B5: General Relativity and Cosmology

Toby Adkins

July 14, 2017

Contents

1. *An Introduction to Curved Spacetime*

This section aims to introduce the concepts underpinning General Relativity, including:

- Newtonian Gravity
- The Metric and Geodesics
- Curvature
- Einstein's Field Equations

General Relativity really is one of the crowning jewels of early 20th Century physics; one does not really appreciate Einstein's genius until they have examined the rich complexity involved with this theory. This chapter shall serve as a basic introduction to the theory, before we go on to examine some of its interesting consequences.

It is recommended that readers are very familiar with index notation; this author shall not clarify steps of working unless something out of the ordinary has been done. There is a lot of index re-labelling, which proves a useful exercise for the reader to follow. We will also no longer refer to four-vectors; as we never really deal with three-vectors in General Relativity, we shall use the term 'vector' to refer to a four-vector.

1.1 Newtonian Gravity

Before diving in immediately with General Relativity, let take a second to examine our old friend of Newtonian gravity, something which should be familiar to most readers of this text.

Newton's law of universal gravitation states that the force between two bodies of m_1 and m_2 located at $\mathbf{x}_1$ and $\mathbf{x}_2$ is given by

$$\mathbf{F} = Gm_1m_2\frac{\mathbf{x}_2 - \mathbf{x}_1}{|\mathbf{x}_2 - \mathbf{x}_1|^3} \tag{1.1}$$

where $G = 6.672 \times 10^{-11}\ \mathrm{m^3kg^{-1}s^{-2}}$ is the universal gravitational constant. The resulting gravitational acceleration felt by one of the masses (say, m_1) is given by

$$\mathbf{g} = Gm_2\frac{\mathbf{x}_2 - \mathbf{x}_1}{|\mathbf{x}_2 - \mathbf{x}_1|^3} \tag{1.2}$$

Now, suppose that this acceleration (force per unit mass) can be written as the gradient of some scalar potential $\mathbf{g} = -\nabla\Phi$. The potential associated with a mass distribution is the superposition of the potentials resulting from point masses, given by

$$\Phi(\mathbf{x}) = -\sum_i \frac{Gm_i}{|\mathbf{x} - \mathbf{x}_i|} \tag{1.3}$$

We can then generalise this to a continuous distribution by integrating over some closed volume V containing the masses:

$$\Phi(\mathbf{x}) = -G\int_V d^3m(\mathbf{x}')\,\frac{1}{|\mathbf{x} - \mathbf{x}'|} = -G\int_V d^3\mathbf{x}'\,\frac{\rho(\mathbf{x}')}{|\mathbf{x} - \mathbf{x}'|} \tag{1.4}$$

for some mass density distribution $\rho(\mathbf{x}')$. Now, consider the Laplacian of both sides of this equation:

$$\boldsymbol{\nabla}^2\Phi(\mathbf{x}) = -G\int_V d^3\mathbf{x}'\,\rho(\mathbf{x}')\nabla^2\left(\frac{1}{|\mathbf{x} - \mathbf{x}'|}\right) = G\int_V d^3\mathbf{x}'\,\rho(\mathbf{x}')\left(4\pi\delta^3(\mathbf{x} - \mathbf{x}')\right) \tag{1.5}$$

This means that we can write

$$\boxed{\nabla^2\Phi = 4\pi G\rho} \tag{1.6}$$

This equation encompasses all of Newtonian gravity. However, it starts to break down if we bring the considerations of relativity to bear. For example, Special Relativity tells us that there is an equivalence between energy and mass. In the above equation, ρ serves as the source term for Φ, but the energy that is associated with Φ has its own mass, which further effects Φ. This means that the study of gravity - unlike electromagnetism, for example - is intrinsically non-linear, which shall be reflected in our study of General Relativity.

1.1.1 The Equivalence Principle

In the above derivation, it was implicitly assumed that the mass in (1.1) and that in Newton's second law were the same. That is, it was assumed that inertial mass and gravitational mass were the same. Is this assumption valid?

Consider two masses made of different materials attached to either end of a straight rod suspended from a string on the surface of the Earth. Each of the masses will be subjected to two forces: the gravitational pull towards the centre of the Earth, and the forces due to the rotation of the Earth. This causes the rod to hang at an angle relative to the vertical direction. The rod will be free to rotate if there is a difference in gravitational acceleration between the masses, which can only occur if there is a difference between the gravitational and inertial masses.

Consider two masses that have gravitational masses m_{G1} and m_{G2}, with inertial masses m_{I1} and m_{I2} respectively. Let g be the component of the gravitational acceleration that causes the rod to twist, and a_1 and a_2 be the associated acceleration of each of the masses, such that

$$m_{I1}a_1 = m_{G1}g, \quad m_{I2}a_2 = m_{G2}g \tag{1.7}$$

If the ratio of the inertial to gravitational mass are the same for both bodies, they will experience the same acceleration. Consider the dimensionless quantity:

$$\varsigma = \frac{a_1 - a_2}{a_1 + a_2} = \frac{(m_{G1}/m_{I1} - m_{G2}/m_{I2})}{(m_{G1}/m_{I1} + m_{G2}/m_{I2})} \approx (0.3 \pm 1.8) \times 10^{-13} \tag{1.8}$$

The numerical values follow from experiments using beryllium and titanium. The fact that the gravitational and inertial masses are the same (to within 10^{-13}) gives rise to the equivalence principle (EP), which can be stated in three ways:

- Weak Equivalence Principle (WEP) - All uncharged, freely falling test particles follow the same trajectories, once an initial position and velocity have been prescribed
- Einstein Equivalence Principle (EEP) - The WEP is valid, and furthermore in all freely falling frames one recovers (locally, and up to tidal gravitational forces) the same laws of special relativistic physics, independent of position or velocity
- Strong Equivalence Principle (SEP) - The WEP is valid for massive gravitating objects, as well as test particles, and in all freely falling frames one recovers (locally, and up to tidal gravitational forces) the same special relativistic physics, independent of position or velocity

Essentially, the equivalence principle encapsulates the idea that gravitational fields are indistinguishable from acceleration; or rather, being in a gravitational field is equivalent to being in an accelerated frame, and vice-versa.

1.1.2 Gravitational Redshift

A consequence that can be derived directly from the equivalence principle is that of gravitational red-shift, though we will derive this later through more sophisticated methods. Consider an observer A at some radial distance h above the surface of the earth, and another observer B on the surface of the Earth. By the equivalence principle, we can regard both observers as moving with some acceleration g, and so their positions can be written as

$$z_A(t) = \frac{1}{2}gt^2 + h, \quad z_B(t) = \frac{1}{2}gt^2 \tag{1.9}$$

Now, consider a light pulse emitted at t_0 by A, and received at some time t_1 by B. A subsequent pulse is then emmited at a time Δt_A by A, and received at a time $t_1 + \Delta t_B$. We then have that

$$z_A(0) - z_B(t_1) = h - \frac{1}{2}gt_1^2 = ct_1 \tag{1.10}$$

$$\begin{aligned} z_A(\Delta t_A) - z_B(t_1 + \Delta t_B) &= h + \frac{1}{2}g\Delta t_A^2 - \frac{1}{2}g(t_1 + \Delta t_B)^2 \\ &\approx h - \frac{1}{2}gt_1^2 - gt_1\Delta t_B = c(t_1 + \Delta t_B - \Delta t_A) \end{aligned} \tag{1.11}$$

where we have neglected terms of order Δt^2. Combining the two equations, and recognising that $t_1 \approx h/c$, we have that

$$\Delta t_A = \left(1 + \frac{gh}{c^2}\right) \Delta t_B \tag{1.12}$$

In other words, there is local gravitational time dilation due to the difference in the gravitational potentials at the two points A and B, meaning that we can in general write that

$$\boxed{\text{Rate Received} = \left(1 - \frac{\Phi_2 - \Phi_1}{c^2}\right) \times \text{Rate Emitted}} \tag{1.13}$$

where Φ_1 and Φ_2 are the local Newtonian gravitational potentials at two points. This formula is know as the *gravitational redshift* of light. It mus be stressed that this result is only locally valid, as we are using Φ. To tackle this result more generally, we first need to build some more sophisticated machinery.

1.2 The Metric and Geodesics

In Special Relativity, we were working in flat spacetime, where the metric was $\eta_{\mu\nu} = \text{diag}(-1,1,1,1)$. In order to upgrade our theory, we want to be able to work in any set of coordinates. Let ξ denote our *local intertial coordinates*, such that we can write the proper interval as

$$-c^2 d\tau^2 = \eta_{\alpha\beta} d\xi^\alpha d\xi^\beta \tag{1.14}$$

With a trivial application of the chain rule, we can generalise this result to an arbitrary coordinate system:

$$\boxed{-c^2 d\tau^2 = g_{\mu\nu} dx^\mu dx^\nu, \quad g_{\mu\nu} = \eta_{\alpha\beta} \frac{\partial \xi^\alpha}{\partial x^\mu} \frac{\partial \xi^\beta}{\partial x^\nu}} \tag{1.15}$$

Our treatment of systems in General Relativity completely revolves around the metric; once the metric is found, the behaviour of the system can be predicted. The metric is simply the coordinate transformation that must be applied to flat spacetime in order to determine the system in question. In section 1.3, we shall see how the metric influences the curvature of spacetime, while in chapter 2 we will look at a few particular cases of the metric. As in Special Relativity, the metric still satisfies the following relationships

$$g_{\mu\nu} = g_{\nu\mu}, \quad g_{\mu\nu} g^{\nu\rho} = \delta_\mu{}^\rho, \quad \partial_\mu x_\nu = g_{\mu\nu} \tag{1.16}$$

1.2.1 The Geodesic Equation

Given the equivalence principle, we can consider the trajectories of particles moving in the curved spacetime associated with a gravitational field as being equivalent to particles moving in a 'straight line' in an accelerated frame. Particles travel along *geodesics*, the shortest path between two points in curved spacetime. In Minkowski space, this is simply a straight line.

Within our overall space, adopt local inertial coordinates ξ^α. Then, as there are no forces acting, we have that

$$\frac{d^2 \xi^\alpha}{d\tau^2} = 0 \tag{1.17}$$

Now, we want to transform this back to our general coordinates. Using the chain rule:

$$0 = \frac{d}{d\tau}\left(\frac{\partial \xi^\alpha}{\partial x^\mu}\frac{dx^\mu}{d\tau}\right) = \frac{\partial \xi^\alpha}{\partial x^\mu}\ddot{x}^\mu + \dot{x}^\mu \dot{x}^\nu \frac{\partial^2 \xi^\alpha}{\partial x^\mu \partial x^\nu} = \frac{\partial \xi^\alpha}{\partial x^\mu}\frac{\partial x^\rho}{\partial \xi^\alpha}\ddot{x}^\mu + \dot{x}^\mu \dot{x}^\nu \frac{\partial^2 \xi^\alpha}{\partial x^\mu \partial x^\nu}\frac{\partial x^\rho}{\partial \xi^\alpha} \tag{1.18}$$

where the dot refers to a derivative with respect to the proper time parameter. The last expression as been obtained by multiplying throughout by $\partial x^\rho / \partial \xi^\alpha$. Now, we recognise that the first term can be written as

$$\frac{\partial \xi^\alpha}{\partial x^\mu}\frac{\partial x^\rho}{\partial \xi^\alpha}\ddot{x}^\mu = \frac{\partial x^\rho}{\partial x^\mu}\ddot{x}^\mu = \delta^\rho{}_\mu \ddot{x}^\mu = \ddot{x}^\rho \tag{1.19}$$

We thus arrive at the *geodesic equation*:

$$\boxed{\frac{d^2 x^\rho}{d\tau^2} + \Gamma^\rho_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = 0, \quad \Gamma^\rho_{\mu\nu} = \frac{\partial x^\rho}{\partial \xi^\alpha}\frac{\partial^2 \xi^\alpha}{\partial x^\mu \partial x^\nu}} \tag{1.20}$$

This equation describes the trajectory taken by particles through curved spacetime. The symbol $\Gamma^\rho_{\mu\nu}$ is known as the *affine connection*, which (despite its apparent form) is not actually a three rank tensor. However, it is clear that the affine connection must contain information about the curvature of the spacetime in which we are working, and so must be in some way related to the metric $g_{\mu\nu}$.

The Affine Connection

Let us investigate the form of the affine connection $\Gamma^\rho_{\mu\nu}$. Motivated by the fact that it must be related to $g_{\mu\nu}$, take the derivative of the second expression in (1.15):

$$\frac{\partial g_{\mu\nu}}{\partial x^\lambda} = \eta_{\alpha\beta}\left[\frac{\partial^2 \xi^\alpha}{\partial x^\lambda \partial x^\nu}\frac{\partial \xi^\beta}{\partial x^\nu} + \frac{\partial \xi^\alpha}{\partial x^\mu}\frac{\partial^2 \xi^\beta}{\partial x^\lambda \partial x^\nu}\right] \tag{1.21}$$

Now, we can use the definition of the affine connection in (1.20) to write that

$$\frac{\partial g_{\mu\nu}}{\partial x^\lambda} = \eta_{\alpha\beta}\left[\frac{\partial \xi^\beta}{\partial x^\nu}\frac{\partial \xi^\alpha}{\partial x^\rho}\Gamma^\rho_{\mu\lambda} + \frac{\partial \xi^\alpha}{\partial x^\mu}\frac{\partial \xi^\beta}{\partial x^\rho}\Gamma^\rho_{\nu\lambda}\right] = g_{\nu\rho}\Gamma^\rho_{\mu\lambda} + g_{\mu\rho}\Gamma^\rho_{\lambda\nu} \tag{1.22}$$

We note that the inertial coordinates ξ have completely disappeared from our expression. Recalling the fact that the metric is symmetric in its lower indices, it is clear that the affine connections must be symmetric in their lower indices in order for this expression to be non-vanishing. Connections satisfying such a condition are said to be *torsion free*. Permuting indices, we can combine expressions of the form (1.22) such that only one connection remains:

$$\frac{\partial g_{\mu\nu}}{\partial x^\lambda} + \frac{\partial g_{\lambda\nu}}{\partial x^\mu} - \frac{\partial g_{\mu\lambda}}{\partial x^\nu} = 2g_{\rho\nu}\Gamma^\rho_{\lambda\nu} \tag{1.23}$$

Inverting this expression using (1.16), and re-labelling some indices, we have that

$$\boxed{\Gamma^\rho_{\mu\nu} = \frac{1}{2}g^{\rho\sigma}\left(\partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - \partial_\sigma g_{\mu\nu}\right)} \tag{1.24}$$

As expected, the connections depend on the metric through its derivatives. In local inertial coordinates, the first derivatives of $g_{\mu\nu}$ vanish, meaning that the $\Gamma^\rho{}_{\mu\nu} = 0$, allowing us to re-obtain flat spacetime.

It is easy to show that for a diagonal metric g_{ab}, the affine connections satisfy the following useful identities:

$$\Gamma^a_{ba} = \Gamma^a_{ab} = \frac{1}{2g_{aa}}\frac{\partial g_{aa}}{\partial x^b} \quad (a = b \text{ permitted, no sum}) \tag{1.25}$$

$$\Gamma^a_{bb} = -\frac{1}{2g_{aa}}\frac{\partial g_{bb}}{\partial x^a} \quad (a \neq b, \text{ no sum}) \tag{1.26}$$

$$\Gamma^a_{bc} = 0 \tag{1.27}$$

As a Variation

In principle, we should be able to re-obtain the geodesic equation as the minimisation of an action, as with Lagrangian mechanics. In this case, the action that we are interested in is the proper time elapsed between two points in spacetime:

$$\mathcal{S} = \int_A^B d\tau = \int_{\lambda_1}^{\lambda_2} d\lambda\,\frac{d\tau}{d\lambda} = \int_{\lambda_1}^{\lambda_2} d\lambda\,\mathcal{L}(x,\dot{x},\lambda), \quad \mathcal{L}(x,\dot{x},\lambda) = \frac{d\tau}{d\lambda} \tag{1.28}$$

Note that we have introduced a parameter λ that increases monotonically along the path, and whose start and end values are fixed at λ_1 and λ_2. This is to avoid the problem that integrals over $d\tau$ can take different values depending on the specific choice of coordinates (frame). We can then write the variation of our Lagrangian $\mathcal{L}$ with respect to our spacetime coordinates x and $\dot{x}$ as

$$\delta\mathcal{L} = \mathcal{L}(x+\delta x, \dot{x}+\delta x, \lambda) - \mathcal{L}(x,\dot{x},\lambda) = \frac{\partial \mathcal{L}}{\partial x^\rho}\delta x^\rho + \frac{\partial \mathcal{L}}{\partial \dot{x}^\rho}\delta\dot{x}^\rho \tag{1.29}$$

Then, the variation in our action is

$$\delta\mathcal{S} = \int_{\lambda_1}^{\lambda_2} d\lambda \left[\frac{\partial\mathcal{L}}{\partial x^\rho}\delta x^\rho + \frac{\partial\mathcal{L}}{\partial \dot{x}^\rho}\delta\dot{x}^\rho\right] = \int_{\lambda_1}^{\lambda_2} d\lambda\, \delta x^\rho \left[\frac{\partial\mathcal{L}}{\partial x^\rho} - \frac{d}{d\lambda}\left(\frac{\partial\mathcal{L}}{\partial\dot{x}^\rho}\right)\right] = 0 \tag{1.30}$$

where we have integrated the second term by parts, and assumed that $\mathcal{L}$ is well defined such that the total derivative vanishes at the limits of integration. According to *Hamiltons principle*, the path taken is that for which the action is stationary with respect to variations in our coordinates, meaning that they must satisfy the *Euler-Lagrange equations*:

$$\boxed{\frac{d}{d\lambda}\left(\frac{\partial\mathcal{L}}{\partial\dot{x}^\rho}\right) - \frac{\partial\mathcal{L}}{\partial x^\rho} = 0} \tag{1.31}$$

Note that we are at liberty to choose the parameter λ, as it is entirely arbitrary; the most convenient choice is $\lambda = \tau$. Now, by (1.28), this Lagrangian is given explicitly by

$$\mathcal{L} = \frac{d\tau}{d\lambda} = \left(-g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}\right)^{1/2} \tag{1.32}$$

Now, extremising $\mathcal{L}$ is equivalent to extremising $\mathcal{L}^2$, and so consider instead the Lagrangian

$$\widetilde{\mathcal{L}} = -g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = -g_{\mu\nu}\dot{x}^\mu\dot{x}^\nu \tag{1.33}$$

Substitute this into (1.31):

$$\frac{\partial\mathcal{L}}{\partial x^\rho} = \frac{\partial g_{\mu\nu}}{\partial x^\rho}\dot{x}^\mu\dot{x}^\nu, \quad \frac{\partial\mathcal{L}}{\partial\dot{x}^\rho} = 2g_{\mu\rho}\dot{x}^\mu \tag{1.34}$$

This means that the path satisfied the equation

$$\boxed{\frac{d}{d\lambda}\left(g_{\mu\rho}\frac{dx^\mu}{d\tau}\right) - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\rho}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = 0} \tag{1.35}$$

This turns out to be the covariant form of the geodesic equation, which proves incredibly useful for finding conserved quantities, as we shall see in section 2.4. We can demonstrate this explicitly:

$$\begin{aligned} 0 &= \frac{d}{d\tau}(g_{\mu\rho}\dot{x}^\mu) - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\rho}\dot{x}^\mu\dot{x}^\nu = \frac{\partial g_{\mu\rho}}{\partial x^\nu}\dot{x}^\nu\dot{x}^\mu + g_{\mu\rho}\ddot{x}^\mu - \frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\rho}\dot{x}^\mu\dot{x}^\nu \\ &= g_{\mu\rho}\ddot{x}^\mu + \frac{1}{2}\left(\partial_\nu g_{\mu\rho} + \partial_\mu g_{\nu\rho} - \partial_\rho g_{\mu\nu}\right)\dot{x}^\mu\dot{x}^\nu = g_{\mu\rho}\ddot{x}^\mu + \Gamma_{\rho\mu\nu}\dot{x}^\mu\dot{x}^\nu \end{aligned} \tag{1.36}$$

where we have used the fact that $d/d\tau = x^\nu\partial_\nu$ to obtain the second expression. This is clearly the geodesic equation except with lowered indices in the affine connection, meaning that (1.20) and (1.35) are entirely equivalent.

