha_k v_k(\eta) + \beta_k v_k^*(\eta)$$

  where α_k and β_k are complex constants. By linearity, u_k also satisfies the appropriate differential equation, and the normalization conditions are satisfied if

$$|\alpha_k|^2 - |\beta_k|^2 = 1.$$

  This is a familiar condition: a Bogoliubov transformation is a lot like a Lorentz transformation, as both preserve an indefinite metric.

- In terms of the new mode functions u_k the field is

$$\chi(\mathbf{x},\eta) = \frac{1}{\sqrt{2}} \int d\!\!{}^{-}\mathbf{k}\, (b^-_{\mathbf{k}} u^*_k(\eta) e^{i\mathbf{k}\cdot\mathbf{x}} + b^+_{\mathbf{k}} u_k(\eta) e^{-i\mathbf{k}\cdot\mathbf{x}})$$

where we have defined a new set of creation and annihilation operators which also obey the standard commutation relations. Unlike the flat spacetime case, we do not get further constraints by demanding the vacuum is an eigenstate of the Hamiltonian, because the Hamiltonian is time-dependent. We will discuss the choice of physical vacuum further below.

- These two fields are equal, so the integrands are the same, so the creation and annihilation operators are related by the Bogoliubov transformation

$$a^-_{\mathbf{k}} = \alpha^*_k b^-_{\mathbf{k}} + \beta_k b^+_{-\mathbf{k}}, \quad a^+_{\mathbf{k}} = \alpha_k b^+_{\mathbf{k}} + \beta^*_k b^-_{-\mathbf{k}}.$$

The inverse of this relation is

$$b^-_{\mathbf{k}} = \alpha_k a^-_{\mathbf{k}} - \beta_k a^+_{-\mathbf{k}}, \quad b^+_{\mathbf{k}} = \alpha^*_k a^+_{\mathbf{k}} - \beta^*_k a^-_{-\mathbf{k}}.$$

Note that this is a special case of the most general possible transformation, which relates a b operator to all of the a operators. We have this restricted form here because of momentum conservation/translational invariance.

- We can construct a Fock space using either the $a^\pm_{\mathbf{k}}$ or the $b^\pm_{\mathbf{k}}$. We define the vacuum state $|0_a\rangle$ and $|0_b\rangle$ to be annihilated by the $a^-_{\mathbf{k}}$ or $b^-_{\mathbf{k}}$ respectively.

- As an example, the b vacuum contains a particles, as

$$\langle 0_b | N^a_{\mathbf{k}} | 0_b \rangle = \langle 0_b | a^+_{\mathbf{k}} a^-_{\mathbf{k}} | 0_b \rangle = \langle 0_b | \beta^*_k b^-_{-\mathbf{k}} \beta_k b^+_{-\mathbf{k}} | 0_b \rangle = |\beta_k|^2.$$

The total number density of a particles, and the energy density relative to the a vacuum, are

$$n = \int d\!\!{}^{-}\mathbf{k}\, |\beta_k|^2, \quad \rho = \int d\!\!{}^{-}\mathbf{k}\, \omega_{\mathbf{k}} |\beta_{\mathbf{k}}|^2$$

and the energy density is finite only if $|\beta_k|$ decays faster than $1/k^2$.

- The b vacuum can be expressed in terms of a superposition of a-particle states as

$$|0_b\rangle = \prod_{\mathbf{k}} \left[\frac{1}{|\alpha_k|^{1/2}} \exp\left(\frac{\beta_k}{2\alpha_k} a^+_{\mathbf{k}} a^+_{-\mathbf{k}} \right) \right] |0_a\rangle.$$

This is straightforward to derive by focusing on one pair of $\mathbf{k}$ and $-\mathbf{k}$ modes, and the factor of two comes from summing over all $\mathbf{k}$ instead of over all distinct pairs. If we had instead used a complex scalar field, we would find that the b vacuum contains particle-antiparticle a pairs. The normalization factor $\prod |\alpha_k|^{1/2}$ only converges if $|\beta_k|^2$ decays faster than k^{-3}. Also note that formally, the $\mathbf{k} = \mathbf{0}$ mode has the form of a squeezed state.

- Note that by the definition of the Bogoliubov transformation, $|\beta_k/\alpha_k| < 1$. Indeed, in the limit $|\beta_k/\alpha_k| \to 1$, we have $|\beta_k| \to \infty$, so the energy density diverges, and the state above is not even normalizable for $|\beta_k/\alpha_k| \geq 1$. For $|\beta_{\mathbf{0}}/\alpha_{\mathbf{0}}| = 1$, the $\mathbf{k} = \mathbf{0}$ mode is squeezed all the way to a momentum eigenstate.

Note. More generally, we require our spacetime to be globally hyperbolic, so that a mode function is defined by initial data on a timeslice. The inner product must also be suitably generalized, replacing derivatives with covariant derivatives.

Next, we discuss the choice of physical vacuum.

- Conceptually, the indeterminate choice of vacuum comes from the absence of a timelike Killing vector in the original curved spacetime, which means we have no reference for positive frequency. Detectors will measure particles using a positive frequency reference that matches with the flow of their proper time, and hence may disagree on the number of particles.

- In the case of flat spacetime treated earlier, the vacuum state was the one with minimum possible energy. But in this case, the Hamiltonian is time-dependent and hence does not have time-independent eigenvectors. We can perform the same procedure for a Hamiltonian at a particular time η_0, giving the instantaneous lowest energy state $|0_{\eta_0}\rangle$, but this state may not have a useful physical meaning.

- For an arbitrary set of mode functions $v_k(\eta)$ we have
$$H(\eta) = \frac{1}{4}\int d\!\!\!^{-}\mathbf{k}\, a^-_{\mathbf{k}} a^-_{-\mathbf{k}} F^*_k + a^+_{\mathbf{k}} a^+_{-\mathbf{k}} F_k + (2a^+_{\mathbf{k}} a^-_{\mathbf{k}} + \delta(\mathbf{0}))E_k$$
where
$$E_k(\eta) = |v'_k|^2 + \omega_k^2(\eta)|v_k|^2, \quad F_k(\eta) = v'^2_k + \omega^2_k(\eta) v_k^2.$$
Therefore, the energy density of the associated vacuum $|0_v\rangle$ at time η_0 is
$$\epsilon(\eta_0) = \frac{1}{4}\int d\!\!\!^{-}\mathbf{k}\, E_k(\eta_0).$$
The lowest energy state is then found by minimizing $E_k(\eta_0)$ for each k individually. Note that we are minimizing the vacuum zero-point energy, which we instantly threw away in flat spacetime, because it is not a constant in this context; it depends on the mode functions.

- Dropping the k subscript, and setting $v = re^{i\alpha}$ for real r and α, the normalization condition is $r^2\alpha' = 1$, and
$$E(\eta_0) = |v'|^2 + \omega^2 |v|^2 = r'^2 + r^2\alpha'^2 + \omega^2 r^2 = r'^2 + \frac{1}{r^2} + \omega^2 r^2.$$
This is minimized when $r'(\eta_0) = 0$ and $r(\eta_0) = \omega^{-1/2}$, giving
$$v_k(\eta_0) = \frac{1}{\sqrt{\omega_k(\eta_0)}}, \quad v'_k(\eta_0) = i\omega_k v_k(\eta_0)$$
where we set the arbitrary phase $\alpha_k(\eta_0)$ to zero.

- At this moment, we have
$$E_k(\eta_0) = 2\omega_k(\eta_0), \quad F_k(\eta_0) = 0, \quad H(\eta_0) = \int d\!\!\!^{-}\mathbf{k}\, \omega_k(\eta_0)\left(a^+_{\mathbf{k}} a^-_{\mathbf{k}} + \frac{1}{2}\delta(\mathbf{0})\right)$$
so the Fock space constructed from the creation operators from these mode functions instantaneously diagonalize the Hamiltonian.

- However, there could be and often are modes with negative $\omega_k^2(\eta_0)$. Then the instantaneous Hamiltonian is not bounded below, and the idea of defining the instantaneous number of particles through excitations of the instantaneous lowest energy state breaks down completely.

- Abandoning the notion of defining a number of particles at every moment in time, we can consider the useful special case where $\omega_k(\eta)$ tends to a constant for low and high η. In this case, we can unambiguously define 'in' and 'out' vacua and determine how many 'out' particles were produced during the process, given that we start in the 'in' vacuum. However, the number of particles at any point during the process is not well-defined.

- Conceptually, the number of particles is well-defined for modes with $k \gg a^{-1}$, where a is the curvature scale. Such modes do not "feel" the curvature and are approximately plane waves.

Note. A basic summary of the above is that a set of mode functions, obeying the appropriate normalization, corresponds to a set of creation and annihilation operators obeying the standard commutation relations. These operators then define a Heisenberg vacuum state (which is not necessarily the same as the lowest energy state at any time, which is a notion that may not even exist), along with a ladder of states built upon it.

The ambiguity of mode functions can appear even in flat spacetime, if we allow for boundary conditions, as discussed in the notes on Optics, and physically means that different detectors may disagree on photon statistics, even if they share an inertial frame.

Note. In situations where ω_k changes slowly, the mode equation of motion

$$v_k'' + \omega_k^2(\eta) v_k = 0.$$

Intuitively, this has the form of a time-independent Schrodinger equation, where η plays the role of the "position", and $\omega_k^2(\eta)$ is the "potential". Just like the Schrodinger equation, we can solve this by the WKB approximation,

$$v_k(\eta) \approx \frac{1}{\sqrt{\omega_k(\eta)}} \exp\left(i \int^\eta \omega_k(\eta)\, d\eta \right).$$

We may use these approximate modes to define an instantaneous vacuum state, called the adiabatic vacuum. This can be a useful tool because it allows us to separate the effects of adiabatic changes in $\omega_k(\eta)$, with particle production effects associated with error terms in the adiabatic theorem. Higher-order WKB approximations yield higher-order adiabatic vacua.