1.2.2 The Newtonian Limit

Thus far, we have just done a lot of algebra, and it is not plainly obvious that anything we are doing makes any sense. In order to demonstrate the fact that it indeed does, let us examine the Newtonian limit of the geodesic equation. In General Relativity, the Newtonian limit corresponds to

- Weak Gravitation - The metric is approximately Minkowski, meaning that we write

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu}, \quad |h_{\mu\nu}| \ll 1 \tag{1.37}$$

  When working within this limit, the aim is to find the small correction $h_{\mu\nu}$

- Slow Motion - Objects move slowly compared to the speed of light, such that $\mathbf{v} \ll c$, or alternatively that $cdt \gg dx^i$, were the index i refers to the spatial components of the metric (we shall adopt i and j for this purpose throughout the remainder of this text). This means that in the Newtonian limit, $\tau \approx t$
- Slowly Varying Gravitational Fields - This means that any time derivatives of the metric can be ignored

Applying this to the geodesic equation, the only component of the metric that we are interested in is g_{00}, such that the only affine connection that is non-vanishing is Γ^ρ_{00}:

$$\frac{d^2x^\rho}{d\tau^2} + \Gamma^\rho_{00}\left(\frac{dx^0}{d\tau}\right)^2 \approx 0 \tag{1.38}$$

Using the fact that the gravitational field is roughly stationary, and (1.37), we can re-write the connection as

$$\Gamma^\rho_{00} = -\frac{1}{2}g^{\rho\mu}\frac{\partial g_{00}}{\partial x^\mu} \approx -\frac{1}{2}\eta^{\rho\mu}\frac{\partial h_{00}}{\partial x^\mu} = -\frac{1}{2}\frac{\partial h_{00}}{\partial x^\rho} \tag{1.39}$$

where we have made use of the fact that only spatial components of $\eta^{\rho\mu}$ remain, which are equal to unity. Then, the geodesic equation becomes

$$\frac{d^2x^\rho}{d\tau^2} = \frac{1}{2}\left(\frac{dx^0}{d\tau}\right)^2\frac{\partial h_{00}}{\partial x^\rho} \quad \longrightarrow \quad \frac{d^2\mathbf{x}}{d\tau^2} = \frac{1}{2}\left(\frac{dx^0}{d\tau}\right)^2 \nabla h_{00} \tag{1.40}$$

Comparing this with the Newtonian result $d^2\mathbf{x}/dt^2 = -\nabla\Phi$, we see that in the Newtonian limit

$$\boxed{g_{00} = -\left(1 + \frac{2\Phi}{c^2}\right), \quad h_{00} = -\frac{2\Phi}{c^2}} \tag{1.41}$$

Note that for metrics containing $\Phi(r) = -GM/r$, this expression can be found by expanding them for small Φ.

1.2.3 Time Dilation

We can write any metric explicitly in the form

$$-c^2d\tau^2 = g_{00}\,c^2dt^2 + g_{ij}dx^idx^j \tag{1.42}$$

where again the indices i and j are used to refer to the spatial components of the metric. Then, it is clear that the proper time, and the local inertial time are related by

$$\boxed{\frac{dt}{d\tau} = \left(-g_{00} - \frac{1}{c^2}g_{ij}\frac{dx^i}{dt}\frac{dx^j}{dt}\right)^{-1/2}} \tag{1.43}$$

This is the analogue of time dilation in General Relativity. In particular, if the clock measuring t is at rest in the corresponding inertial frame, this becomes

$$\frac{dt}{d\tau} = (-g_{00})^{-1/2} \tag{1.44}$$

Now, consider two observes A and B that are tangent to $K^0 = (1,0,0,0)$, such that they are both at rest in some frame. For each observer, it is clear that

$$dt = d\tau_A(-g_{00}(A))^{-1/2}, \quad dt = d\tau_B(-g_{00}(B))^{-1/2} \tag{1.45}$$

This means that the observed frequencies are related by

$$\frac{\nu_B}{\nu_A} = \frac{d\tau_A}{d\tau_B} = \sqrt{\frac{g_{00}(A)}{g_{00}(B)}} \approx 1 - \frac{\Phi_B - \Phi_A}{c^2} \tag{1.46}$$

where we have assumed that we are in the Newtonian limit; we have thus obtained the previous gravitational redshift result directly from the metric.

1.3 Curvature

Throughout the previous section, we made many references to 'curved spacetime' without ever having actually demonstrating that spacetime is curved. In this section, we aim to demonstrate how this curvature is related to the energy and momentum of matter and radiation present within the system.

1.3.1 The Covariant Derivative

How quantities change in curved space is not a trivial affair, as one has to take into account of both how the quantity varies, but also how the metric changes as this quantity is varied. As such, we need a different concept of the 'derivative' operation for curved spacetime. Consider the contravariant and covariant coordinate transformations

$$\text{Contravariant:} \quad dx'^{\mu} = \frac{\partial x'^{\mu}}{\partial x^{\nu}} dx^{\nu} \tag{1.47}$$

$$\text{Covariant:} \quad dx'_{\mu} = \frac{\partial x^{\nu}}{\partial x'^{\mu}} dx_{\nu} \tag{1.48}$$

This further means that contravariant and covariant quantities will also have different derivative type operations in this space.

Now, define $\mathsf{U}^{\rho} = dx^{\rho}/d\tau$ in the geodesic equation (1.20):

$$0 = \frac{d\mathsf{U}^{\rho}}{d\tau} + \Gamma^{\rho}_{\mu\nu}\mathsf{U}^{\mu}\mathsf{U}^{\nu} = \frac{dx^{\mu}}{d\tau}\frac{\partial \mathsf{U}^{\rho}}{\partial x^{\mu}}\Gamma^{\rho}_{\mu\nu}\mathsf{U}^{\mu}\mathsf{U}^{\nu} = \mathsf{U}^{\mu}\left(\frac{\partial \mathsf{U}^{\rho}}{\partial x^{\mu}} + \Gamma^{\rho}_{\mu\nu}\mathsf{U}^{\nu}\right) \tag{1.49}$$

Since U^{μ} is a well-defined vector, the term in brackets must be must be a tensor, as it contracts with this four-vector to produce another vector (that is identically equal to zero). Now, the connection $\Gamma^{\rho}_{\mu\nu}$ vanishes in local inertial coordinates, in which partial derivatives of four-vectors are indeed tensors. This leads us to define the *covariant derivative* as

$$\boxed{\nabla_{\mu}\mathsf{U}^{\rho} = \partial_{\mu}\mathsf{U}^{\rho} + \Gamma^{\rho}_{\mu\nu}\mathsf{U}^{\nu}} \tag{1.50}$$

The covariant derivative is sometimes notated as $\nabla_{\mu}\mathsf{U}^{\rho} \equiv \mathsf{U}^{\rho}_{;\mu}$, with normal particle derivatives being notated by $\partial_{\mu}\mathsf{U}^{\rho} \equiv \mathsf{U}^{\rho}_{,\mu}$. If we require successive derivatives, we can notate these by including more indices after the comma.

As anticipated above, the covariant derivative takes a different form when operating on covariant objects, or tensors of mixed rank. Though we will not show it here, it is easy to show that the covariant derivative behaves as

$$\nabla_{\mu}\mathsf{U}_{\rho} = \partial_{\mu}\mathsf{U}_{\rho} - \Gamma^{\nu}_{\mu\rho}\mathsf{U}_{\nu} \tag{1.51}$$

$$\nabla_{\mu}\mathbb{T}^{\rho\sigma} = \partial_{\mu}\mathbb{T}^{\rho\sigma} + \Gamma^{\rho}_{\mu\nu}\mathbb{T}^{\nu\sigma} + \Gamma^{\sigma}_{\mu\nu}\mathbb{T}^{\rho\nu} \tag{1.52}$$

$$\nabla_{\mu}\mathbb{T}^{\rho}{}_{\sigma} = \partial_{\mu}\mathbb{T}^{\rho}{}_{\sigma} + \Gamma^{\rho}_{\mu\nu}\mathbb{T}^{\nu}{}_{\sigma} - \Gamma^{\nu}_{\mu\sigma}\mathbb{T}^{\rho}{}_{\nu} \tag{1.53}$$

$$\nabla_{\mu}\mathbb{T}_{\rho\sigma} = \partial_{\mu}\mathbb{T}_{\rho\sigma} - \Gamma^{\nu}_{\mu\rho}\mathbb{T}_{\nu\sigma} - \Gamma^{\nu}_{\mu\sigma}\mathbb{T}_{\rho\nu} \tag{1.54}$$

Note that the covariant derivative is constructed in such a way that is satisfies the *metricity* condition that $\nabla_{\rho}g_{\mu\nu} = 0$.

The importance of covariant differentiation arises from two of its properties: that is converts tensors to other tensors, and that it reduces to ordinary differentiation in the absence of gravitation (where the connections vanish). These properties suggest the following algorithm for inserting the effects of gravitation on physical systems: write the appropriate special-relativistic equations that hold in the absence of gravitation, replace $\eta_{\mu\nu}$ with $g_{\mu\nu}$, and promote all derivatives to covariant derivatives. The resulting equations will be generally covariant and true in the absence of gravitation, and therefore (according to the principle of General Covariance), they will be true in the presence of gravitational fields, provided that we always work on a spacetime scale sufficiently small compared with the scale of variation of the gravitational field.

More on the Covariant Derivative

In a similar way to partial derivatives, we can also define the *covariant divergence* as

$$\nabla_\mu \mathsf{U}^\mu = \partial_\mu \mathsf{U}^\mu + \Gamma^\mu_{\mu\nu}\mathsf{U}^\nu \tag{1.55}$$

Recalling (1.25), we can write that

$$\Gamma^\mu_{\mu\nu} = \frac{1}{2g_{\mu\rho}}\partial_\nu g_{\rho\mu} = \frac{1}{2}\partial_\nu \log(-g) \tag{1.56}$$

where g is the determinant of $g_{\mu\nu}$, which for diagonal metrics is simply the product of entries. Though the above equation seems specific to diagonal metrics, it is in fact a general result. Then, our covariant divergence becomes

$$\boxed{\nabla_\mu \mathsf{U}^\mu = \frac{1}{\sqrt{-g}}\partial_\mu\left(\sqrt{-g}\mathsf{U}^\mu\right)} \tag{1.57}$$

We can write a similar expression for the covariant divergence of a tensor

$$\nabla_\mu \mathbb{T}^{\mu\nu} = \frac{1}{\sqrt{-g}}\partial_\mu\left(\sqrt{-g}\mathbb{T}^{\mu\nu}\right) + \Gamma^\nu_{\mu\rho}\mathbb{T}^{\mu\rho} \tag{1.58}$$

For an antisymmetric tensor, the last term disappears as the connection is symmetric in its lower indices. We shall make use of this expression in due course.

Parallel Transport

Given that $\mathsf{U}^\mu = dx^\mu/d\tau$ is a tangent vector to our spacetime coordinates, we define the absolute derivative of another vector V^ρ along a path $x^\mu(\tau)$ as

$$\frac{D\mathsf{V}^\rho}{D\tau} = \mathsf{U}^\mu\nabla_\mu\mathsf{V}^\rho \equiv \nabla_{\mathsf{U}}\mathsf{V}^\rho, \quad \frac{D}{D\tau} = \mathsf{U}^\mu\nabla_\mu = \nabla_{\mathsf{U}} \tag{1.59}$$

We say that the vector V^μ undergoes *parallel transport* when moving along this curve in the case that $D/D\tau = 0$. For example, a vector may always point along $\mathbf{e}_y$ as we move it in the xy plane, but its corresponding r and θ components will constantly have to change to keep this true. This is embodied in the parallel transport condition. Now, if we apply this to the tangent vector U^μ, then we have that

$$\mathsf{U}^\mu\nabla_\mu\mathsf{U}^\rho = \mathsf{U}^\mu\left(\partial_\mu\mathsf{U}^\rho + \Gamma^\rho_{\mu\nu}\mathsf{U}^\nu\right) = \frac{D\mathsf{U}^\rho}{D\tau} = 0 \tag{1.60}$$

by the definition of U^μ. However, we can recognise the second expression as the geodesic equation; this means that an alternative way of writing it is

$$\boxed{\mathsf{U}^\mu\nabla_\mu\mathsf{U}^\rho = 0, \quad \mathsf{U}^\rho = \frac{dx^\rho}{d\tau}} \tag{1.61}$$

Volume Elements

How do volumes transform within our space? Suppose that we have two sets of coordinates denoted by x and x'. Then, the volume elements transform according to

$$d^4x' = \left|\frac{\partial x'}{\partial x}\right| d^4x \tag{1.62}$$

where we have used $|\partial x'/\partial x|$ to denote the Jacobean transformation between the two sets of coordinates. Now consider the transformation of the metric:

$$g'_{\mu\nu} = g_{\rho\sigma}\frac{\partial x^\rho}{\partial x'^\mu}\frac{\partial x^\sigma}{\partial x'^\nu} \quad \longrightarrow \quad g' = \left|\frac{\partial x'}{\partial x}\right|^2 g \tag{1.63}$$

where the second expression follows from taking the determinant of both sides of the first. Then, we have that

$$\sqrt{-g'}\, d^4x' = \sqrt{-g}\left|\frac{\partial x}{\partial x'}\right|\left|\frac{\partial x'}{\partial x}\right| d^4x = \sqrt{-g}\, d^4x \tag{1.64}$$

This means that $\sqrt{-g}d^4x$ is the invariant volume associated with curved spacetime. In Minkowski space, the metric does not transform, and so the usual volume element is itself invariant.

1.3.2 Stress Energy and Hydrostatic Equilibrium

As in electromagnetism in Special Relativity, the stress energy tensor $\mathbb{T}^{\mu\nu}$ serves as the source term for gravitation within General Relativity. In many cases, we are interested in the stress-energy tensor that is associated with an ideal, isotropic fluid. Suppose that U^μ is the four-velocity of a particle within the fluid, and ρ is the local mass density. Then, we can write that

$$\mathbb{T}^{\mu\nu} = \rho\,\langle \mathsf{U}^\mu \mathsf{U}^\nu\rangle \tag{1.65}$$

The brackets denote an ensemble average, which is the same as the time average if we assume that the system is ergodic. Assuming that the gas is isotropic, all off-diagonal terms must be zero, so we can write the stress energy tensor in the rest frame as

$$\mathbb{T}^{\mu\nu} = \begin{pmatrix} \rho c^2 & 0 & 0 & 0 \\ 0 & p & 0 & 0 \\ 0 & 0 & p & 0 \\ 0 & 0 & 0 & p \end{pmatrix} \tag{1.66}$$

where p is the pressure of the fluid. If we can find a tensor that agrees with $\mathbb{T}^{\mu\nu}$ in one frame, then it must agree in all frames, as the tensor that corresponds to the difference between these two tensors must remain zero under transformations. We propose a solution of the form

$$\mathbb{T}^{\mu\nu} = c_1 g^{\mu\nu} + c_2 \mathsf{U}^\mu \mathsf{U}^\nu \tag{1.67}$$

with c_1 and c_2 constants to be determined. Evaluating this in the rest frame, it is clear that the spatial components allow us to restrict $c_1 = p$. Looking at the time component, it is clear that

$$\rho c^2 = -p + c_2 c^2 \quad \longrightarrow \quad c_2 = \rho + \frac{p}{c^2} \tag{1.68}$$

This means that our expression for the stress energy tensor of an *isotropic, ideal fluid* is given by

$$\boxed{\mathbb{T}^{\mu\nu} = p g^{\mu\nu} + \left(\rho + \frac{p}{c^2}\right)\mathsf{U}^\mu \mathsf{U}^\nu} \tag{1.69}$$

Hydrostatic Equilibrium

In order for it to obey energy conservation, the covariant divergence of this tensor must be zero, such that

$$0 = \mathbb{T}^{\mu\nu}_{;\mu} = g^{\mu\nu}\partial_\mu p + \left[\left(\rho + \frac{p}{c^2}\right)\mathsf{U}^\mu\mathsf{U}^\nu\right]_{;\mu} \tag{1.70}$$

We can make use of (1.58) to re-write the second term

$$0 = g^{\mu\nu}\partial_\mu p + \frac{1}{\sqrt{-g}}\partial_\mu\left[\sqrt{-g}\left(\rho + \frac{p}{c^2}\right)\mathsf{U}^\mu\mathsf{U}^\nu\right] + \Gamma^\nu_{\mu\lambda}\left(\rho + \frac{p}{c^2}\right)\mathsf{U}^\mu\mathsf{U}^\lambda \tag{1.71}$$

By the definition of static equilibrium, all components of U^μ and U^ν must vanish except for U^0. We know that $g_{\mu\nu}\mathsf{U}^\mu\mathsf{U}^\nu = -c^2$, we can write that

$$\left(\mathsf{U}^0\right)^2 = -\frac{c^2}{g_{00}}, \quad \Gamma^\nu_{00} = -\frac{1}{2}g^{\mu\nu}\partial_\mu g_{00} \tag{1.72}$$

Plugging these expressions into (1.71), we have that

$$0 = g^{\mu\nu}\left[\partial_\mu p + (\rho c^2 + p)\partial_\mu \log\sqrt{-g_{00}}\right] \tag{1.73}$$

However, as $g^{\mu\nu}$ has a well-defined inverse, this means that the term in brackets must be zero, namely that

$$\boxed{\partial_\mu p + (\rho c^2 + p)\partial_\mu \log\sqrt{-g_{00}} = 0} \tag{1.74}$$

This is known as the *equation of hydrostatic equilibrium.* In the Newtonian limit, this clearly reduces to

$$\partial_i p + \rho\partial_i\Phi = 0 \tag{1.75}$$

which is the familiar equation for hydrostatic equilibrium.

Model a neutron star atmosphere by the simple equation of state $p = K\rho^\gamma$, where γ is the adiabatic index, and K a constant. If $\rho = \rho_0$ at the surface $r = r_0$, solve the equation of hydrostatic equilibrium. What is the Newtonian limit of the expression you obtain?

Given that $p = p(\rho)$, we can use the chain rule to write the equation for hydrostatic equilibrium as

$$\partial_\mu\left[d\rho\,\frac{\partial p/\partial\rho}{(\rho c^2 + p)} + \log\sqrt{-g_{00}}\right] = 0 \quad \longrightarrow \quad \int d\rho\,\frac{\partial p/\partial\rho}{(\rho c^2 + p)} + \log\sqrt{-g_{00}} = 0 \tag{1.76}$$

Performing the integration of the first expression under the given equation of state gives

$$\log\left(1 + \frac{K}{c^2}\rho^{\gamma-1}\right) + c_1 = \log\left(\frac{1}{(1 - r_s/r)^\alpha}\right) \tag{1.77}$$

where we have defined $\alpha = (\gamma - 1)/2\gamma$, and $r_s = 2GM/c^2$ (the *Schwarzchild radius*, a quantity that we shall meet again later). Solving for the constant using the constraints, our equation becomes

$$\frac{1 + c_s^2/\gamma c^2}{1 + c_{s0}^2/\gamma c^2} = \left[\frac{1 - r_s/r_0}{1 - r_s/r}\right]^\alpha \tag{1.78}$$

where we have introduced the speed of sound $c_s = (\gamma p/\rho)^{1/2}$, $c_{s0} = (\gamma p_0/\rho_0)^{1/2}$. In the Newtonian limit, we expand both sides for $c \gg c_s$ and $r_s \ll r$. Keeping only first order terms, we obtain the expression:

$$c_s^2 - c_{s0}^2 = (\gamma - 1)\left[\Phi(r) - \Phi(r_0)\right], \quad \Phi(r) = \frac{GM}{r} \tag{1.79}$$

1.3.3 The Curvature Tensor

Like partial derivatives, the covariant derivative is a linear operator, and it obeys the product rule. However, where it differs from ordinary differentiation is in the way that it commutes. In fact, we have that

$$[\nabla_c, \nabla_d]\, \mathsf{V}^a = (\nabla_c\nabla_d - \nabla_d\nabla_c)\mathsf{V}^a = R^a{}_{bcd}\mathsf{V}^b \tag{1.80}$$

where $R^a{}_{bcd}$ in the *Riemann curvature tensor* defined by

$$\boxed{R^a{}_{bcd} = \partial_c\Gamma^a_{bd} - \partial_d\Gamma^a_{bc} + \Gamma^a_{ce}\Gamma^e_{bd} - \Gamma^a_{de}\Gamma^e_{bc}} \tag{1.81}$$

We have chosen to use Roman letters to make it easier to remember the ordering of terms. To gain an understanding of the meaning of this tensor, consider a vector V_σ that undergoes parallel transport, allowing us to use (1.59) to write that

$$\frac{d\mathsf{V}_b}{d\tau} = \Gamma^e_{bc}\frac{dx^c}{d\tau}\mathsf{V}_e \quad \longrightarrow \quad \oint d\mathsf{V}_b = \oint dx^c\, \Gamma^e_{cd}\mathsf{V}_e \tag{1.82}$$

where we have integrated over a small closed curve in the vicinity of some spacetime point x_0. If our integration path remains close to this point over its entirety, we can expand the terms on the right-hand side, keeping only first order terms in the derivatives of the connections:

$$\Gamma^e_{bc}(x) = \Gamma^e_{bc}(x_0) + (x^d - x^d_0)\partial_d\Gamma^e_{bc}(x_0) + \dots \tag{1.83}$$

$$\mathsf{V}_e(x) = \mathsf{V}_e(x_0) + (x^d - x^d_0)\mathsf{V}_a(x_0)\Gamma^a_{ed}(x_0) + \dots \tag{1.84}$$

where in the second expression we have again used the parallel transport equation to evaluate the derivative of V_κ. The product of these two terms to first order is then

$$\Gamma^e_{bc}(x)\mathsf{V}_e(x) = \Gamma^e_{bc}(x_0)\mathsf{V}_e(x_0) + (x^d - x^d_0)\mathsf{V}_a(x_0)\left[\partial_d\Gamma^e_{bc}(x_0) + \Gamma^e_{bc}(x_0)\Gamma^a_{ed}(x_0)\right] \tag{1.85}$$

Then, the only non-vanishing term in the integral is given by

$$\oint d\mathsf{V}_b = \mathsf{V}_a\left[\partial_d\Gamma^e_{bc} + \Gamma^e_{bc}\Gamma^a_{ed}\right]\oint dx^c\, x^d \tag{1.86}$$

where the terms outside the integral on the right-hand side are evaluated at x_0. The terms inside the integral are antisymmetric in their indices, meaning that we can add any symmetric quantities to the terms in brackets, as these will evaluate to zero upon product with the integral. This means that we can in fact construct the Riemann tensor, such that

$$\oint d\mathsf{V}_b = \frac{1}{2}R^a{}_{bcd}\mathsf{V}_a\oint dx^c\, x^d \tag{1.87}$$

This means that the parallel transport of a vector around a closed curve does not change the vector unless the enclosed area contains a non-vanishing curvature tensor. This means that the departure of spacetime from a Minkowski structure reveals itself through the existence of the Riemann curvature tensor, as the name suggests.

We can also define some important associated quantities:

$$\text{Covariant Riemann tensor:} \quad R_{abcd} = g_{ae}R^e{}_{bcd} \tag{1.88}$$

$$\text{Ricci Tensor:} \quad R_{ab} = R^c{}_{acb} \tag{1.89}$$

$$\text{Ricci Scalar:} \quad R = g^{ab}R_{ab} \tag{1.90}$$

Suppose that we are in local inertial coordinates. Then, g_{cd} and $\partial_a g_{cd}$ vanish, meaning that all affine connections will be zero. However, the derivatives of the affine connections themselves are not necessarily zero; otherwise, we have that $R^a{}_{bcd} = 0 \longrightarrow R = 0$ in local inertial coordinates. As R is coordinate free, this means that the whole space must be flat. Thus, covariant derivatives only commute when the space is globally flat, as otherwise we still have remaining derivative terms in $R^a{}_{bcd}$.

Bianchi Identities

The Riemann tensor satisfies the a series of symmetry relationships, the last two of which are known as the *Bianchi identities.* In fully covariant form, these can be written as

$$R_{abcd} = R_{cbad} = R_{cdab} \tag{1.91}$$
$$R_{abcd} = -R_{bacd} = -R_{abdc} \tag{1.92}$$
$$R_{a[bcd]} = R_{abcd} + R_{acdb} + R_{adbc} = 0 \tag{1.93}$$
$$R_{ab[cd;e]} = R_{abcd;e} + R_{abde;c} + R_{abec;d} = 0 \tag{1.94}$$

where we have again adopted Roman letters as indices to make the properties of these identities clear.Note that the first of these implies that the Ricci tensor is symmetric in its indices. Of particular in interest to us is the last of these identities. By applying the metric throughout, raise a and contract it with c:

$$R_{bd;e} - R_{be;d} + R^c{}_{bde;c} = 0 \tag{1.95}$$

Next, raise b and contract it with d, remembering to make use of the above symmetry properties:

$$R_{;e} - R^d{}_{e;d} - R^c{}_{e;c} = R_{;e} - 2R^c{}_{e;c} = 0 \tag{1.96}$$

The last expression was obtained by making use of the fact that the labelling of our indices is arbitrary. This means that we can write this identity (with another relabelling) as

$$\left(\delta^\mu{}_\rho R - 2R^\mu{}_\rho\right)_{;\mu} \tag{1.97}$$

Raising ρ, recalling the metricity condition $g_{\rho\nu;\mu} = 0$, and diving throughout by -2, we arrive at the 'zero-divergence' form of the last Bianchi identity

$$\boxed{\left(R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R\right)_{;\mu} = 0} \tag{1.98}$$

This sort of combination of tensors crops up frequently within our study of General Relativity, as it satisfies the conservation condition that is implied when something has zero divergence.

Geodesic Deviation Equation

Suppose that we have a series of neighbouring geodesics that we shall denote by $x^\rho(\tau, s)$, where s is some parameter that describes which geodesic that we are on. Now, we define some *connecting vector* N as

$$\mathsf{N}^\rho = \frac{\partial x^\rho}{\partial s} \tag{1.99}$$

that describes how the 'separation' of neighbouring geodesics changes. Recalling $\mathsf{U} = \partial x^\rho / \partial \tau$, it is clear that N^ρ satisfies:

$$[\mathsf{U}^\rho, \mathsf{N}^\rho] \quad \longrightarrow \quad \nabla_{\mathsf{N}} \mathsf{U}^\rho - \nabla_{\mathsf{U}} \mathsf{N}^\rho = \mathsf{N}^\sigma \nabla_\sigma \mathsf{U}^\rho - \mathsf{U}^\sigma \nabla_\sigma \mathsf{N}^\rho = 0 \tag{1.100}$$

We are interested in how this quantity evolves with time, as this will allow us to find the rate at which objects separate or move closer together, and other such quantities. We thus write that

$$\frac{D^2 \mathsf{N}^\rho}{D\tau^2} = \nabla_{\mathsf{U}} \nabla_{\mathsf{U}} \mathsf{N}^\rho \tag{1.101}$$

We can simplify this further, making use of (1.61) and (1.100):

$$\begin{aligned}\nabla_{\mathsf{U}}\nabla_{\mathsf{U}}\mathsf{N}^\rho &= \nabla_{\mathsf{U}}\nabla_{\mathsf{N}}\mathsf{U}^\rho \\ &= \nabla_{\mathsf{U}}(\mathsf{N}^\mu\nabla_\mu\mathsf{U}^\rho) - \underbrace{\mathsf{N}^\mu\nabla_\mu\left(\mathsf{U}^\nu\nabla_\nu\mathsf{U}^\rho\right)}_{=0\text{ by geodesic equation}} + \underbrace{(\mathsf{N}^\mu\nabla_\mu\mathsf{U}^\nu)\nabla_\nu\mathsf{U}^\rho - \left(\mathsf{U}^\nu\nabla_\nu\mathsf{N}^\mu\right)\nabla_\mu\mathsf{U}^\rho}_{=0\text{ by }\nabla_{\mathsf{N}}\mathsf{U}=\nabla_{\mathsf{U}}\mathsf{N}} \\ &= \mathsf{N}^\mu\mathsf{U}^\nu\nabla_\nu\nabla_\mu\mathsf{U}^\rho - \mathsf{N}^\mu\mathsf{U}^\nu\nabla_\mu\nabla_\nu\mathsf{U}^\rho \\ &= -\mathsf{N}^\mu\mathsf{U}^\nu\left[\nabla_\mu,\nabla_\nu\right]\mathsf{U}^\rho \\ &= -\mathsf{R}^\rho{}_{\sigma\mu\nu}\mathsf{U}^\sigma\mathsf{N}^\mu\mathsf{U}^\nu \end{aligned} \tag{1.102}$$

Combining (1.101) and (1.102), we have that

$$\boxed{\frac{D^2\mathsf{N}^\rho}{D\tau^2} = -R^\rho{}_{\sigma\mu\nu}\mathsf{U}^\sigma\mathsf{N}^\mu\mathsf{U}^\nu} \tag{1.103}$$

This is known as the *geodesic deviation equation*. It is thus clear that the evolution of the connecting vector N^ρ is related to the curvature of the space. This is to be expected; if we consider two neighbouring geodesics moving in a very curved space, their separation will increase much more quickly than if the space was flat.

Let us consider the Newtonian limit of this equation. In this limit, the timelike components of U^σ and U^μ dominate, such that the equation becomes

$$\frac{D^2\mathsf{N}^\rho}{D\tau^2} = -c^2R^\rho{}_{0\mu0}\mathsf{N}^\mu \approx -c^2\partial_\mu\Gamma^\rho_{00}\mathsf{N}^\mu = -c^2\partial_\mu\left(\frac{1}{2}g^{\rho\sigma}\partial_\sigma g_{00}\right)\mathsf{N}^\mu = -\frac{1}{2}c^2\partial_\mu\partial^\rho g_{00}\mathsf{N}^\mu \tag{1.104}$$

Now, the time-derivative of g_{00} is zero, meaning that we are only left with the spatial components. This means that the Newtonian limit of the geodesic deviation equation can be written as

$$\frac{D^2\mathsf{N}^i}{D\tau^2} = \left(\partial_j\partial^i\Phi\right)\mathsf{N}^j \tag{1.105}$$

where we have used (1.41). This is in fact the *tidal force* that describes the difference in the way in which the gravitational potential Φ effects particles undergoing motion along separated geodesics.

1.3.4 Killing Vectors

Let K^ρ be a *Killing vector* that satisfies Killing's equation

$$\nabla_\mu K_\rho + \nabla_\rho K_\mu = 0 \tag{1.106}$$

This has a much more rigorous mathematical meaning, but here we can interpret this condition as implying that if K^ρ is indeed Killing, then the metric is independent of the corresponding quantity. For example, let us suppose that $g_{\mu\nu} = \eta_{\mu\nu}$ such that

$$-c^2d\tau^2 = -c^2dt^2 + dx^2 + dy^2 + dz^2 \tag{1.107}$$

It is clear that the metric is independent of time and space; this means that $K^0 = (1,0,0,0)$, $K^1 = (0,1,0,0)$, $K^2 = (0,0,1,0)$ and $K^3 = (0,0,0,1)$ are all Killing vectors on this metric.

Killing vectors have a particularly useful property. Consider $g_{\rho\sigma}\mathsf{U}^\sigma K^\rho$, and take its derivative along the wordline:

$$\nabla_{\mathsf{U}}(g_{\rho\sigma}\mathsf{U}^\sigma K^\rho) = \mathsf{U}^\mu\nabla_\mu\left(K_\rho\mathsf{U}^\rho\right) = \underbrace{K_\rho\mathsf{U}^\mu\nabla_\mu\mathsf{U}^\rho}_{=0\text{ by geodesic equation}} + \mathsf{U}^\mu\mathsf{U}^\rho\nabla_\mu K_\rho \tag{1.108}$$

This second term is also zero as it is a product of a symmetric $\mathsf{U}^\mu \mathsf{U}^\rho$ and an antisymmetric $\nabla_\mu K_\rho$ quantity. This allows us to conclude that along a given geodesic,

$$\boxed{g_{\mu\nu} K^\mu \mathsf{U}^\nu = \text{constant}} \tag{1.109}$$

assuming that K^μ is a Killing vector for the metric $g_{\mu\nu}$. Given a particular metric, it then becomes very easy to find constants of the motion, a technique that we shall make extensive use of in chapter 2. Note that derivative notation is often used as a shorthand to denote Killing vectors. For example, one could write $K^0 = (1, 0, 0, 0) = \partial/\partial t$.

1.4 Einstein's Field Equations

We spent the last section investigating how to formalise the way in which spacetime curves through the Riemann curvature tensor. However, we have not yet outlined how this curvature relates to the presence of mass. This relationship is embodied in the *Einstein field equations*, which we will now seek to derive using this vast theoretical apparatus that we have constructed.