Example. Consider the simple case

$$m_{\text{eff}}^2(\eta) = \begin{cases} m_0^2 & \eta < 0, \eta > \eta_1, \\ -m_0^2 & 0 < \eta < \eta_1 \end{cases}$$

which allows us to define 'in' and 'out' vacua. We let the mode functions be

$$v_k^{\text{in}}(\eta) = \frac{e^{i\omega_k \eta}}{\sqrt{\omega_k}}, \quad v_k^{\text{out}}(\eta) = \frac{e^{i(\eta-\eta_1)\omega_k}}{\sqrt{\omega_k}}, \quad \omega_k = \sqrt{k^2 + m_0^2}$$

where the expressions above are valid in the 'in' and 'out' regions, respectively, and may be used to define creation and annihilation operators, and hence a vacuum, in each region. We may straightforwardly solve the equation of motion for the mode function $v_k^{\text{in}}(\eta)$, by taking exponentials in each region and matching by demanding continuity of $v_k^{\text{in}}(\eta)$ and its derivative. This gives

$$v_k^{\text{in}}(\eta) = \frac{1}{\sqrt{\omega_k}} \left(\alpha_k^* e^{i\omega_k(\eta-\eta_1)} + \beta_k^* e^{-i\omega_k(\eta-\eta_1)} \right)$$

in the 'out' region, where the Bogoliubov coefficients are

$$\alpha_k = \frac{e^{-i\Omega_k\eta_1}}{4} \left(\sqrt{\frac{\omega_k}{\Omega_k}} + \sqrt{\frac{\Omega_k}{\omega_k}} \right)^2 - \frac{e^{i\Omega_k\eta_1}}{4} \left(\sqrt{\frac{\omega_k}{\Omega_k}} - \sqrt{\frac{\Omega_k}{\omega_k}} \right)^2, \quad \beta_k = \frac{1}{2} \left(\frac{\Omega_k}{\omega_k} - \frac{\omega_k}{\Omega_k} \right) \sin(\Omega_k\eta_1)$$

where $\Omega_k = \sqrt{k^2 - m_0^2}$. The final density of 'out' particles starting from the 'in' vacuum is

$$n_k = |\beta_k|^2 = \frac{m_0^4}{|k^4 - m_0^4|} \left| \sin\left(\eta_1 \sqrt{k^2 - m_0^2} \right) \right|^2.$$

The above expressions hold for all k, where for $k < m_0$ the argument of the sine is imaginary. For $k \gg m_0$, we clearly have $n_k \sim k^{-4} \ll 1$, but for $k \ll m_0$ we have $\sqrt{k^2 - m_0^2} \approx i m_0$, and hence

$$n_k \sim \sinh^2(m_0\eta_1).$$

Assuming $m_0\eta_1 \gg 1$, this is exponentially large. The particle energy density is

$$\rho = \int d\!\!\!{}^-\mathbf{k}\, n_k \omega_k$$

which is logarithmically divergent for high k. This is because of the unphysical, discontinuous change of $m_{\text{eff}}^2(\eta)$, and can be removed by applying an ultraviolet cutoff. Then as long as the cutoff is not extremely high, the dominant contribution to the integral comes from low frequencies $k \lesssim m_0$, so a rough estimate is

$$\rho \sim m_0 \int_0^{m_0} dk\, k^2 \exp(2m_0\eta_1) \sim m_0^4 \exp(2m_0\eta_1).$$

Example. Consider a linearly changing mass. Like the previous example, this could happen in a completely flat spacetime where the background dynamics have nontrivial time-dependence, so we write the dependent variable as t, and let $m(t) = m_0 t$. The mode functions obey

$$(-\partial_t^2 - m_0^2 t^2) u_k = k^2 u_k.$$

This is analogous to the Schrodinger problem of tunneling through an upside-down quadratic barrier, where the tunneling amplitude gives β_k, and there is an exact solution in terms of parabolic cylinder functions. For demonstration purposes, we will use the WKB approximation. **(finish)**

10.3 The Unruh Effect

So far we haven't discussed what set of particles a particle detector detects. Roughly speaking, 'positive frequency' is defined with respect to the detector's proper time. In particular, a uniformly accelerated observer can detect particles even in the Minkowski vacuum, in the Unruh effect.

- We work in two-dimensional Minkowski space, $ds^2 = dt^2 - dx^2$. A uniformly accelerated observer has $a^\mu a_\mu = -a^2$ for a constant a. For simplicity, we use lightcone coordinates,

$$u = t - x, \quad v = t + x, \quad ds^2 = du dv.$$

  In particular, Lorentz transformations take the form $u \to \alpha u$, $v \to v/\alpha$.