As previously suggested, the source term for gravitation in General Relativity is the stress energy tensor $\mathbb{T}^{\mu\nu}$. Now, we must relate curvature to this in a way that means that the conservation condition $\mathbb{T}^{\mu\nu}_{;\mu} = 0$ is still satisfied. We already have a neat form for this given by (1.98). We thus propose the solution of the form

$$R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R = c_1\mathbb{T}^{\mu\nu} + c_2 g^{\mu\nu} \tag{1.110}$$

for constants c_1 and c_2. We can determine these constants by looking at the Newtonian limit of this equation. Once again, the dominant terms are $\mathbb{T}^{00} = \rho c^2$ and g^{00}, and we can use the weak field condition (1.37) to write that

$$\Gamma^{\rho}_{\sigma\nu} = \frac{1}{2}\eta^{\rho\mu}(\partial_\sigma h_{\mu\nu} + \partial_\nu h_{\mu\sigma} - \partial_\mu h_{\sigma\nu}) \tag{1.111}$$

$$R^{\rho}{}_{\mu\sigma\nu} = \frac{1}{2}\eta^{\rho\lambda}(\partial_\mu\partial_\sigma h_{\nu\lambda} + \partial_\nu\partial_\lambda h_{\rho\mu} - \partial_\lambda\partial_\sigma h_{\mu\nu} - \partial_\mu\partial_\nu h_{\rho\lambda}) \tag{1.112}$$

$$R_{\mu\nu} = \frac{1}{2}\eta^{\rho\lambda}(\partial_\mu\partial_\rho h_{\nu\lambda} + \partial_\nu\partial_\lambda h_{\rho\mu} - \partial_\lambda\partial_\rho h_{\mu\nu} - \partial_\mu\partial_\nu h_{\rho\lambda}) \tag{1.113}$$

To find c_1, we focus on the R^{00} component of the field equation. Assuming the fields are slowly varying, such that all time derivatives of the metric vanish, meaning that we are left with

$$R_{00} = -\frac{1}{2}\partial_\lambda\partial_\rho h_{00} = -\frac{1}{2}\nabla^2 h_{00} = \frac{1}{c^2}\nabla^2\Phi \tag{1.114}$$

where we have made use of (1.41). For the Ricci scalar, we let $c_2 = 0$, and assume that $\mathbb{T}_{ij} \approx 0$ (as once again the timelike component is dominant), such that

$$R_{ij} = \frac{1}{2}g_{ij}R = \frac{1}{2}\delta_{ij}R \tag{1.115}$$

Then, the Ricci scalar is

$$R = g^{\mu\nu}R_{\mu\nu} \approx \eta^{\mu\nu}R_{\mu\nu} = -R_{00} + R_{ii} = -R_{00} + \frac{3}{2}R \quad \longrightarrow \quad R = 2R_{00} \tag{1.116}$$

This means that the timelike component of the field equation becomes

$$2R_{00} = c_1\mathbb{T}_{00} \quad \longrightarrow \quad \nabla^2\Phi = c_1\frac{\rho c^4}{2} \tag{1.117}$$

Comparison with $\nabla^2\Phi = 4\pi G\rho$ means that our constant $c_1 = 8\pi G/c^4$. It is convention to define $c_2 = -\Lambda$, known as the *cosmological constant.* Finally, we have arrived at the Einstein field equations:

$$\boxed{G^{\mu\nu} = \frac{8\pi G}{c^4}\mathbb{T}^{\mu\nu} - \Lambda g^{\mu\nu}, \quad G^{\mu\nu} = R^{\mu\nu} - \frac{1}{2}g^{\mu\nu}R} \tag{1.118}$$

where we have introduced the field tensor $G^{\mu\nu}$. The field equations completely encapsulate the interaction between matter and spacetime curvature, and vice-versa. In essence, it describes how "space tells matter how to move, and matter tells space how to move", in a beautiful summary of the non-linear theory that is General Relativity

2. *Point Masses and Black Holes*

This chapter aims to further our study of General Relativity by examining:

- The Schwarzchild Solution
- Orbits around a Point Mass
- Black Holes
- Tests of General Relativity

Now that we have established the equations that define our study of General Relativity, it is time to examine some of their consequences. In this chapter, we will be mostly interested in how objects behave in a particular curved metric, that of the Schwarzchild metric.

2.1 The Schwarzchild Solution

Let us re-write the field equation in terms of the source term $\mathbb{T}^{\mu\nu}$:

$$R^{\mu\nu} = \frac{8\pi G}{c^4}\mathbb{S}^{\mu\nu} - \Lambda g^{\mu\nu}, \quad \mathbb{S}^{\mu\nu} = \left(\mathbb{T}^{\mu\nu} - \frac{1}{2}g^{\mu\nu}T\right) \tag{2.1}$$

where $T = \mathbb{T}^{\mu}_{\mu}$. We shall assume that we are in a vacuum far from source terms ($\mathbb{S}^{\mu\nu} = 0$), and shall ignore the cosmological constant factor Λ (as this only becomes relevant when working on cosmological scales, as in chapter 4). This means that our curvature is described by the vacuum field equation:

$$R_{\mu\nu} = 0 \tag{2.2}$$

Suppose that we want to find a spherically symmetric metric that satisfies the above equation vacuum equation, and is asymptotically flat as $r \to \infty$. We thus propose a diagonal metric of the form

$$-c^2 d\tau^2 = -B(r,t)c^2dt^2 + A(r,t)dr^2 + r^2d\theta^2 + r^2\sin^2\theta d\phi^2 \tag{2.3}$$

where A and B are arbitrary functions of r and t. Now, in order to find functional forms for A and B, we must impose the constraints given by the components of (2.2), for which we need to find the Ricci tensor. Let us begin by computing all the non-zero components of the affine connection using (1.25) and (1.26):

$$\begin{aligned}
\Gamma^r_{00} &= -\frac{1}{2}g^{rr}\partial_r g_{00} = \frac{B'}{2A} \\
\Gamma^r_{rr} &= -\frac{1}{2}g^{rr}\partial_r g_{rr} = \frac{A'}{2A} \\
\Gamma^r_{\theta\theta} &= -\frac{1}{2}g^{rr}\partial_r g_{\theta\theta} = -\frac{r}{A} \\
\Gamma^r_{\phi\phi} &= -\frac{1}{2}g^{rr}\partial_r g_{\phi\phi} = -\frac{r\sin^2\theta}{A} \\
\Gamma^r_{r0} &= \frac{1}{2}g^{rr}\partial_0 g_{rr} = \frac{\dot{A}}{2A}
\end{aligned} \tag{2.4}$$

$$\begin{aligned}
\Gamma^\theta_{\theta r} &= \frac{1}{2}g^{\theta\theta}\partial_r g_{\theta\theta} = \frac{1}{r} \\
\Gamma^\theta_{\phi\phi} &= -\frac{1}{2}g^{\theta\theta}\partial_r g_{\phi\phi} = \sin\theta\cos\theta \\
\Gamma^\phi_{\phi r} &= \frac{1}{2}g^{\phi\phi}\partial_r g_{\phi\phi} = \frac{1}{r} \\
\Gamma^\phi_{\phi\theta} &= -\frac{1}{2}g^{\phi\phi}\partial_\theta g_{\phi\phi} = \frac{\cos\theta}{\sin\theta}
\end{aligned}$$

$$\Gamma^0_{00} = \frac{1}{2}g^{00}\partial_0 g_{00} = \frac{\dot{B}}{2B} \tag{2.5}$$

$$\begin{aligned}
\Gamma^0_{0r} &= \frac{1}{2}g^{00}\partial_r g_{00} = -\frac{B'}{2B} \\
\Gamma^0_{rr} &= \frac{1}{2}g^{00}\partial_0 g_{rr} = \frac{\dot{A}}{2B}
\end{aligned} \tag{2.6}$$

Note that we have used a dash to refer to a derivative with respect to r, while a dot indicates a derivative with respect to t. Note that the only connections where a time derivative enters are (2.4), (2.5) and (2.6).

Using (1.25) and (1.26), we can write the Ricci tensor (1.89) in the useful form

$$\boxed{R_{\mu\nu} = \frac{1}{2}\partial_\mu\partial_\nu \log(-g) - \partial_\rho \Gamma^\rho_{\mu\nu} + \Gamma^\lambda_{\mu\rho}\Gamma^\rho_{\nu\lambda} - \frac{1}{2}\Gamma^\lambda_{\mu\nu}\partial_\lambda \log(-g)} \tag{2.7}$$

where as usual $g = -ABr^4 \sin^2\theta$ is the determinant of $g_{\mu\nu}$. Plugging the affine connections into this (the algebra is left as an exercise for the reader; had to get that phrase in here somewhere), we find that

$$R_{00} = -\frac{B''}{2A} + \frac{B'}{4A}\left[\frac{B'}{B} + \frac{A'}{A}\right] - \frac{B'}{rA} + \frac{1}{2}\left[\frac{\ddot{A}}{A} - \frac{\dot{A}^2}{A^2} + \frac{\dot{A}\dot{B}}{2AB} + \frac{\dot{A}B'}{2AB} + \frac{\dot{A}A'}{2A^2} + \frac{\dot{B}^2}{2AB}\right] \tag{2.8}$$

$$R_{rr} = \frac{B''}{2B} - \frac{A'B'}{4AB} - \frac{B'^2}{4B^2} - \frac{A'}{rA} - \frac{1}{2}\left[\frac{\dot{A}}{2B}\left(\frac{\dot{A}}{A} + \frac{\dot{B}}{B}\right) - \frac{\dot{A}B'}{AB} + \frac{\dot{B}^2}{2B^2}\right] \tag{2.9}$$

$$R_{0r} = -\frac{\dot{A}}{2A} \tag{2.10}$$

$$R_{0\theta} = R_{0\phi} = R_{r\phi} = R_{r\theta} = 0 \tag{2.11}$$

Note that we have not evaluated $R_{\theta\theta}$ and $R_{\phi\phi}$ as these do not actually provide constraints on A or B. This could have been guessed at by the form of the metric (2.3). Now, all of these terms have to be identically zero according to (2.2). This immediately implies that $\dot{A} = 0$, such that R_{00} and R_{rr} become

$$R_{00} = -\frac{B''}{2A} + \frac{B'}{4A}\left[\frac{B'}{B} + \frac{A'}{A}\right] - \frac{B'}{rA} + \frac{\dot{B}^2}{4AB} \tag{2.12}$$

$$R_{rr} = \frac{B''}{2B} - \frac{A'B'}{4AB} - \frac{B'^2}{4B^2} - \frac{A'}{rA} - \frac{\dot{B}^2}{4B^2} \tag{2.13}$$

Now, consider the following combination:

$$\frac{R_{00}}{B} + \frac{R_{rr}}{A} = -\frac{1}{rA}\left(\frac{A'}{A} + \frac{B'}{B}\right) = 0 \tag{2.14}$$

The remaining term involving time derivatives of B has vanished. This equation implies that

$$AB = \text{constant} = 1 \tag{2.15}$$

where the constant has been fixed by arguing that both A and B should return to their Minkowski values as $r \to \infty$. As $B = A^{-1}$, this further means that B must also be independent of time. This is statement of *Birkhoff's theorem*: that any spherically symmetric solution of the vacuum field equation (2.2) must be static, and asymptotically flat. This is interesting; even when we introduced a possible time dependence into our metric, we were forced by the nature of the equations describing spacetime to abandon it!

The condition $R_{00} = 0$ now gives us that

$$B'' + \frac{2B'}{r} = 0 \tag{2.16}$$

meaning that B is a linear combination of a constant and another constant times $1/r$. However, we have the additional constraint that B must approach the Newtonian limit (1.41) for $r \to \infty$. This leads us to conclude that

$$B(r) = 1 - \frac{2GM}{rc^2}, \quad A(r) = \left(1 - \frac{2GM}{rc^2}\right)^{-1} \tag{2.17}$$

2.1.1 The Schwarzchild Metric

All our work in the previous section allows us to finally write down the *Schwarzchild metric*

$$\boxed{-c^2d\tau^2 = -\left(1-\frac{2GM}{rc^2}\right)c^2dt^2 + \left(1-\frac{2GM}{rc^2}\right)^{-1}dr^2 + r^2d\theta^2 + r^2\sin^2\theta d\phi^2} \tag{2.18}$$

We can also write this explicitly as a matrix:

$$g_{\mu\nu} = \begin{pmatrix} -\left(1-\frac{2GM}{rc^2}\right) & 0 & 0 & 0 \\ 0 & \left(1-\frac{2GM}{rc^2}\right)^{-1} & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2\sin^2\theta \end{pmatrix} \tag{2.19}$$

The Schwarzchild metric is the unique spherically symmetric vacuum metric, as per Birkhoff's theorem. As such, it sees a lot of use in General Relativity when dealing with objects that can be modelled as point masses. It is often useful to introduce the *Schwarzchild radius* defined by

$$\boxed{r_s = \frac{2GM}{c^2}} \tag{2.20}$$

where M corresponds to the mass contained inside the radius r_s. It will not have escaped the readers attention that the metric is apparently singular at $r = r_s$. However, this is not a true singularity; it is simply a function of our choice of coordinates. For example, consider the metric for the surface of the unit sphere

$$-c^2d\tau^2 = d\theta^2 + \sin^2\theta d\phi^2 \tag{2.21}$$

and change coordinates to $x = \sin\theta$, such that

$$-c^2d\tau^2 = \frac{dx^2}{1-x^2} + x^2d\phi^2 \tag{2.22}$$

This has an apparent singularity at $x = 1$, but in reality this is not the case. Since x is simply the cylindrical radius, the singularity simply reflects that at the equator, x has reached its maximum value. This is how we must interpret the Schwartzhcild radius (as a coordinate singularity), though it does have some interesting properties that we shall examine in section 2.3.

How large exactly is this critical radius r_s? We can re-formulate it in terms of the solar mass $M_\odot$ as

$$r_s = 2.95(M/M_\odot)\,\text{km} \tag{2.23}$$

which is well inside any normal star or massive object. In fact, we can define a black hole as being a compact object whose Schwarzchild radius is part of the external vacuum, rather than residing in its interior. This gives rise to a true singularity at $r = 0$. We shall investigate the significance of the Schwarzchild radius for black holes further in section 2.3.

2.2 Orbits around a Point Mass

Now that we have an explicit metric to work with, we can begin investigating how objects move within said metric. In most cases, we will be considering the trajectories of light objects moving in the gravitational field of much larger objects, such as stars or black holes.

2.2.1 Classical Orbits

Before tackling the problem of orbits in the Schwarzchild metric, let us refresh our knowledge of classical orbits. Recall the classical orbit equation

$$\frac{1}{2}\left(\frac{dr}{dt}\right)^2 + \frac{J^2}{2r^2} - \frac{GM}{r} = \mathcal{E}_0 \tag{2.24}$$

which we can derive from energy conservation considerations. $J = r^2 d\phi/dt$ is the (constant) angular momentum per unit mass of the orbiting body, while $\mathcal{E}_0$ is some constant energy corresponding to the initial kinetic energy per unit mass of the orbiting body. We can then write that

$$\frac{dr}{dt} = \frac{dr}{d\phi}\frac{d\phi}{dt} = \frac{J}{r^2}\frac{dr}{d\phi} = \pm\left[2\mathcal{E}_0 + \frac{2GM}{r} - \frac{J^2}{r^2}\right]^{1/2} \tag{2.25}$$

We can then integrate both sides of this equation:

$$\pm\,\phi = \int dr\,\frac{J}{r^2\left[2\mathcal{E}_0 + \frac{2GM}{r} - \frac{J^2}{r^2}\right]^{1/2}} = \int du\,\frac{1}{\left[\frac{2\mathcal{E}_0}{J^2} + \frac{2GM}{J^2}u - u^2\right]^{1/2}} \tag{2.26}$$

The second expression follows from making the usual substitution $u = r^{-1}$. With some re-arrangement, this becomes a recognisable integral

$$\pm\,\phi = \int du\,\frac{1}{\left[\frac{2\mathcal{E}_0}{J^2} + \frac{G^2M^2}{J^4} - \left(u - \frac{GM}{J^2}\right)^2\right]^{1/2}} = \cos^{-1}\left[\frac{u - \frac{GM}{J^2}}{\left(\frac{2\mathcal{E}_0}{J^2} + \frac{G^2M^2}{J^4}\right)^{1/2}}\right] \tag{2.27}$$

meaning that we can write this as

$$u - \frac{GM}{J^2} = \frac{GM}{J^2}\left(1 + \frac{2\mathcal{E}_0 J^2}{G^2M^2}\right)^{1/2}\cos\phi \tag{2.28}$$

Defining the *elliptical eccentricity* e, we can recover the solution for r in terms of constants of the motion, and the phase angle $\phi = \Omega t$.

$$\boxed{r = \frac{J^2}{GM(1 + e\cos\phi)}, \quad e = \left(1 + \frac{2\mathcal{E}_0 J^2}{G^2M^2}\right)^{1/2}} \tag{2.29}$$

Note that we often define the *latus-rectum* ℓ such that

$$r = \frac{\ell}{1 + e\cos\phi}, \quad r_+ = \frac{\ell}{1+e}, \quad r_- = \frac{\ell}{1-e} \tag{2.30}$$

Then, supposing that the *semi-major* axis of the ellipse is a, it follows that

$$a = \frac{1}{2}(r_+ + r_-) \quad\longrightarrow\quad \ell = a(1-e^2), \quad J = [GMa(1-e^2)]^{1/2} \tag{2.31}$$

This means that we can write the rate of change of ϕ in terms of constants of the motion, and r:

$$\boxed{\Omega = \frac{d\phi}{dt} = \frac{J}{r^2} = \frac{[GMa(1-e^2)]^{1/2}}{r^2}} \tag{2.32}$$

2.2.2 Orbits in Schwarzchild

Let us now consider orbits in Schwarzchild spacetime. The key to constructing an orbit equation is to find the constants of the motion. Now, we can do this through multiple ways: finding an appropriate Lagrangian and applying the Euler-Lagrange equations (1.31), integration the covariant form of the geodesic equation (1.35), or by making use of what we learnt about Killing vectors in section 1.3.4. We shall choose to do the latter.

The Schwarzchild metric (2.18) is clearly independent of both time and ϕ, meaning that there are two constants of the motion associated with the Killing vectors

$$K^0 = (1,0,0,0) \quad \text{and} \quad K^\phi = (0,0,0,1) \tag{2.33}$$

We can employ (1.109) for each of these in turn, as detailed below.

- $K^0 = (1,0,0,0)$:

$$g_{00}K^0\mathsf{U}^0 = \left(1-\frac{r_s}{r}\right)\frac{dt}{d\lambda} = E \tag{2.34}$$

  where E is some constant of the motion related to the energy that the body has for $r \to \infty$. Note that in asymptotically flat space-time, $E \to 1$.

- $K^\phi = (0,0,0,1)$:

$$g_{\phi\phi}K^\phi\mathsf{U}^\phi = r^2\sin^2\theta\frac{d\phi}{d\lambda} = \text{constant} \tag{2.35}$$

  Without loss of generality, restrict our consideration to the orbital plane at $\theta = \pi/2$, such that $d\theta/d\lambda = 0$. The constant is clearly related to angular momentum, and so we define

$$r^2\frac{d\phi}{d\lambda} = J \tag{2.36}$$

  as the angular momentum per unit mass, in direct analogy with the classical case

To find the orbital equation, we make use of the fact that the interval is also an invariant of the motion. Taking the derivatives with respect to our affine parameter λ throughout, we find that

$$-c^2\left(\frac{d\tau}{d\lambda}\right)^2 = -c^2\left(1-\frac{r_s}{r}\right)\left(\frac{dt}{d\lambda}\right)^2 + \left(1-\frac{r_s}{r}\right)^{-1}\left(\frac{dr}{d\lambda}\right)^2 + \cancel{r^2\left(\frac{d\theta}{d\lambda}\right)^2} + r^2\sin^2\theta\left(\frac{d\phi}{d\lambda}\right)^2 \tag{2.37}$$

We now define another constant $k = d\tau/d\lambda$, which has value $k=1$ for massive geodesics, and $k=0$ for null geodesics. Furthermore, letting $u = r^{-1}$, we have that

$$\frac{dr}{d\tau} = \frac{J}{r^2}\frac{dr}{d\phi} = -J\frac{du}{d\phi} \tag{2.38}$$

such that our orbit equation becomes

$$\boxed{\left(\frac{du}{d\phi}\right)^2 + u^2(1-r_s u) - \frac{c^2k^2}{J^2}r_s u = \frac{c^2}{J^2}(E^2-k^2)} \tag{2.39}$$

Note that this equation is valid for all types of geodesics, whether massive or null, as we have encoded this information in our constant k.