- Applying $u^\mu u_\mu = 1$ and $a^\mu a_\mu = -a^2$ in these coordinates,

$$\dot{u}\dot{v} = 1, \quad \ddot{u}\ddot{v} = -a^2$$

  where the dots are derivatives with respect to proper time τ. Straightforwardly solving the equations, performing a Lorentz transformation, and shifting the origin we have

$$u(\tau) = -\frac{1}{a} e^{-a\tau}, \quad v(\tau) = \frac{1}{a} e^{a\tau}$$

  which in the original coordinates gives

$$x(\tau) = \frac{1}{a} \cosh a\tau, \quad t(\tau) = \frac{1}{a} \sinh a\tau.$$

  The worldline is a hyperbola, and the observer is at rest at time $t = 0$ at $x = a^{-1}$.

- Next, we switch to a coordinate system adapted to the accelerating observer. We define

$$u = -\frac{1}{a} e^{-a\tilde{u}}, \quad v = \frac{1}{a} e^{a\tilde{v}}, \quad ds^2 = e^{a(\tilde{v}-\tilde{u})} d\tilde{u} d\tilde{v}$$

  so the trajectory is simply $\tilde{u}(\tau) = \tilde{v}(\tau) = \tau$. Switching back to spacelike and timelike coordinates, we have

$$\tilde{u} = \xi^0 - \xi^1, \quad \tilde{v} = \xi^0 + \xi^1, \quad ds^2 = e^{2a\xi^1}((d\xi^0)^2 - (d\xi^1)^2).$$

  All metrics in two dimensions are conformally flat, but this choice is particularly nice because it is manifestly conformally flat.

- Both coordinates ξ^0, ξ^1 cover the range $(-\infty, \infty)$, but only cover the wedge $x > |t|$. Hence we say the metric instead yields Rindler space, which is a subset of Minkowski space.

- Physically, inertial observers in Minkowski space follow orbits of the time translation Killing vector field, or combinations of it with space translation. By contrast, Rindler observers follow the timelike Killing vector field associated with boosts.

Note. There are a variety of possible coordinates for Rindler space. As just one further example, we can define coordinates (ρ, τ) by

$$t = \left(\rho + \frac{1}{a}\right) \sinh(a\tau), \quad x = \left(\rho + \frac{1}{a}\right) \cosh(a\tau) - \frac{1}{a}.$$

These are defined so that ρ is constant for each uniformly accelerated path, and that for fixed τ, the spacetime interval between points at ρ and $\rho + d\rho$ is just $d\rho$. These are Kottler-Moller coordinates, and have metric

$$ds^2 = (1 + a\rho)^2 d\tau^2 - d\rho^2$$

which is precisely that of a uniform gravitational field.

Next, we perform canonical quantization.

- In $1+1$ dimensions, the action for a massless scalar field is conformally invariant,

$$S[\phi] = \frac{1}{2}\int d^2x \sqrt{-g}\, g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi$$

  because the transformation of $\sqrt{-g}$ exactly cancels that of $g^{\mu\nu}$. Thus the action looks very similar in the (u,v) coordinates and the $(\tilde{u},\tilde{v})$ coordinates,

$$S = 2\int \partial_u\phi\partial_v\phi\, du dv = 2\int \partial_{\tilde{u}}\phi\partial_{\tilde{v}}\phi\, d\tilde{u}d\tilde{v}.$$

- The field equations take the form

$$\partial_u\partial_v\phi = 0, \quad \partial_{\tilde{u}}\partial_{\tilde{v}}\phi = 0$$

  which have the simple solutions

$$\phi(u,v) = A(u) + B(v), \quad \phi(\tilde{u},\tilde{v}) = \tilde{A}(\tilde{u}) + \tilde{B}(\tilde{v}).$$

  In particular, the positive frequency modes in the (u,v) coordinates are simply $e^{-i\omega u}$, giving a right-moving mode, and $e^{-i\omega v}$, giving a left-moving mode, with similar expressions for $(\tilde{u},\tilde{v})$.

- Therefore, we have the two mode expansions

$$\phi = \int_0^\infty \frac{d\omega}{\sqrt{2\pi}}\frac{1}{\sqrt{2\omega}}(e^{-i\omega u}a_\omega^- + e^{i\omega u}a_\omega^+) + \text{left} = \int_0^\infty \frac{d\Omega}{\sqrt{2\pi}}\frac{1}{\sqrt{2\Omega}}(e^{-i\Omega\tilde{u}}b_\Omega^- + e^{i\Omega\tilde{u}}b_\Omega^+) + \text{left}$$

  where we've left the left-moving modes implicit, and the latter is valid for $x > |t|$. These define the Minkowski vacuum $|0_M\rangle$ and the Rindler vacuum $|0_R\rangle$ respectively.