Analysing the Effective Potential

We can define the effective potential for the Schwarzchild orbit equation as

$$\boxed{V_{\text{eff}}(u) = u^2(1 - r_s u) - \frac{c^2 k^2}{J^2} r_s u} \tag{2.40}$$

It is clear that this is the same as the classical effective potential, but with an extra correction due to the effects of General Relativity. Now, we know that the classical effective potential always has a minimum (which can be verified by simple differentiation). What about this potential?

$$\frac{\partial V_{\text{eff}}}{\partial u} = 0 \quad \longrightarrow \quad u^2 - \frac{2}{3r_s} u + \frac{c^2 k^2}{3J^2} = 0 \tag{2.41}$$

This has solution

$$u = \frac{1}{r} = \frac{1}{3r_s}\left(1 \pm \sqrt{1 - 3\left(\frac{ckr_s}{J}\right)^2}\right) \tag{2.42}$$

Let us examine the consequences of this equation for massive and null geodesics:

- Massive geodesics ($k = 1$) - This clearly has two solutions, corresponding to the extrema of an elliptical orbit, as in the the last potential curve in figure 2.1. For the special value of $J^2 = 3c^2 r_s^2$, there is only one solution, which corresponds to the minimum stable circular orbit of $r = 3r_s$. If the expression inside the square root is negative (corresponding to small J), then there will be no stable orbit, corresponding to plunging and scattering orbits.
- Null geodesics ($k = 0$) - In this case, we only have two possible solutions; $u = 0$ or $u = 2/3r_s$. This means that null orbits are inherently unstable; we have actually found the unstable maximum of the potential curve for null orbits.

It is interesting that in both cases, there is no stable orbit below a particular value, or none at all in the case of null orbits. This is a departure from the classical case, where we are used to being able to achieve orbit at any radius given sufficiently large values of E and J.

Circular Orbits

Let us now consider the case of circular orbits. As $dr = d\theta = 0$, we can write from (2.37) that

$$-c^2 = -c^2\left(1 - \frac{r_s}{r}\right)\left(\frac{dt}{d\tau}\right)^2 + r^2\left(\frac{d\phi}{d\tau}\right)^2 \tag{2.43}$$

Differentiating with respect to r:

$$0 = -\frac{c^2 r_s}{r^2}\left(\frac{dt}{d\tau}\right)^2 + 2r\left(\frac{d\phi}{d\tau}\right)^2 \tag{2.44}$$

This can be re-arranged to give

$$\boxed{\Omega = \frac{d\phi}{dt} = \sqrt{\frac{GM}{r^3}}} \tag{2.45}$$

Interestingly, this means that outside the minimum stable circular orbit, circular orbits in Schwarzchild satisfy Kepler's second law. We can use the definition of J (2.36) to write it in terms of Ω:

$$J = r^2 \frac{d\phi}{d\tau} = r^2 \frac{dt}{d\tau}\Omega \tag{2.46}$$

However, recalling (2.34), we have that

$$\frac{dt}{d\tau} = \left(1 - \frac{r_s}{r}\right)^{-1} \frac{E}{mc^2} \tag{2.47}$$

such that we can write J in terms of E and r as

$$J = \frac{r^2 \Omega E}{c^2(1 - r_s/r)} \tag{2.48}$$

Substituting this into (2.39) for $du/d\phi = 0$, one can show (with some algebra) that the energy of circular orbits corresponds to

$$\boxed{E = \frac{mc^2}{\left(1 - \frac{3r_s}{r}\right)^{1/2}}} \tag{2.49}$$

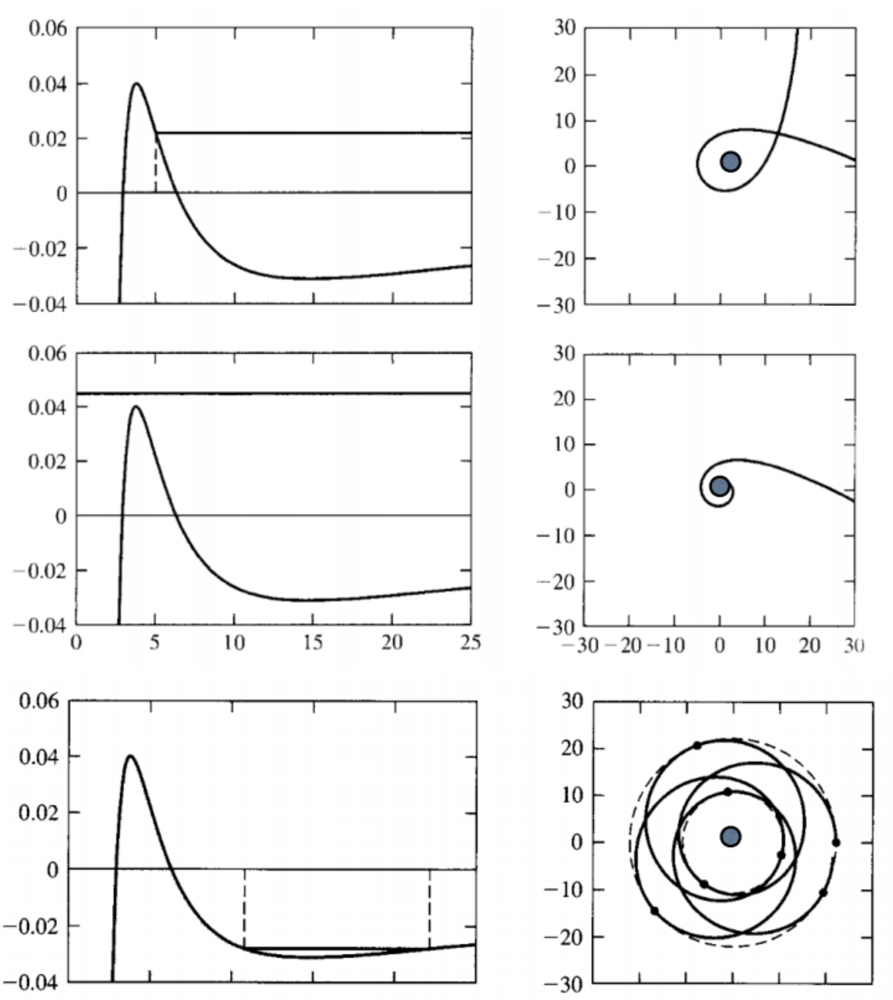

Figure 2.1: Effective potential energy curves and possible trajectories for different values of J. The first describes a scattering orbit, the second a plunging orbit, while the third elliptical (and precessing) orbits. Note the form of the effective potential curve; it does not diverge to infinity as $r \to 0$, but instead becomes quickly negative around for $r < 3r_s$

2.3 Black Holes

A black hole is a compact object whose Schwarzchild radius is part of the external vacuum, rather than residing in its interior, and contain a true singularity at $r = 0$ at which the mass is concentrated. Let us examine these entities, and in particular investigate the nature of the coordinate singularity that we have encountered at $r = r_s$. Evidently, the material derived in the previous section is still valid for black holes.

2.3.1 Massive Radial Geodesics

Initially, we consider how massive particles behave when dropped from some fixed point (r_0, θ_0, ϕ_0) far from $r = r_s$ such that they fall radially inwards. Consider (2.37) for $d\theta = d\phi = 0$:

$$\left(1 - \frac{r_s}{r}\right)\left(\frac{dt}{d\tau}\right)^2 - \frac{1}{c^2}\left(1 - \frac{r_s}{r}\right)^{-1}\left(\frac{dr}{d\tau}\right)^2 = 1 \tag{2.50}$$

From (2.34), we also have that

$$\left(1 - \frac{r_s}{r}\right)\frac{dt}{d\tau} = 1 \tag{2.51}$$

as initially $E = mc^2$ far away from the black hole (or rather, we have simply chosen our integration constant to be one). Substituting (2.51) into (2.50), we have that

$$\left(1 - \frac{r_s}{r}\right)^{-1} - \frac{1}{c^2}\left(1 - \frac{r_s}{r}\right)^{-1}\left(\frac{dr}{d\tau}\right)^2 = 1 \tag{2.52}$$

which can be solved to give

$$\boxed{\left(\frac{dr}{d\tau}\right)^2 = \frac{c^2 r_s}{r}} \tag{2.53}$$

Taking the negative solution for an inwards falling particle, we write that

$$d\tau = -\left(\frac{r}{c^2 r_s}\right)^{1/2} dr \quad \longrightarrow \quad \tau - \tau_0 = \frac{2}{3(c^2 r_s)^{1/2}}\left(r_0^{3/2} - r^{3/2}\right) \tag{2.54}$$

where r_0 and τ_0 are the initial spacetime coordinates in the frame of the particle when it was released. It is clear that there is no temporal problem with an object falling through the Schwarzchild radius, as it will reach the coordinate singularity in a finite time as measured in its own rest frame. However, the observer sitting far away from the black hole sees a different story. Consider instead

$$\frac{dr}{dt} = \frac{dr}{d\tau}\frac{d\tau}{dt} = -\left(1 - \frac{r_s}{r}\right)\left(\frac{c^2 r_s}{r}\right)^{1/2} \tag{2.55}$$

where again we have taken the negative solution for an inwardly falling particle. We can re-arrange this into an integral expression:

$$\int dt = -\int dr \, \frac{1}{\left(1 - \frac{r_s}{r}\right)\left(\frac{c^2 r_s}{r}\right)^{1/2}} \tag{2.56}$$

The integral on the right-hand side clearly diverges as $r \to r_s$. This means that from the perspective of the external observer, the particle can never actually fall past the Schwarzchild radius. It is clear that something interesting is going on at $r = r_s$. To investigate further, let us examine how light behaves as it approaches the Schwarzchild radius.

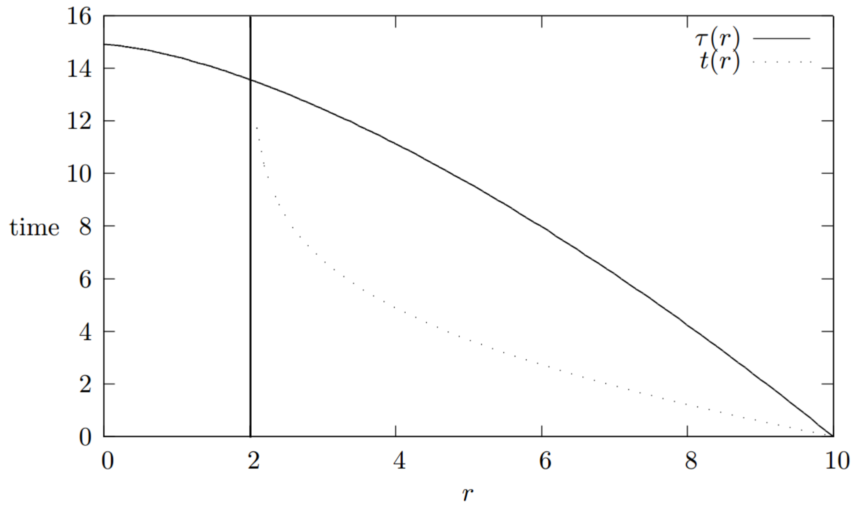

Figure 2.2: Plots of the proper time and observer time as a function of the radius, in units of GM/c^2

2.3.2 Null Radial Geodesics

Consider the metric for a radially propagating light ray

$$0 = -c^2 d\tau^2 = -\left(1 - \frac{r_s}{r}\right) c^2 dt^2 + \left(1 - \frac{r_s}{r}\right)^{-1} dr^2 \tag{2.57}$$

such that the radial null geodesic can be written as

$$\frac{dt}{dr} = \pm\frac{1}{c}\left(1 - \frac{r_s}{r}\right)^{-1} \tag{2.58}$$

This can be integrated to give

$$\boxed{\mp c(t - t_0) = (r - r_0) + r_s \log\left|\frac{r - r_s}{r_0 - r_s}\right|} \tag{2.59}$$

where r_0 and t_0 are time and radial position of the photon at time of emission. The positive and negative solutions corresponds to the outwards and inwards propagating solutions respectively. These null geodesics are plotted in figure 2.3. It is worth taking a moment to examine some of the consequences of this expression. As a particle approaches $r = r_s$, its light cones experience a spacetime equivalent of aberration, with the opening angle of the cone tending to zero, corresponding to longer times for the light to propagate to some distant observer. The observer thus sees an infinite redshift as the particle approaches r_s, something that we will later demonstrate explicitly.

Now consider light cones that are interior to r_s. From the structure of the time and radial components of (2.18), we can see that dr and dt reverse their character at the horizon, as g_{00} and g_{rr} switch signs at that point. This corresponds to a $\pi/2$ rotation of the light cones within the space. This means that inside the Schwarzchild radius, all timelike and null paths are bounded, and must encounter the true singularity at $r = 0$. This means that both light and particles are causally bound within $r = r_s$, which is the reason why this is often referred to as the *event horizon* for a black hole.

This is the first example that we have encountered where the distortion of spacetime can lead to very real, observable effects. It is quite a wonder that something so simple - yet so complex - as the distortion of spacetime can have an effect on the behaviour of physical systems. Anyway, enough reflection; onwards into more mathematical and physical rigour.

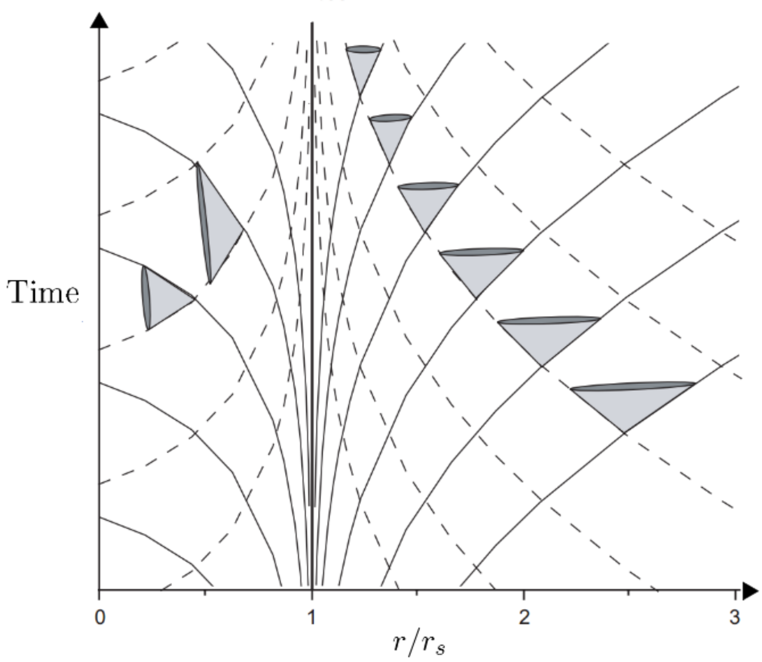

Figure 2.3: Plots of outwards (solid line) and inwards (dashed line) propagating null geodesics, showing the local light cones

Eddington-Finkelstein Coordinates

Let us re-write (2.59) in the following way

$$ct \pm (r + r_s \log|r - r_s|) = ct \pm r_* = \text{constant} \tag{2.60}$$

where we have grouped all constants on the right-hand side. Let us define a new coordinate v by

$$v = ct + r_* \tag{2.61}$$

so that v is constant along ingoing radial null geodesics. Then, we can eliminate t from our equations by $t = v - r_*$ such that

$$cdt = dv - \frac{dr}{\left(1 - \frac{r_s}{r}\right)} \tag{2.62}$$

These new coordinates are known as *Eddington-Finkelstein (EF) coordinates*. Substituting (2.62) into the Schwarzchild metric, we have that

$$-c^2 d\tau^2 = -\left(1 - \frac{r_s}{r}\right) dv^2 + 2dvdr + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2 \tag{2.63}$$

or explicitly as a matrix:

$$g_{\mu\nu} = \begin{pmatrix} -\left(1 - \frac{r_s}{r}\right) & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2\theta \end{pmatrix} \tag{2.64}$$

Unlike the metric components in Schwarzchild coordinates, the components of the metric are smooth for all $r > 0$, and in particular at $r = r_s$, confirming our prior statement in section 2.1.1 that this apparent singularity was simply due to the choice of coordinates.

The Schwarzchild spacetime can now be extended through the surface $r = r_s$ into the new region $r < r_s$. The new metric is still indeed a solution to the vacuum field equation, as the new metric is simply obtained by an analytic transformation. We note that in the new region $0 < r < r_s$, the metric is spherically symmetric. How is this consistent with Birkhoff's theorem? Recall that the Schwarzchild metric had Killing vector $K^0 = \partial/\partial t$. In EF coordinates, we have that

$$K^0 = \frac{\partial}{\partial t} = \frac{\partial x^\mu}{\partial t}\frac{\partial}{\partial x^\mu} = c\frac{\partial}{\partial v} \tag{2.65}$$

since EF coordinates are independent of t apart from through the definition $v = ct + r_*$. This can be used to extend the Schwarzchild Killing vector into $r < r_s$. Note that $(K^0)^2 = g_{vv}$, such that K^0 is null at $r = r_s$, and spacelike for $0 < r < r_s$. This means that this new extended solution is only static for $r > r_s$, meaning that Birkhoff's theorem is still satisfied.

Let us now apply our new coordinate transformation to radial null geodesics. From the EF metric (2.63), we thus have that

$$0 = -c^2 d\tau^2 = -\left(1 - \frac{r_s}{r}\right) dv^2 + 2dvdr \tag{2.66}$$

This has three solutions

- $dv = 0$, so $v =$ constant. This corresponds to ingoing light rays that follow trajectories of constant v
- $dv \neq 0$, such that we can write

$$\frac{dv}{dr} = 2\left(1 - \frac{r_s}{r}\right)^{-1} \tag{2.67}$$

  which yields upon integration

$$v - 2\left(r + r_s \log|r - r_s|\right) = \text{constant} \tag{2.68}$$

  This solution corresponds to 'outgoing' geodesics. However, the solution changes behaviour at $r = r_s$; for $r > r_s$, the solution is r increases as v increases as expected, but for $r < r_s$, r decreases as v increases. This means that for $r < r_s$, r decreases for both families of geodesics. Once again, null geodesics cannot escape from within the event horizon
- In the special case that $r = r_s$, we have that $dvdr = 0$, corresponding to light that is trapped on the surface of the event horizon

Together, these solutions allow us to view the event horizon at the Schwarzchild radius as a null surface generated by the radial light rays that can neither escape to infinity nor fall into the singularity.