- We claim that the uniformly accelerated observer detects Rindler particles, and hence sees the Minkowski vacuum as containing particles. To find the particle spectrum, we compute the Bogoliubov transformation,

$$b_\Omega^- = \int_0^\infty d\omega\, a_{\Omega\omega}a_\omega^- - \beta_{\Omega\omega}a_\omega^+$$

  where maintaining the commutation relations requires

$$\int_0^\infty d\omega\, a_{\Omega\omega}a^*_{\Omega'\omega} - \beta_{\Omega\omega}\beta^*_{\Omega'\omega} = \delta(\Omega - \Omega').$$

- A rather complicated calculation then yields

$$|a_{\Omega\omega}|^2 = e^{2\pi\Omega/a}|\beta_{\Omega\omega}|^2.$$

  In particular, the number of particles with frequency Ω is

$$\langle N_\Omega\rangle = \int d\omega\, |\beta_{\omega\Omega}|^2 = \int d\omega\, \frac{|a_{\Omega\omega}|^2 - |b_{\Omega\omega}|^2}{e^{2\pi\Omega/a} - 1} = \frac{\delta(0)}{e^{2\pi\Omega/a} - 1}$$

  where we used the normalization condition. Since $\delta(0) = V$, we have computed a number density that obeys the Planck distribution for $T = a/2\pi$. More detailed calculations shows that the radiation observed by the accelerated observer is precisely thermal.

- Note that the number of particles falls off exponentially when $\Omega \gtrsim a$. Generally, this occurs when Ω is greater than the timescale of the variations.
- It can be shown that the Rindler vacuum is physically singular; it requires an infinite amount of energy to be prepared from the Minkowski vacuum.
- Note that the Minkowski vacuum has zero energy density, $\langle T^{\mu\nu} \rangle = 0$ according to both inertial and Rindler observers. The energy associated with the particles detected by the Rindler observer does not come from the vacuum, but rather from whatever gives the observer a constant acceleration.

Finally, we turn to the issue of what a particle detector measures.

- This problem is difficult because the notion of a particle is 'global', as it requires mode functions that are defined on at least a large patch of spacetime. This is in contrast to quantities such as $\langle T_{\mu\nu}(x) \rangle$ which can be defined locally. In practice, this means that the response of a detector will depend on its entire past history.
- We consider a detector with worldline $x^\mu(\tau)$ in the Minkowski vacuum, with interaction
$$L_{\text{int}} = cm(\tau)\phi(x(\tau))$$
where $m(\tau)$ is the detector's monopole moment operator, and c is a small parameter. The states of the detector are parametrized by their energy E.
- By first order perturbation theory, the amplitude for a transition is
$$ic\langle \psi, E| \int m(\tau)\phi(x(\tau))\, d\tau |0_M, E_0\rangle = ic\langle E|m(0)|E_0\rangle \int e^{i(E-E_0)\tau} \langle \psi|\phi(x)|0_M\rangle\, d\tau$$
where $|\psi\rangle$ is the final state of the field and we used $m(\tau) = e^{iH_0\tau} m(0) e^{-iH_0\tau}$. The first factor depends on the internal details of the detector.
- Since we're working to first order, $|\psi\rangle$ must be a one-particle state $|\mathbf{k}\rangle$, and
$$\langle \mathbf{k}|\phi(x)|0_M\rangle \propto e^{-i\mathbf{k}\cdot\mathbf{x}(\tau)+i\omega t(\tau)}$$
and the final integral can be computed explicitly. For example, for a detector moving a uniform velocity, the transition amplitude vanishes, as we'd expect.
- For a general path, it's useful to consider the total transition probability, which is
$$P \sim \int d\tau\, d\tau'\, e^{-iE(\tau-\tau')} G^+(x(\tau), x(\tau')), \quad G^+(x, x') = \langle 0_M|\phi(x)\phi(x')|0_M\rangle.$$
In particular, if we use a uniformly accelerated detector, the resulting transition rate is the same as what we would compute for an inertial observing using the thermal Green's function at temperature $T = a/2\pi$. On the other hand, the transition rate is exactly zero for all uniformly accelerated detectors if we use the Rindler vacuum.
- Now consider the same process of particle absorption in the inertial frame. In this case, the detector instead emits particles, since it is an accelerating charged object, and this causes a radiation reaction force. The energy from whatever accelerates the detector both goes into the emitted particles and into the internal energy of the detector.

- Next, we return to the 'in/out' situation considered for the expanding universe. To detect the particle creation operationally, we take an unexcited detector in the 'in' state. Before the transition begins, we turn off the detector's coupling to the field adiabatically, and turn it back on after the transition ends.

- The detector will then see the number density of particles we computed earlier. Explicitly,

$$G^+(x,x')_{\text{in}} = \langle 0,\text{in}|\phi(x)\phi(x')|0,\text{in}\rangle = \int u^{\text{in}}_{\mathbf{k}}(x)u^{\text{in}*}_{\mathbf{k}}(x')\,d^{n-1}k$$

  where x and x' are in the 'out' region. This integral is difficult to evaluate since $u^{\text{in}}_{\mathbf{k}}$ is complicated in this region, so we perform a Bogoliubov transformation to the 'out' modes, which are just plane waves. The only term that doesn't vanish upon the τ and τ' integration is the one proportional to $|\beta_{\mathbf{k}}|^2$, giving exactly the integral we found earlier.

10.4 Hawking Radiation and Black Hole Thermodynamics

Next, we turn to a simple explanation of the Hawking effect. We begin by discussing negative energy particles.

- First, we define the energy of a particle in general relativity. A reasonable definition is $E = -K^\mu p_\mu$ where K^μ is a timelike Killing vector; this is conserved along geodesics and it is indeed the energy that would be measured by an observer with four-velocity along K^μ. Generally K^μ is not unique, but neither is the definition of energy, which is observer-dependent.

- For a rotating black hole, there is a region called the ergosphere, outside of the event horizon, where K^μ becomes spacelike and the energy can be negative. Then an object can fall into the ergosphere and split into two pieces, one with negative energy.

- It is difficult to interpret what 'negative energy' means when inside the black hole, as no observer would measure E to be the energy. Indeed a local observer should only measure positive energy because locally spacetime looks flat.

- However, we can simply look at the final condition: it is possible for the negative energy particle to fall into the event horizon and for the positive energy particle to come out. Note that the negative energy particle can't escape the ergosphere, as we know that far away from the black hole there are only positive energy particles.

- Once the positive energy particle escapes, we can measure its energy and see that it has increased; hence energy has been extracted from the black hole. This is called the Penrose process, and more generally superradiance.

- Similarly, a rotating black hole will spontaneously emit radiation, as virtual particle pairs formed in the ergosphere can be real, with opposite energy and momentum.

- Hawking radiation comes from static black holes, where there is no ergosphere; however, the time translation Killing vector is still spacelike inside the horizon. Instead, Hawking radiation is mediated by quantum tunneling: a positive energy particle can tunnel out, and a negative energy particle can tunnel in.

Next, we repeat the calculations of the previous section for a simplified black hole.

- We will work with a two-dimensional Schwarzschild black hole, by simply dropping the angular coordinates for the Schwarzschild metric,

$$ds^2 = \left(1 - \frac{r_g}{r}\right) dt^2 - \frac{dr^2}{1 - r_g/r}, \quad r_g = 2M.$$

We introduce the tortoise coordinate

$$dr^* = \frac{dr}{1 - r_g/r}, \quad ds^2 = \left(1 - \frac{r_g}{r}\right)(dt^2 - dr^{*2}).$$

Then we switch to lightcone coordinates,

$$ds^2 = \left(1 - \frac{r_g}{r}\right) d\tilde{u}d\tilde{v}, \quad \tilde{u} = t - r^*, \quad \tilde{v} = t + r^*.$$

- Finally, we introduce the Kruskal–Szekeres coordinates

$$u = -2r_g e^{-\tilde{u}/2r_g}, \quad v = 2r_g e^{-\tilde{v}/2r_g}, \quad ds^2 = \frac{r_g}{r} \exp\left(1 - \frac{r}{r_g}\right) dudv.$$

Everything is now regular at $r = r_g$, so we may extend the domain of u and v beyond $u < 0$ and $v > 0$ to all reals. We may also switch to timelike and spacelike coordinates T and R.

- Next, we again consider a massless scalar field. By conformal invariance the solutions are

$$\phi(u, v) = A(u) + B(v), \quad \phi(\tilde{u}, \tilde{v}) = \tilde{A}(\tilde{u}) + \tilde{B}(\tilde{v})$$

just as for the Unruh effect. For example, $\phi \propto e^{-i\omega\tilde{u}} = e^{-i\omega(t-r^*)}$ describes a positive-frequency mode with respect to time t moving away from the black hole.

- Therefore, we define the Boulware vacuum $|0_B\rangle$ for the $\tilde{u}/\tilde{v}$ coordinates and the Kruskal vacuum $|0_K\rangle$ for the u/v coordinates. The Boulware vacuum defines positive frequency with respect to t, which is the proper time for a distant observer; hence an static observer at infinity sees no particles in this vacuum. However, the Boulware vacuum is singular on the black hole horizon, with a diverging energy density; this makes it physically unacceptable since we are assuming the black hole is weakly perturbed. It is analogous to the Rindler vacuum.

- The Kruskal vacuum defines positive frequency with respect to T, and it has finite energy density everywhere besides the singularities; hence it is a better candidate for the 'physical' vacuum. It roughly corresponds to the vacuum state for a freely falling observer and is analogous to the Minkowski vacuum.