2.3.3 Gravitational Redshift

Another important problem associated with black holes is that of the gravitational redshift that we associate with the Schwarzchild metric. Suppose that we have an isotropic gas that is orbiting a black hole, and emits at a particular spectrum in its own rest frame. How is this spectrum broadened for a distant observer sitting at rest relative to the black hole?

We shall move back to the Schwarzchild metric to tackle this problem. For simplicity, assume that the orbit of the gas is circular. A distant observer with coordinate time t will observe the rotating gas moving with a velocity

$$v = r\frac{d\phi}{dt} = c\sqrt{\frac{r_s}{2r}} \tag{2.69}$$

where we have made use of (2.45). Then, the local Doppler frequency shift due to this motion is given by

$$\frac{\omega_r}{\omega_e} = \sqrt{\frac{c \pm v}{c \mp v}} \tag{2.70}$$

where ω_r is the frequency observed at some radius r near the black hole, and ω_e is the emitted frequency of the radiation in the rest frame of the source. The positive and negative signs correspond to matter moving towards and away from the observer along the line of sight. For light, the (normalised) tangent to the null geodesic is given by

$$\mathsf{U}^\mu = \left(1 - \frac{r_s}{r}\right)^{-1/2}(1,0,0,0) \tag{2.71}$$

as $K^0 = \partial/\partial t$ is Killing in Schwarzchild. Let P^μ be the four-momentum of the emitted radiation. Then, we can write the invariant quantity

$$\mathsf{P}_\mu \mathsf{U}^\mu = \left(1 - \frac{r_s}{r}\right)^{-1/2}\mathsf{P}^0 \tag{2.72}$$

From this, it clearly follows that

$$\frac{\omega_\infty}{\omega_r} = \left(1 - \frac{r_s}{r}\right)^{1/2} \tag{2.73}$$

where ω_∞ is the frequency observed at $r \to \infty$. Thus, we find an equation for the relationship between the frequency emitted in the rest frame of the gas located at some radius r, and that observed by a distant observer far from the black hole:

$$\boxed{\frac{\omega_\infty}{\omega_e} = \left[\left(1 - \frac{r_s}{r}\right)\left(\frac{1 \pm \sqrt{r_s/2r}}{1 \mp \sqrt{r_s/2r}}\right)\right]^{1/2}} \tag{2.74}$$

where again the positive and negative signs correspond to matter moving towards and away from the observer along the line of sight. It is then easy to show that the line broadening is given by

$$\frac{\Delta\omega}{\omega_e} = 2\left(1 - \frac{r_s}{r}\right)^{1/2}\left(\frac{2r}{r_s} - 1\right)^{-1/2} \tag{2.75}$$

Suppose that the gas is located at the minimum stable circular orbit, so $r = 3r_s$, and thus $\Delta\omega/\omega_e = (8/15)^{1/2}$. The broadening is thus comparable to the emitted frequency close to the black hole.

2.4 Tests of General Relativity

Let us now examine some predictions made by General Relativity in the context of the Schwarzchild metric that can be measured, and thus used to test the validity of the theory. We will be making extensive use of the material covered in section 2.2.

2.4.1 Gravitational Deflection of Light

Consider a light ray originating at $r = -\infty$ that passes by a massive object of mass M at a distance of closest approach b. Classically, the light will remain undeflected, meaning that $\Delta\phi = \pi$ between $r = \pm\infty$. However, this is not the case when we take into account the effects of General Relativity.

We shall again consider the orbit equation (2.39). Differentiating throughout with respect to ϕ, and dividing by $du/d\phi$, we find that

$$\boxed{u'' + u = \frac{3GM}{c^2}u^2 + \frac{GM}{J^2}k^2} \tag{2.76}$$

where the dash indicates a derivative with respect to ϕ. For the gravitational deflection of light, we are interested in null orbits, and so we set $k = 0$:

$$u'' + u = \alpha u^2, \quad \alpha = \frac{3GM}{c^2} \tag{2.77}$$

We now let

$$u = u_0 + \delta u, \quad u_0 = \frac{\sin\phi}{b} \tag{2.78}$$

where u_0 is the classical solution for $\alpha \to 0$, corresponding to no deflection. We thus find that

$$\delta u'' + \delta u = \alpha u_0^2 = \frac{\alpha}{2b^2}(1 - \cos 2\phi) \tag{2.79}$$

This has solution

$$\delta u = \frac{\alpha}{2b^2}\left(1 + \frac{1}{3}\cos 2\phi\right) \tag{2.80}$$

meaning that

$$u = \frac{\sin\phi}{b} + \frac{\alpha}{2b^2}\left(1 + \frac{1}{3}\cos 2\phi\right) \tag{2.81}$$

As $r \to \infty$, $u \to 0$, and $\phi \to 0$, π. Under these conditions, we find that

$$\sin\phi \approx \phi = -\frac{2\alpha}{3b} \quad \text{or} \quad \frac{2\alpha}{3b} + \pi \tag{2.82}$$

This means that the overall deflection of the light over the interval $[0, \pi]$ is given by

$$\boxed{\Delta\phi = \frac{4\alpha}{3b} = \frac{4GM}{bc^2}} \tag{2.83}$$

For the sun, the correction is approximately 1.75 arseconds, famously measured by Arthur Eddington during his Eclipse expedition in 1919. We have thus demonstrated one of the more well known results of General Relativity; that the path of light rays is curved by massive objects.

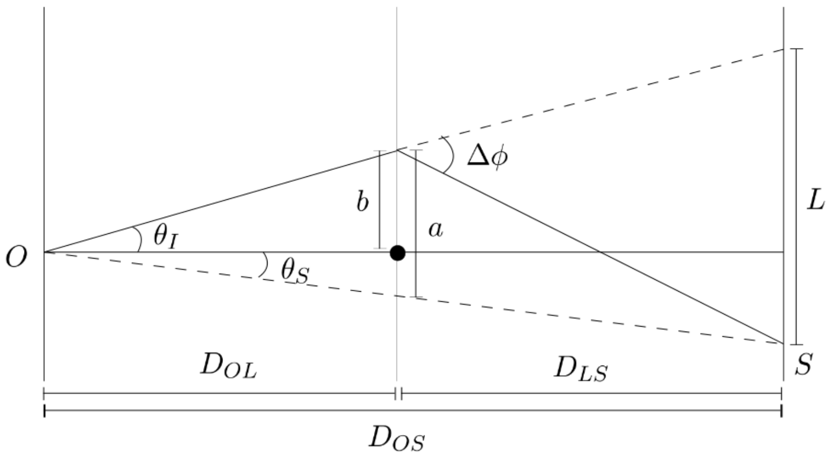

Figure 2.4: The geometry associated with gravitational lensing

Modelling Gravitational Lensing

This gravitational deflection of light in the context of astronomy is often referred to as gravitational lensing, which we will now seek to model.

Consider figure 2.4. Suppose that the source of the light is located at S, and the observer is located at O, while the massive object of mass M is indicated by the large dot in the centre. We now need to do some geometrical optics. We shall work in the small angle approximation, such that $\sin\theta \approx \theta$. From the diagram, we have that

$$\theta_I D_{OL} = b \tag{2.84}$$

$$\theta_S D_{OL} = a - b \tag{2.85}$$

$$(\theta_I + \theta_S)(D_{OL} + D_{LS}) = L \tag{2.86}$$

However, we also have that $\Delta\phi D_{LS} = L$. Combining the first two of these equations yields

$$(\theta_I + \theta_S)D_{OL} = b + (a - b) = a \tag{2.87}$$

Then:

$$(\theta_I + \theta_S)(D_{OL} + D_{LS}) = \frac{a}{D_{OL}}(D_{OL} + D_{LS}) = \Delta\phi D_{LS} \tag{2.88}$$

It follows that the distance a between the observed position of the source and its actual position in the lens plane (at D_{OL} from the observer) is given by

$$a = \frac{D_{OL}D_{LS}}{D_{OS}}\Delta\phi \tag{2.89}$$

Suppose that in the lens plane, we have N point masses of individual masses M_i. Let $\mathbf{x}$ be the point at which the light rays intersect the lens plane. Then, the total deflection will be

$$\mathbf{a}(\mathbf{x}) = \sum_i^N \Delta\phi_i \,\frac{D_{OL}D_{LS}}{D_{OS}} = \frac{D_{OL}D_{LS}}{D_{OS}}\,\frac{4G}{c^2}\sum_i^N M_i\,\frac{\mathbf{x} - \mathbf{x}'_i}{|\mathbf{x} - \mathbf{x}'_i|^2} \tag{2.90}$$

Note that the vector expression denotes the distance of closest approach of the light to a given mass i. Now, if we have some arbitrary density profile $\rho(\mathbf{x})$ in the plane of the lens, we can write this expression as an integral:

$$\mathbf{a}(\mathbf{x}) = \frac{D_{OL}D_{LS}}{D_{OS}}\,\frac{4G}{c^2}\int d^3\mathbf{x}'\,\rho(\mathbf{x}')\frac{\mathbf{x} - \mathbf{x}'}{|\mathbf{x} - \mathbf{x}'|^2} \tag{2.91}$$

In order to infer $\rho(\mathbf{x}')$, we need to actually know the source distribution, as we can observe $\mathbf{a}(\mathbf{x})$, and then invert the above expression. However, this is difficult to do in practise.

When the source is directly behind the lens plane, we obtain an *Einstein ring*, which is where the light from the source is deformed into a ring around the objects at the centre of the lens. Again looking at figure 2.4, we have that

$$(\theta_I + \theta_S)\, D_{OS} = \Delta\phi D_{LS} \quad \longrightarrow \quad \theta_I + \theta_S = \frac{4GM}{c^2\theta_I}\frac{D_{LS}}{D_{OL}D_{OS}} \tag{2.92}$$

The Einstein ring occurs at $\theta_S = 0$, which allows us to determine the angular radius of the Einstein ring:

$$\boxed{\theta_I = \left(\frac{4GM}{c^2}\frac{D_{LS}}{D_{OL}D_{OS}}\right)^{1/2}} \tag{2.93}$$

2.4.2 The Perihelion Advance of Mercury

Departures from a $1/r$ gravitational potential cause elliptical orbits not to close, which gives rise to the advance of the perihelion of the orbit. We can calculate the rate at which this occurs for massive objects using General Relativity.

In this case, we are interested in massive orbits, and so we consider (2.76) with $k = 1$:

$$u'' + u = \frac{3GM}{c^2}u^2 + \frac{GM}{J^2} \tag{2.94}$$

The Newtonian limit of this equation corresponds to ignoring the first term on the right-hand side. In this case, we are already familiar with the solution given by (2.29):

$$u_0 = \frac{GM}{J^2}(1 + e\cos\phi) \tag{2.95}$$

We thus solve (2.94) perturbatively by letting $u = u_0 + \delta u$. Making this substitution, it becomes clear that δu satisfies the differential equation

$$\frac{d^2\delta u}{d\phi^2} + \delta u = a(1 + 2e\cos\phi + e^2\cos^2\phi), \quad a = \frac{3(GM)^3}{c^2J^4} \tag{2.96}$$

We now write this as the real part of a complex differential equation

$$\frac{d^2\delta\widetilde{u}}{d\phi^2} + \delta\widetilde{u} = a\left(1 + \frac{e^2}{2} + 2ee^{i\phi} + \frac{e^2}{2}e^{2i\phi}\right) \tag{2.97}$$

where we have defined $\delta\widetilde{u}$ such that $\mathrm{Re}(\delta\widetilde{u}) = \delta u$. Using the trial substitution

$$\delta\widetilde{u} = A_0 + A_1\phi e^{i\phi} + A_2 e^{2i\phi} \tag{2.98}$$

we find that

$$\delta\widetilde{u} = a\left(1 + \frac{e^2}{2}\right) + ea\phi e^{i(\phi-\pi/2)} - \frac{e^2a}{6}e^{2i\phi} \tag{2.99}$$

Taking the real part, we can construct out final solution as

$$u \approx \frac{GM}{J^2}\left(1 + e\cos\phi + \alpha e\phi\sin\phi\right), \quad \alpha = 3(GM/Jc)^2 \ll 1 \tag{2.100}$$

where we have only kept contributions to δu proportional to ϕ, as these will dominate over the other terms. This expression corresponds to the Taylor expansion of

$$u = \frac{GM}{J^2}\left(1 + e\cos\left[\phi(1-\alpha)\right]\right) \tag{2.101}$$

This means that the period of the orbit becomes

$$T = \frac{2\pi}{1-\alpha} \approx 2\pi + 2\pi\alpha \tag{2.102}$$

which corresponds to the advance of the perihelion of the object by an amount

$$\boxed{\Delta\phi = 2\pi\alpha = 6\pi\left(\frac{GM}{Jc}\right)^2 = 6\pi\left(\frac{GM}{c^2\ell}\right)} \tag{2.103}$$

This is the classic result as per Einstein. For example, the semi-latus rectum for Mercury is $\ell = 5.546 \times 10^{10}$ m, such that $\Delta\phi = 43$ arcseconds of arc per century.

2.4.3 The Shapiro Time Delay

The Shapiro time delay was an experiment designed to test the effects of general relativity. Proposed in 1964 by Irwin Shapiro, the experiment involves sending a radio signal from Earth, that bounces off Mercury and returns. One performs the experiment when Mercury is at its closed point to the Earth, and then repeats the experiment when the planet is on the far side of its orbit. There should be an additional delay of the pulses when Mercury is on the far side of the sun because of the traversal of the radio waves across the Schwarzchild geometry of the sun. We shall investigate this effect.

As usual, we consider the metric for null geodesics:

$$0 = -c^2 d\tau^2 = -\left(1 - \frac{r_s}{r}\right)c^2dt^2 + \left(1 - \frac{r_s}{r}\right)^{-1}dr^2 + r^2d\phi^2 \tag{2.104}$$

Dividing throughout by dt^2, and using the normal constant of motion $r^2 d\phi/dt = J$, we find that

$$\left(\frac{dr}{dt}\right)^2 + \left(1 - \frac{r_s}{r}\right)\frac{J^2}{r^2} = c^2 \tag{2.105}$$

In this problem, it is conventional to notate the distance of closest approach by r_0. At $r = r_0$, $dr/dt = 0$, such that we can write

$$\left(1 - \frac{r_s}{r_0}\right)\frac{J^2}{r_0^2} = c^2 \quad\longrightarrow\quad J^2 = c^2 r_0^2\left(1 - \frac{r_s}{r_0}\right)^{-1} \tag{2.106}$$

such that our equation becomes

$$\frac{1}{c^2}\left(\frac{dr}{dt}\right)^2 + \frac{r_0^2}{r^2}\left[\frac{1 - r_s/r}{1 - r_s/r_0}\right] = 1 \tag{2.107}$$

Now, as the Schwarzchild radius for the Sun is very small, we expand this to first order in $r_s \ll r, r_0$:

$$\frac{r_0^2}{r^2}\left[\frac{1 - r_s/r}{1 - r_s/r_0}\right] \approx \frac{r_0^2}{r^2}\left[1 - r_s\left(\frac{1}{r} - \frac{1}{r_0}\right)\right] = \left(1 - \frac{r_0^2}{r^2}\right)\left(1 - \frac{r_s r_0}{r(r+r_0)}\right) \tag{2.108}$$

The relevant integral then becomes

$$\begin{aligned} t &= \frac{1}{c}\int_{r_0}^{r} dr\,\left(1-\frac{r_0^2}{r^2}\right)^{-1/2}\left(1-\frac{r_s r_0}{r(r+r_0)}\right)^{-1/2} \\ &\approx \frac{1}{c}\int_{r_0}^{r} dr\,\left(1-\frac{r_0^2}{r^2}\right)^{-1/2}\left(1+\frac{r_s}{r}+\frac{r_s r_0}{r(r+r_0)}\right) \end{aligned} \tag{2.109}$$

This then evaluates to

$$ct(r,r_0) = \left(r^2-r_0^2\right)^{1/2} + r_s\log\left[\frac{r}{r_0}+\left(\frac{r^2}{r_0^2}-1\right)^{1/2}\right] + r_s\left(\frac{r-r_0}{r+r_0}\right)^{1/2} \tag{2.110}$$

We are thus interested in the quantity $2ct(r_1,r_0) \pm 2t(r_2,r_0)$ for a path from the Earth at r_1 that is reflected from a planet at r_2. The $\pm$ sign depends on whether the planet is on the far or near side of the sun. It can be shown that the time delay is (to about 5 significant figures) the same as that given by this prediction of General Relativity.

3. *Linearised Gravity*

This chapter aims to cover the basics of linearised gravity, including

- Gravitational Waves
- Gravitational Radiation

In the previous chapter, we dealt with Schwarzchild spacetime, in which massive objects cause large distortions in the geometry of the space. Moving to the other asymptotic limit, we are going to consider the case where matter causes only small excursions from Minkowski space. This gives rise to the concept of a gravitational wave, and all its consequences.

3.1 Gravitational Waves

As we are now working in the weak-gravity limit, we adopt the same metric as (1.37)

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \quad g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu}, \quad |h_{\mu\nu}| \ll 1 \tag{3.1}$$

where $\eta_{\mu\nu}$ is the usual Minkowski metric, and $h_{\mu\nu}$ is our small linear perturbation on flat spacetime. In this section, we aim to show that adopting this definition, but without making any of the other assumptions that are associated with the Newtonian limit, leads to gravitational waves as a consequence of the resultant equations.

3.1.1 Linearising the Field Equations

We showed earlier that in the weak field limit, the Ricci tensor has the following form:

$$R_{\mu\nu} = \frac{1}{2}\eta^{\rho\lambda}(\partial_\mu\partial_\rho h_{\nu\lambda} + \partial_\nu\partial_\lambda h_{\rho\mu} - \partial_\lambda\partial_\rho h_{\mu\nu} - \partial_\mu\partial_\nu h_{\rho\lambda}) \tag{3.2}$$

Defining $\Box = \partial_\rho\partial^\rho$ and $h = h_\rho{}^\rho$, this can also be written as

$$R_{\mu\nu} = \frac{1}{2}\left(\partial_\rho\partial_\mu h^\rho{}_\nu + \partial_\rho\partial_\nu h^\rho{}_\mu - \partial_\mu\partial_\nu h - \Box h_{\mu\nu}\right) \tag{3.3}$$

Contracting with $g^{\mu\nu}$ yields the Ricci scalar:

$$R = \partial_\mu\partial_\nu h^{\mu\nu} - \Box h \tag{3.4}$$

Then, the linearised Einstein field tensor is given by

$$G_{\mu\nu} = \frac{1}{2}\left(\partial_\rho\partial_\nu h^\rho{}_\mu + \partial_\rho\partial_\mu h^\rho{}_\nu - \partial_\mu\partial_\nu h - \Box h_{\mu\nu} - \eta_{\mu\nu}\partial_\mu\partial_\nu h^{\mu\nu} + \eta_{\mu\nu}\Box h\right) \tag{3.5}$$

The linearised field equation is then evidently the above expression equal to the stress energy tensor $\mathbb{T}_{\mu\nu}$ that has been calculated to zeroth order in $h_{\mu\nu}$. We do not include higher order corrections to the stress energy tensor because the amount of energy and momentum associated with it must itself be small for the weak-field limit to apply. In other words, the lowest non-vanishing order in $\mathbb{T}_{\mu\nu}$ is automatically of the same order of the perturbation.

Gauge Invariance

In the same way that we encountered gauge invariance when dealing with electromagnetism in Special Relativity, there is also a gauge invariance associated with our perturbative metric $h_{\mu\nu}$. Consider the gauge transformation

$$h_{\mu\nu} \mapsto h_{\mu\nu} + \partial_\mu\xi_\nu + \partial_\nu\xi_\mu \tag{3.6}$$

for some field ξ^μ. We find that this gauge transformation changes the Ricci tensor by an amount

$$\begin{aligned}\delta R_{\mu\nu\rho\sigma} = \frac{1}{2}(&\partial_\rho\partial_\nu\partial_\mu\xi_\sigma + \partial_\rho\partial_\nu\partial_\sigma\xi_\mu + \partial_\sigma\partial_\mu\partial_\nu\xi_\rho + \partial_\sigma\partial_\mu\partial_\rho\xi_\nu \\ &-\partial_\sigma\partial_\mu\partial_\nu\xi_\rho - \partial_\sigma\partial_\mu\partial_\rho\xi_\nu - \partial_\rho\partial_\nu\partial_\mu\xi_\sigma - \partial_\rho\partial_\nu\partial_\sigma\xi_\mu) = 0\end{aligned} \tag{3.7}$$

This means that in fact (3.6) is in fact the appropriate gauge transformation to adopt because it leaves the curvature (and hence the physical spacetime) unchanged. When faced with a gauge invariance, it is a first instinct to fix a gauge. For this, we adopt the *harmonic gauge*:

$$\boxed{\partial^\mu \bar{h}_{\mu\nu} = \partial^\mu \left(h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h \right) = 0} \tag{3.8}$$

Note that we have defined the quantity

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h \tag{3.9}$$

which is our trace-reversed form of the perturbative metric. Using the harmonic gauge condition, it is a small amount of algebra to show that

$$G_{\mu\nu} = -\frac{1}{2}\Box \bar{h}_{\mu\nu} \tag{3.10}$$

meaning that the fields equations become (for $\Lambda = 0$)

$$\boxed{\Box \bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4}\mathbb{T}_{\mu\nu}} \tag{3.11}$$

This is clearly a wave equation for $\bar{h}_{\mu\nu}$, meaning that there must exist a wave-like solution for the departures from Minkowski spacetime that we call *gravitational waves*.

3.1.2 Vacuum Solution

Let us look for solutions to this equation in the absence of any sources, meaning that we are attempting to solve

$$\Box \bar{h}_{\mu\nu} = 0 \tag{3.12}$$

We shall adopt plane wave solutions of the form

$$\bar{h}_{\mu\nu} = C_{\mu\nu} e^{i\mathsf{K}_\rho x^\rho} \tag{3.13}$$

where $\mathsf{K}^\rho = (\omega, \mathbf{k})$ is the wavevector, and $C_{\mu\nu}$ is a constant, symmetric tensor. Substituting this solution into (3.12):

$$0 = \eta^{\sigma\eta}\partial_\sigma\partial_\eta \bar{h}_{\mu\nu} = -\mathsf{K}_\rho \mathsf{K}^\rho \bar{h}_{\mu\nu} \quad \longrightarrow \quad \mathsf{K}_\rho \mathsf{K}^\rho = 0 \tag{3.14}$$

The plane wave is thus a solution to the linearised equations if the wavevector is null; this corresponds to the statement that gravitational waves propagate at the speed of light.

Degrees of Freedom

There are a number of degrees of freedom that we currently have in our system; ten of the components of $C_{\mu\nu}$ (recalling that its symmetric), and three for the wavevector K. In order for our system to be properly constrained, must eliminate these degrees of freedom.

Imposing the harmonic gauge (3.8), we have that

$$0 = \partial_\mu \left(C^{\mu\nu} e^{i\mathsf{K}_\rho x^\rho} \right) = iC^{\mu\nu}\mathsf{K}_\mu e^{i\mathsf{K}_\rho x^\rho} \quad \longrightarrow \quad \mathsf{K}_\mu C^{\mu\nu} = 0 \tag{3.15}$$

That is, the wavevector is orthogonal to the constant vector. This comprises four equations, which reduces the number of independent components of $C_{\mu\nu}$ from ten to six.

Although we have now imposed the harmonic gauge condition, there is still some coordinate freedom left. Any coordinate transformation of the form

$$x^\mu \mapsto x^\mu + \zeta^\mu \tag{3.16}$$

will leave the harmonic coordinate condition $\Box x^\mu$ satisfied as long as $\Box \zeta^\mu = 0$. This itself is a wave equation for ζ^μ; once we choose a solution, we will have used up all of our coordinate freedom. Let us adopt a solution of the form

$$\zeta_\mu = B_\mu e^{i\mathsf{K}_\rho x^\rho} \tag{3.17}$$

where again K^ρ is the wavevector, and now B_μ is another constant vector. We now claim that this remaining freedom allows us to convert from whatever coefficients we had in $C^{(\text{old}}_{\mu\nu}$ that characterise our gravitational wave to a new set $C^{(\text{new})}_{\mu\nu}$ such that the conditions

$$C^{(\text{new})}_{0\nu} = 0, \quad C^{(\text{new})\mu}_{\mu} = 0 \tag{3.18}$$

are satisfied. This corresponds to the solution being transverse and traceless respectively. Under the coordinate transformation (3.16), the resulting change in our metric perturbation can be written can be written in the from

$$h^{(\text{new})}_{\mu\nu} = h^{(\text{old})}_{\mu\nu} - \partial_\mu \zeta_\nu - \partial_\nu \zeta_\mu \tag{3.19}$$

meaning that the trace-reversed form transforms as

$$\begin{aligned}\bar{h}^{(\text{new})}_{\mu\nu} &= h^{(\text{new})}_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu} h^{(\text{new})} \\ &= h^{(\text{old})}_{\mu\nu} - \partial_\mu \zeta_\nu - \partial_\nu \zeta_\mu - \frac{1}{2}\eta_{\mu\nu}(h^{(\text{old})} - 2\partial_\rho \zeta^\rho) \\ &= h^{(\text{old})}_{\mu\nu} - \partial_\mu \zeta_\nu - \partial_\nu \zeta_\mu + \eta_{\mu\nu}\partial_\rho \zeta^\rho \end{aligned} \tag{3.20}$$

This can be used to show that the conditions on B_μ required such that it satisfies (3.18) are

$$B_0 = -\frac{i}{2\mathsf{K}_0}\left(C^{(\text{old})}_{00} + \frac{1}{2}C^{(\text{old})\mu}_{\mu}\right) \tag{3.21}$$

$$B_j = \frac{i}{2\left(\mathsf{K}_0\right)^2}\left[-2\mathsf{K}_0 C^{(\text{old})}_{0j} + \mathsf{K}_j\left(C^{(\text{old})}_{00} + \frac{1}{2}C^{(\text{old})\mu}_{\mu}\right)\right] \tag{3.22}$$

Assuming this transformation has been made, we shall refer to the new coefficients $C^{(\text{new})}_{\mu\nu} \equiv C^{\mu\nu}$. We have now used up all our possible degrees of freedom, so we now only have two independent components that represent the physical information characterising our plane wave in this gauge.

Polarisation states

Let us now choose our spatial coordinates such that the wave is propagating along $\mathbf{e}_3$, such that

$$\mathsf{K}^\mu = (\omega, 0, 0, \omega) \tag{3.23}$$

In this case, the conditions $\mathsf{K}^\mu C_{\mu\nu} = 0$ and $C_{0\nu} = 0$ together imply that

$$C_{3\nu} = 0 \tag{3.24}$$

The only nonzero components of $C_{\mu\nu}$ are thus C_{11}, C_{22}, C_{12} and C_{21}. However, as $C_{\mu\nu}$ is traceless and symmetric, meaning that in general, we can write that

$$C_{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & C_{11} & C_{12} & 0 \\ 0 & C_{12} & -C_{11} & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{3.25}$$

This means that for a plane wave in this gauge travelling along $\mathbf{e}_3$, the two components C_{11} and C_{12} completely characterise the wave. We are now in the *transverse-traceless (TT) gauge*, for which the metric perturbation is traceless and perpendicular to the wavevector, meaning that $\bar{h}_{\mu\nu}$ and $h_{\mu\nu}$ are equivalent in this gauge.

The part of the wave that is proportional to $C_{11} \equiv C_+$ is known as the *plus polarisation* (denoted $+$), while the part proportional to $C_{12} \equiv C_\times$ is known as the *cross polarisation* (denoted by $\times$). This means that we can decompose our coefficient tensor as

$$C_{\mu\nu} = C^+_{\mu\nu} + C^\times_{\mu\nu} \tag{3.26}$$

These two quantities measure the two independent modes of linear polarisation of the gravitational wave. As we shall demonstrate explicitly in the next section, these cause a ring of particles to bounce back and forth in the shape of an '$+$' and a '$\times$' respectively. We could also consider right- and left-handed circularly polarised modes by defining

$$C_R = \frac{1}{\sqrt{2}}(C_+ + iC_\times), \quad C_L \frac{1}{\sqrt{2}}(C_+ - iC_\times) \tag{3.27}$$

The effect of a pure C_R wave would be to rotate particles in a right-handed sense, and similarly for the left-handed mode C_L. Note that individual particles do not travel around the ring, they simply move in epicycles.

The solution to the linearised vacuum field equation

$$\Box \bar{h}_{\mu\nu} = \Box\left(h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h\right) = 0 \tag{3.28}$$

subject to the transverse-traceless (TT) gauge

$$\partial^\mu \bar{h}_{\mu\nu} = 0, \quad \bar{h}_{\mu\nu} = h_{\mu\nu}, \quad \bar{h}_{0\nu} = 0 \tag{3.29}$$

for a plane wave travelling along $\mathbf{e}_3$ is given by

$$\bar{h}_{\mu\nu} = C_{\mu\nu} e^{i\mathsf{K}_\rho x^\rho} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & C_+ & C_\times & 0 \\ 0 & C_\times & -C_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} e^{i(\mathsf{K}_3 x^3 - \omega t)} \tag{3.30}$$

where the wavevector K^ρ is null, such that the gravitational wave propagates at the speed of light. This has two associated polarisations C_+ and $C_\times$.

Responses to Gravitational Waves

We have already anticipated how particles respond when a gravitational wave passes by, but we shall now demonstrate this behaviour explicitly. In order to do this, let us first find a linear form of the geodesic equation

$$\dot{\mathsf{U}}^\rho + \Gamma^\rho_{\mu\nu}\mathsf{U}^\mu\mathsf{U}^\nu = 0 \tag{3.31}$$

Note that in the weak field limit in which the particles will move slowly, the coordinate time is roughly the proper time, meaning that we can simply think of the dots as denoting coordinate time derivatives (d/dt).

We have already seen that the linearised form of the affine connection is given by

$$\Gamma^{\rho}_{\mu\nu} = \frac{1}{2}\eta^{\lambda\rho}(\partial_\mu h_{\lambda\nu} + \partial_\nu h_{\lambda\mu} - \partial_\lambda h_{\mu\nu}) \tag{3.32}$$

As we are interested in spatial components, we set $\rho = i$, meaning that the affine connection becomes

$$\Gamma^{i}_{\mu\nu} = \frac{1}{2}\eta^{ij}(\partial_\mu h_{j\nu} + \partial_\nu h_{j\mu} - \partial_j h_{\mu\nu}) \tag{3.33}$$

as $\eta_{\mu\nu}$ is diagonal. As the time components of U^μ are much larger than the spatial ones in the weak field limit, we ignore terms that are quadratic in the velocity $u^i = \mathsf{U}^i$. Setting all the time derivatives of $h_{\mu\nu}$ to zero, we then have that

$$\Gamma^{i}_{\mu\nu}\mathsf{U}^\mu\mathsf{U}^\nu \approx -\frac{1}{2}\mathsf{U}^0\mathsf{U}^0\left(\eta^{ij}\partial_j h_{00}\right) + \mathsf{U}^0\mathsf{U}^k\eta^{ij}\left(\partial_k h_{0j} - \partial_j h_{0k}\right) \tag{3.34}$$

The geodesic equation can then be written as

$$\dot{\mathsf{U}}^i = \mathsf{U}^0\left[\mathsf{U}^0\partial^i h_{00} + \mathsf{U}^k\eta^{ij}\left(\partial_k h_{0j} - \partial_j h_{0k}\right)\right] \tag{3.35}$$

As we are considering only spatial components, the linearised spatial geodesic equation becomes

$$\boxed{\frac{d\mathbf{u}}{dt} = -\frac{1}{2}\left[c^2\nabla h_{00} + (\mathbf{u}c)\times(\nabla\times\mathbf{h}_0)\right]} \tag{3.36}$$

where $\mathbf{h}_0 = (h_{01}, h_{02}, h_{03})$. Now, we know from our solution (3.30) that the perturbing metric satisfies $h_{00} = h_{0i} = 0$. This means that, by the above equation, geodesics will be straight lines in space. We can thus use the geodesic deviation equation to find how the separation between these straight lines changes with the gravitational wave.

Consider (1.103). Once again, the timelike component of U^σ will dominate, meaning that we simply need to compute the Riemann tensor to first order, which gives

$$R^i{}_{j00} = -\frac{1}{2}\partial_0\partial_0 h^i_j \tag{3.37}$$

such that the linearised geodesic equation becomes

$$\boxed{\frac{\partial^2\mathsf{N}^i}{\partial t^2} = \frac{1}{2}\mathsf{N}^j\frac{\partial^2}{\partial t^2}(h^i_j)} \tag{3.38}$$

where again we have taken the proper time to be equal to the coordinate time. As our wave (3.30) is travelling along $\mathbf{e}_3$, only N^1 and N^2 will be affected. For the plus polarisation, (3.38) can be trivially integrated to yield

$$\mathsf{N}^1(t) = \left(1 + \frac{1}{2}C_+e^{i(\mathsf{K}_3x^3-\omega t)}\right)\mathsf{N}^1(0), \quad \mathsf{N}^2(t) = \left(1 - \frac{1}{2}C_+e^{i(\mathsf{K}_3x^3-\omega t)}\right)\mathsf{N}^2(0) \tag{3.39}$$

meaning that particles initially separated along $\mathbf{e}_1$ will oscillate back and forth along $\mathbf{e}_2$, and likewise for those initially separated along $\mathbf{e}_2$, except π out of phase. For cross polarisation, we can the same calculation to yield

$$\mathsf{N}^1(t) = \mathsf{N}^1(0) + \frac{1}{2}C_\times e^{i(\mathsf{K}_3x^3-\omega t)}\mathsf{N}^2(0), \quad \mathsf{N}^2(t) = \mathsf{N}^2(0)\frac{1}{2}C_\times e^{i(\mathsf{K}_3x^3-\omega t)}\mathsf{N}^1(0) \tag{3.40}$$

Figure 3.1 shows some examples of the motion of test points under the influence of a gravitational wave. It is clear that both plus and cross polarisation have a similar effect on the test points. In fact, cross polarisation is simply the same as plus polarisation in a coordinate system that is rotated by $\pi/4$.

(a)

(b)

(c)

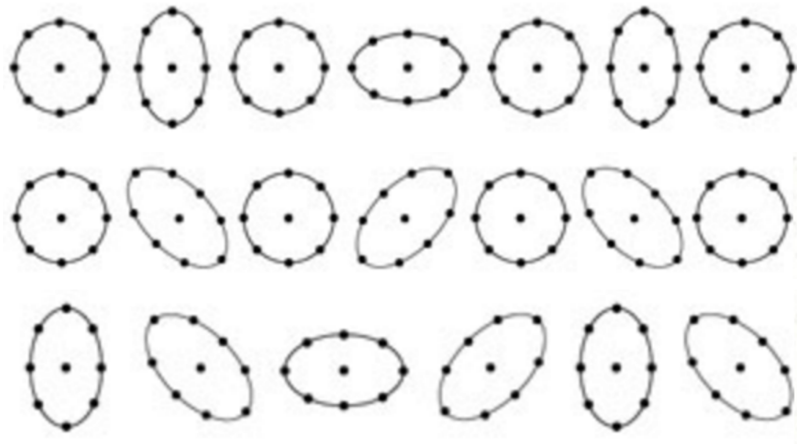

Figure 3.1: Test points subjected to a gravitational wave. (a) Plus polarisation (b) Cross polarisation (c) Right-hand circular polarisation. Note that time is plotted on the horizontal axis

3.2 Gravitational Radiation

Now that we have an understanding of how gravitational waves behave in a vacuum, we can look at how these waves are generated in the first place. For this, we consider the linearised field equation with the source term

$$\Box\bar{h}_{\mu\nu} = -\frac{16\pi G}{c^4}\mathbb{T}_{\mu\nu} \tag{3.41}$$

As usual, the way that we approach equations of this form is to invert the relevant differential operator. We shall begin with this.