- Therefore, to see how many particles are seen by a distant observer, we perform a Bogoliubov transformation from the Kruskal vacuum to the Boulware vacuum. Note that the coordinate transformation is identical to that from Minkowski to Rindler coordinates with the replacement $a \to \kappa = 1/2r_g$. Thus there is a thermal spectrum with temperature

$$T_H = \frac{\kappa}{2\pi} = \frac{1}{8\pi M}.$$

More generally, κ is the surface gravity of the black hole.

We now physically interpret this result.

- Above, we have considered an eternal black hole, and the thermal spectrum contains both inward-moving and outward-moving particles. It follows that for an eternal black hole to exist, it must be placed in a thermal bath of temperature T_H as measured by a distant observer.
- Since a black hole can absorb particles, it must also be able to emit them. Therefore, a non-eternal black hole placed in empty space should emit with temperature T_H. This is the Hawking effect.
- To derive this more rigorously, we would need to choose a vacuum for a non-eternal black hole formed by gravitational collapse. In the far past, we simply have an asymptotically flat spacetime rather than a white hole; thus we can choose the ingoing Boulware modes. In the far future, we have a black hole, so we choose the outgoing Kruskal modes. In this vacuum, a distant observer indeed sees no radiation at early times and outgoing radiation at late times.
- By dimensional analysis, the wavelength of photons produced by Hawking radiation is on the order of the black hole size. All other particles can be produced as well, by pair creation. For example, a proton can only be produced by a black hole about the same size or smaller.
- In four dimensions, we need to account for the angular coordinates. The wave equation is now $\Box\,\phi = 0$, and expanding in partial waves by

$$\phi(t, r, \theta, \phi) = \sum_{\ell m} \phi_{\ell m}(t, r) Y_{\ell m}(\theta, \phi)$$

  the equation for $\phi_{\ell m}$ is

$$\left(\Box^{(2)} + \left(1 - \frac{r_g}{r}\right)\left(\frac{r_g}{r^3} + \frac{\ell(\ell+1)}{r^2}\right)\right)\phi_{\ell m}(t, r) = 0$$

  where $\Box^{(2)}$ is the two-dimensional Laplacian.
- Therefore, we have an additional potential barrier, even for $\ell = 0$, and only a fraction of outgoing or ingoing particles can make it through. This modifies the spectrum by a greybody factor, which also depends on the particle mass and spin, but keeps the temperature the same.
- More generally, we also have a chemical potential for each species of particle; for instance, a charged black hole will preferentially radiate particles of the same charge, quickly becoming neutral.
- By the Stefan–Boltzmann law, the luminosity is proportional to $T^4 \propto 1/M^4$ and to the area $A = 4\pi r_g^2 \propto M^2$. Then

$$\frac{dM}{dt} \propto \frac{1}{M^2}$$

  and the black hole evaporates in a finite time $t \sim M_0^3$, with a distinctive signature at the end of its lifetime. Ordinary black holes would last far too long to be observed this way, but primordial black holes might be detected. An alternative is that something else happens when $M \sim M_{\text{pl}}$ since nonperturbative quantum gravity effects would become dominant.

Finally, we briefly venture into black hole thermodynamics.

- Taking the differential of the expression $A = 16\pi M^2$,
$$dM = \frac{1}{8\pi M} d\left(\frac{A}{4}\right)$$
which is reminiscent of the first law, $dE = TdS$, as long as we identify
$$S_{\mathrm{BH}} = \frac{1}{4}A = 4\pi M^2.$$
That is, a black hole has an extremely high entropy, reflecting the fact that it could have been formed in many ways; the black hole at the center of our galaxy has more entropy than all visible matter. In classical GR, this is puzzling as a black hole is characterized by only a few parameters, but string theory has reproduced S_{BH} by microscopic state counting.

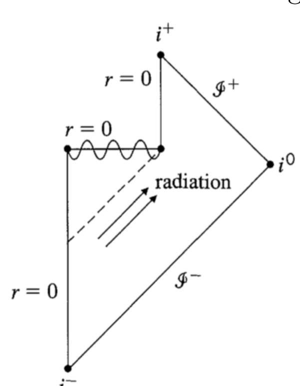
- The second law of black hole thermodynamics states that
$$\delta S = \delta S_{\mathrm{matter}} + \delta S_{\mathrm{BH}} \geq 0.$$
The evidence for this statement is that it holds without black holes by classical thermodynamics, and it holds in classical GR (by the area theorem), and it continues to hold when we add Hawking radiation.

- Since $E(T) = M = 1/8\pi T$, the heat capacity of a black hole is
$$C_{\mathrm{BH}} = -\frac{1}{8\pi T^2}$$
which is negative; a black hole cannot be in stable thermal equilibrium with an infinite reservoir.

- However, a black hole can be in equilibrium with a finite reservoir, as
$$\frac{d^2S}{dQ^2} = -\frac{1}{T^2}\frac{dT}{dQ} = -\frac{1}{CT^2}$$
which indicates that the entropy is maximized in equilibrium,
$$\frac{d^2(S_{\mathrm{res}} + S_{\mathrm{BH}})}{dQ^2} > 0$$
when $C_{\mathrm{res}} < -C_{\mathrm{BH}}$.

- The black hole information loss paradox refers to the fact that it's hard to see how the information in the black hole reflected by S_{BH} can come out; it is likely encoded in subtle correlations, i.e. the spectrum is not perfectly thermal.

- The problem can also be seen in terms of a conformal diagram.



Here, a Cauchy surface after the singularity disappears cannot be used to infer the data on a Cauchy surface before the singularity disappears.

- One obstacle is that since $S_{\mathrm{BH}} \sim A$, we expect the information is encoded on the horizon. But when an object passes the event horizon of a large black hole, nothing special happens; we expect it continues to the singularity unimpeded. Assuming locality, this means the information must come out at late times, but at late times S_{BH} is too small!
- One way to avoid this problem is to give up on locality; this is related to the idea of holography, which is motivated by $S_{\mathrm{BH}} \sim A$.
- In a full theory of quantum gravity, the entropy $S = A/4$ is just the first term is a series of higher curvature corrections; the general formula is called the Wald entropy.

10.5 Spinors in Curved Spacetime

We begin with introducing two-component spinor notation in flat spacetime.

- We may convert between a vector index and two Weyl spinor indices using the Van de Waerden symbols,
$$dx^{AA'} = \sigma_a^{AA'} dx^a = \frac{1}{\sqrt{2}} \begin{pmatrix} dt + dz & dx + idy \\ dx - idy & dt - dz \end{pmatrix}.$$
We will call the two types of spinor indices primed and unprimed.
- We raise and lower the primed and unprimed spinor indices using
$$\epsilon_{AB} = \epsilon_{A'B'} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$
with the sign conventions
$$\psi^A \epsilon_{AB} = \psi_B, \quad \psi^A = \epsilon^{AB} \psi_B$$
and similarly for the primed version.
- Let $\mathbb{S}$ denote the two-dimensional complex vector space of unprimed spinors ψ^A and let $\mathbb{S}'$ be the same for primed spinors $\psi^{A'}$. Then the Van de Waerden symbols give a concrete isomorphism
$$T = \mathbb{S} \otimes \mathbb{S}'$$
where T is the tangent space. Hence in terms of abstract indices only, we can identify $V^a = V^{AA'}$ as both represent the same geometric object.
- Given a Lorentz transformation $L_b^a \in SO_+(1,3)$, there is a corresponding transformation $L_B^A \in SL(2, \mathbb{C})$ so that
$$L_a^b \sigma_b^{AA'} = L_B^A \overline{L}_{B'}^{A'} \sigma_a^{BB'}$$
and hence vectors can be transformed with either the L_b^a or the L_B^A.
- We can also give a correspondence between infinitesimal Lorentz transformations. For an antisymmetric l_{ab} and symmetric l^{AB},
$$l^{ab} \sigma_a^{AA'} \sigma_b^{BB'} = l^{AA'BB'} = \epsilon^{A'B'} l^{AB} + \epsilon^{AB} \overline{l}^{A'B'}, \quad l^{AB} = \frac{1}{2} l^{AA'BB'} \epsilon_{A'B'}$$
which expresses the isomorphism between $\mathfrak{so}(1,3)$ and $\mathfrak{sl}(2, \mathbb{C}) \oplus \overline{\ell(2, \mathbb{C})}$.

- A Dirac spinor is constructed as $\psi^\alpha = (\psi^A, \phi^{A'})$, and the Clifford matrices are

$$\gamma^\alpha_{c\beta} = \sqrt{2}\begin{pmatrix} 0 & {\sigma_{cB'}}^A \\ {\sigma_{cB}}^{A'} & 0 \end{pmatrix}, \quad \{\gamma_a, \gamma_b\} = -2I\eta_{ab}.$$

  The Dirac equation is $\gamma^a \partial_a \psi = m\psi$, which in two-component notation is

$$\partial_{AA'}\psi^A = m\phi_{A'}, \quad \partial_{AA'}\phi^{A'} = m\psi_A.$$

  Here we have implicitly converted spacetime indices to spinor indices, $\partial_{AA'} = \sigma^a_{AA'}\partial_a$, and we will do this implicitly below.

Note. We can construct spin s fields by taking the symmetric tensor product of spinors, $\phi_{(A_1 \ldots A_{2s})}$. Note that these fields are automatically traceless, because contraction is with the antisymmetric ϵ^{AB}. Now, massless field equations for spin s fields take the form

$$\nabla^{A_1}_{A'}\phi_{A_1 \ldots A_{2s}} = 0.$$

Here s is the spin/helicity of the corresponding particle, and $s < 0$ can be reached by complex conjugation. For $s = 0$, this is just the scalar wave equation, and for $s = 1/2$ we get the Weyl equation. The case $s = 1$ is a bit confusing. **(finish)**