3.2.1 Green's Function Solution

The Green's function for the d'Alembertian $\Box = \partial_\mu\partial^\mu$ is defined by

$$\left(-\frac{1}{c^2}\frac{\partial^2}{\partial t^2} + \nabla^2\right) G(\mathbf{x}-\mathbf{x}', t-t') = \delta^3(\mathbf{x}-\mathbf{x}')\delta(t-t') \tag{3.42}$$

Take the time and spatial Fourier transforms of both sides of this equation:

$$\left(\frac{\omega^2}{c^2} - k^2\right) \widetilde{G}(\mathbf{k}-\mathbf{k}', \omega-\omega') = e^{-i\mathbf{k}\cdot\mathbf{x}'}e^{i\omega t'} \quad \longrightarrow \quad \widetilde{G}(\mathbf{k},\omega) = \frac{1}{(\omega/c)^2 - k^2} \tag{3.43}$$

We have chosen to set $\mathbf{k}' = \omega' = 0$ without loss of generality. Transforming back to real space:

$$G(\mathbf{x},t) = \int \frac{d^3\mathbf{k}}{(2\pi)^3}\int \frac{d\omega}{2\pi}\,\frac{1}{(\omega/c)^2 - k^2}e^{i(\mathbf{k}\cdot\mathbf{x}-\omega t)} \tag{3.44}$$

Align our choice of coordinates along $\mathbf{x}$, such that we can write $\mathbf{k}\cdot\mathbf{x} = kr\cos\theta$ for $r = |\mathbf{x}|$. Consider the integral over $\mathbf{k}$ space:

$$\begin{aligned}
\int \frac{d^3\mathbf{k}}{(2\pi)^3}\frac{1}{(\omega/c)^2 - k^2}e^{i\mathbf{k}\cdot\mathbf{x}} &= \frac{1}{(2\pi)^3}\int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin\theta \int_0^\infty dk\, \frac{k^2}{(\omega/c)^2 - k^2}e^{ikr\cos\theta} \\
&= \frac{1}{(2\pi)^2}\int_0^\infty dk\, \frac{k^2}{(\omega/c)^2 - k^2}\int_{-1}^{1} d(\cos\theta)\, e^{ikr\cos\theta} \\
&= \frac{1}{(2\pi)^2}\int_{-\infty}^\infty dk\, \frac{k}{ir}\frac{e^{ikr}}{(\omega/c)^2 - k^2}
\end{aligned} \tag{3.45}$$

Then, our full integral expression becomes

$$G(\mathbf{x},t) = \frac{1}{(2\pi)^3}\frac{1}{ir}\int_{-\infty}^{\infty} dk\, ke^{ikr}\int_{-\infty}^{\infty} d\omega\, \frac{e^{-i\omega t}}{((\omega/c)-k)\,((\omega/c)+k)} \tag{3.46}$$

We now have to invoke some maths from complex analysis. Namely, *Cauchy's Residue Theorem* states that for a complex valued function $f(z)$ with poles at $z = z_n$, the value its integral around some closed contour Γ is given by

$$\oint_\Gamma dz\, f(z) = 2\pi i\sum_n \mathrm{Res}(f(z=z_n)) \tag{3.47}$$

where $\mathrm{Res}(f(z = z_n))$ is a residue contained within the contour Γ. (3.46) has poles at $\omega = \pm c|k|$, along the real axis. Choosing a semicircular contour containing both poles closed in the lower half plane, we can evaluate this integral as

$$G(\mathbf{x},t) = -\frac{2\pi i}{(2\pi)^3}\frac{c^2}{ir}\int_{-\infty}^{\infty} dk\, ke^{ikr}\left[\frac{e^{-ikct}}{2kc}-\frac{e^{ikct}}{2kc}\right] = \frac{c}{4\pi r}\left[\delta(r+ct)-\delta(r-ct)\right] \tag{3.48}$$

Re-introducing $r = |\mathbf{x}|$, $\mathbf{x}'$ and $\mathbf{t}'$, we can write the Greens function for the wave equation as

$$G_\pm(\mathbf{x}-\mathbf{x}', t-t') = -\frac{c}{4\pi\left|\mathbf{x}-\mathbf{x}'\right|}\delta\left(\left|\mathbf{x}-\mathbf{x}'\right| \pm c\left|t-t'\right|\right) \tag{3.49}$$

The positive and negative solutions correspond to $t < t'$ and $t > t'$ respectively. This means we can re-construct our final solution

$$\bar{h}_{\mu\nu} = \int d^3\mathbf{x}'\int dt' G_\pm\left(-\frac{16\pi G}{c^4}\mathbb{T}_{\mu\nu}\right) = \frac{4G}{c^4}\int d^3\mathbf{x}'\,\frac{1}{\left|\mathbf{x}-\mathbf{x}'\right|}\mathbb{T}_{\mu\nu}\left(ct-\left|\mathbf{x}-\mathbf{x}'\right|, \mathbf{x}'\right) \tag{3.50}$$

where we have chosen the negative solution as this corresponds to propagation away from the source. We note that the stress energy tensor is a function of both the position $\mathbf{x}'$ of the source of the gravitational waves, and the corresponding retarded source time $t' = t - |\mathbf{x}-\mathbf{x}'|$.

3.2.2 The Quadrupole Formula

(3.50) is still difficult to solve in full generality. Instead, we consider the case where the observer at $\mathbf{x}$ is very far away from the source, such that $|\mathbf{x}| \gg |\mathbf{x}'|$. This is known as the *compact source approximation*. We can expand the denominator as

$$\frac{1}{|\mathbf{x}-\mathbf{x}'|} = \frac{1}{|\mathbf{x}|}\left(1+\frac{|\mathbf{x}'|^2}{|\mathbf{x}|^2}-2\frac{\mathbf{x}\cdot\mathbf{x}'}{|\mathbf{x}|^2}\right)^{-1/2} \approx \frac{1}{|\mathbf{x}|}\left(1+\frac{\mathbf{x}\cdot\mathbf{x}'}{|\mathbf{x}|}+\dots\right) \tag{3.51}$$

Letting $r = |\mathbf{x}|$, we have we have to first order that

$$\bar{h}_{\mu\nu} = \frac{4G}{c^4}\frac{1}{r}\int d^3\mathbf{x}'\,\mathbb{T}_{\mu\nu} \tag{3.52}$$

We now adopt the TT gauge, meaning that we are only interested in the spatial parts of $\bar{h}_{ij}$, since all time indices vanish. The above integral can be simplified by considering

$$\left(\mathbb{T}^{\mu\nu}x^ix^j\right)_{,\mu\nu} = \left(\mathbb{T}^{00}x^ix^j\right)_{,00} + \left(\mathbb{T}^{ij}x^ix^j\right)_{,ij} \tag{3.53}$$

The left-hand side of this equation can be re-written as

$$\left(\mathbb{T}^{\mu\nu}x^ix^j\right)_{\mu\nu} = \partial_\mu\partial_\nu\left(x^ix^j\mathbb{T}^{\mu\nu}\right) = 2\delta_\mu{}^i\delta_\nu{}^j\mathbb{T}^{\mu\nu} = 2\mathbb{T}^{ij} \tag{3.54}$$

such that

$$\mathbb{T}^{ij} = \frac{1}{2}\left[\left(\mathbb{T}^{00}x^i x^j\right)_{,00} + \left(\mathbb{T}^{ij}x^i x^j\right)_{,ij}\right] \tag{3.55}$$

Then, our integral expression can be written as

$$\bar{h}_{ij} = \frac{4G}{c^4}\frac{1}{r}\int d^3\mathbf{x}'\,\mathbb{T}_{ij} = \frac{2G}{c^4}\frac{1}{r}\int d^3\mathbf{x}'\,\left(\mathbb{T}^{00}x^i x^j\right)_{,00} + \left(\mathbb{T}^{ij}x^i x^j\right)_{,ij} \tag{3.56}$$

As we are in the weak-field limit, covariant derivatives become partial derivatives (to first order), while proper time derivatives simply become normal time derivatives. This means that the integral

$$\int d^3\mathbf{x}'\,\left(\mathbb{T}^{ij}x^i x^j\right)_{,ij} \stackrel{!}{=} 0 \tag{3.57}$$

must vanish over all space by energy conservation. This allows us to arrive at the *quadrupole formula*:

$$\boxed{\bar{h}_{ij}(t,\mathbf{r}) = \frac{2G}{c^4}\frac{\ddot{I}_{ij}}{r}, \quad I_{ij} = \frac{1}{c^2}\int d^3\mathbf{r}_s\, r_s^i r_s^j\,\mathbb{T}^{00}(t_s,\mathbf{r}_s)} \tag{3.58}$$

Once again, the field event (at which the gravitational waves are measured) is at $(ct, \mathbf{r})$, while the source event (at which the gravitational waves are emitted) is at $(ct_s, \mathbf{r}_s)$, where $t_s = t - r/c$ is the retarded source time. This is a very neat result; given a particular (time-dependent) density distribution $\mathbb{T}^{00} = \rho c^2$, we can find the local perturbation from the Minkowski metric, with the perturbation depending on the mass quadrupole moment. Note that the time derivatives are with respect to the source time t_s.

One can also show - with an extensive amount of algebra - that the total gravitational luminosity of the source is given by the beautifully simple formula

$$\boxed{L_{GW} = \frac{G}{5c^5}\dddot{J}_{ij}\dddot{J}_{ij}, \quad J_{ij} = I_{ij} - \frac{1}{3}\delta_{ij}\delta^{mn}I_{mn}} \tag{3.59}$$

We will make use of both this and the quadrupole formula in the next section.

3.2.3 Compact Binary System

The gravitational radiation observation event at LIGO was of a wave emitted by such a system consisting of two massive black holes in a binary orbit. Due to their mutual attraction, these black holes spiralled in towards each other - at an ever increasing rate - until they collided, releasing an unimaginable amount of energy in the form of gravitational radiation. We can gain an insight into this significant observation by examining the compact binary system.

Let us describe the orbital mechanics of the two massive objects classically. Evidently, this is only an approximation, but it works quite well. We have already derived the main results associated with classical orbits in section 2.2.1, but we shall state the important ones here:

$$r = \frac{a(1-e^2)}{1+e\cos\phi}, \quad \frac{d\phi}{dt} = \frac{[GMa(1-e^2)]^{1/2}}{r^2} \tag{3.60}$$

Note that e is the orbital eccentricity ($e = 0$ for circular orbits, $e = 1$ for scattering orbits), a is the semi-major axis of the orbit, and $M = m_1 + m_2$ where m_1 and m_2 are the masses of the two bodies that make up the binary. Assume that both bodies have motion only in the $\theta = \pi/2$ plane, such that the (time-dependant) mass density can be written as

$$\rho = \delta(z)\left[m_1\delta(x - r_1\cos\phi)\delta(y - r_1\sin\phi) + m_2\delta(x + r_2\cos\phi)\delta(y + r_2\sin\phi)\right] \tag{3.61}$$

where

$$r_1 = \frac{m_2}{M} r, \quad r_2 = \frac{m_1}{M} r, \quad \mu = \frac{m_1 m_2}{m_1 + m_2} \tag{3.62}$$

We can now calculate the non-vanishing moments of the quadrupole tensor. Consider I_{xx}:

$$I_{xx} = \int d^3\mathbf{x}\, \rho x^2 = m_1 r_1^2 \cos^2\phi + m_2 r_2^2 \cos^2\phi = \mu r^2 \cos^2\phi \tag{3.63}$$

Similarly, we have that

$$I_{yy} = \mu r^2 \sin^2\phi \tag{3.64}$$

$$I_{xy} = I_{yx} = \mu r^2 \sin\phi\cos\phi \tag{3.65}$$

$$I_{ii} = I_{xx} + I_{yy} = \mu r^2 \tag{3.66}$$

Now, this is far as we can proceed in full generality while keeping the solution simple. As such, we shall treat the simple case of the circular binary, and then sketch an outline of the general result involving eccentric orbits.

Circular Orbits

For circular orbits, $e = 0$ and the radius remains fixed at $r = a$, meaning that the orbital frequency becomes

$$\omega = \frac{d\phi}{dt} = \left(\frac{GM}{a^3}\right)^{1/2} \tag{3.67}$$

Once again, this is just an approximation; the radius must change as the system looses energy by emitting gravitational radiation, meaning that the frequency will also change. However, we will stick with it for the sake of this example.

It is straightforward to calculate the trace-reversed quadrupole tensor from the quadrupole moments listed above:

$$J_{ij} = I_{ij} - \frac{1}{3}\delta_{ij} I_{kk} = \frac{1}{2}\mu a^2 \begin{pmatrix} \cos 2\omega t + 1/3 & \sin 2\omega t & 0 \\ \sin 2\omega t & -\cos 2\omega t + 1/3 & 0 \\ 0 & 0 & -2/3 \end{pmatrix} \tag{3.68}$$

Taking time derivatives of this expression is straightforward, given that only ϕ is time dependent. Taking three time derivatives, this becomes

$$\dddot{J}_{ij} = 4\mu a^2 \omega^3 \begin{pmatrix} \sin 2\omega t & -\cos 2\omega t & 0 \\ -\cos 2\omega t & -\sin 2\omega t & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{3.69}$$

Then:

$$\dddot{J}_{ij}\dddot{J}_{ij} = 16\mu^2 a^4 \omega^6 \,\mathrm{Tr} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} = 32\mu^2 a^4 \omega^6 \tag{3.70}$$

Then, the gravitational luminosity of the binary is given by (3.59):

$$L_{GW} = \frac{32}{5}\frac{G}{c^5}\mu^2 a^4 \omega^6 = \frac{32}{5}\frac{G^4}{c^5}\frac{m_1^2 m_2^2 (m_1 + m_2)}{a^5} \tag{3.71}$$

By the virial theorem, the total energy of the binary is

$$E_{\text{tot}} = \frac{1}{2}\langle V \rangle = \frac{1}{2}\frac{G m_1 m_2}{r} \tag{3.72}$$

The energy lost as gravitational radiation can only come from the total energy of the binary. We can use this to provide a very rough estimate for the time for the two bodies to collide if they are initially at some separation r_0. Equating the gravitational luminosity with the rate of change of the energy:

$$\frac{dE_{\text{tot}}}{dt} = \frac{1}{2}Gm_1m_2\frac{d}{dt}\left(\frac{1}{r}\right) = L_{GW} \quad \longrightarrow \quad \frac{dr}{dt} = -\frac{64}{5}\frac{G^3}{c^5}\frac{m_1m_2(m_1+m_2)}{r^3} \tag{3.73}$$

This differential equation can be solved subject to the initial condition to give

$$r^4 = r_0^4 - \frac{256}{5}\frac{G^3}{c^5}m_1m_2(m_1+m_2)t \tag{3.74}$$

The time at which the binaries merge is thus given my

$$t_{\text{merge}} = \frac{5}{256}\frac{c^5}{G^3}\frac{r_0^4}{m_1m_2(m_1+m_2)} \tag{3.75}$$

This scales as we would expect; the greater the initial separation, the larger time till the merger, but the greater the mass, the shorter the time as more energy is lost as gravitational radiation.

Eccentric Orbits

The case of eccentric orbits is significantly more complicated. The full treatment of the problem is covered in 'Peters and Mathews 1963', though we shall sketch a derivation of the key results here.

The components of the quadrupole tensor (3.63), (3.65) and (3.64) remain the starting point in this case, except that when taking derivatives, one needs to take into account that both r and ϕ have time dependence. Doing this, we obtain the components

$$\frac{d^3I_{xx}}{dt^3} = \beta(1+e\cos\phi)^2\left[2\sin 2\phi + 3e\sin\phi\cos^2\phi\right] \tag{3.76}$$

$$\frac{d^3I_{yy}}{dt^3} = -\beta(1+e\cos\phi)^2\left[2\sin 2\phi + e\sin\phi(1+3\cos^2\phi)\right] \tag{3.77}$$

$$\frac{d^3I_{xy}}{dt^3} = \frac{d^3I_{yx}}{dt^3} = -\beta(1+e\cos\phi)^2\left[2\sin 2\phi - e\cos\phi(1-3\cos^2\phi)\right] \tag{3.78}$$

where

$$\beta^2 = 4G^3\frac{m_1^2m_2^2(m_1+m_2)}{a^5(1-e^2)^5} \tag{3.79}$$

Using (3.59), we find that

$$L_{GW} = \frac{32}{5}\frac{G^4}{c^5}\frac{m_1^2m_2^2(m_1+m_2)}{a^5(1-e^2)^5}(1+e\cos\phi)^4\left[(1+e\cos\phi)^2 + \frac{e^2}{12}\sin^2\phi\right] \tag{3.80}$$

However, we now need to average L_{GW} over an orbit. However, this is not as simple as simply integrating over $d\phi/2\pi$, as this assumes that ϕ remains constant over the period. Instead we must integrate over time $d\phi/\dot{\phi}$, and then divide by the total period to give the time average. Skipping over the painstaking detail, the result is

$$\boxed{\langle L_{GW}\rangle = \frac{32}{5}\frac{G^4}{c^5}\frac{m_1^2m_2^2(m_1+m_2)}{a^5(1-e^2)^{7/2}}\left(1+\frac{73}{24}e^2+\frac{37}{96}e^4\right)} \tag{3.81}$$

It is clear that this reduces to (3.71) in the limit that $e \to 0$. The advantage of this result is that it is applicable to all types of orbits, even very eccentric ones. For example, we can consider a single parabolic encounter between two massive bodies. Such a scattering event corresponds to $e \to 1$ in such a way that $\ell = a(1-e^2) = a(1-e)(1+e)$ remains finite. The distance of closest approach is given by $b = a(1-e)$, while the period of the orbit is given by

$$T = 2\pi\sqrt{\frac{a^2}{GM}} \tag{3.82}$$

The total gravitational energy emitted in the encounter is given by

$$\Delta E_{GM} = \lim_{e\to 1} T \langle L_{GW} \rangle \tag{3.83}$$

which is give explicitly by

$$\Delta E_{GW} = \frac{84\sqrt{2}\pi}{24}\frac{G^{7/2}m_1^2 m_2^2(m_1+m_2)}{c^5 b^{7/2}} \tag{3.84}$$

Limits on Gravitational Radiation

According to Hawking (1971), when two black holes of masses m_1 and m_2 collide to form a single large black hole of mass M, the total area of the event horizon must increase. This is due to causality arguments. We know from our work with null radial geodesics in section 2.3.2 that points inside Schwarzchild radii of the original two black holes cannot lie outside of the Schwarzchild radius of the final black hole, as otherwise this would correspond to the propagation of something outwards across the event horizon. One can show that this places the constraint on the merger that the total area of the event horizon must increase.

We know that the invariant volume is given by

$$d^4V = \sqrt{-g}\, d^4x = \sqrt{-g}\, dtdrd\theta d\phi \tag{3.85}$$

Let us adopt the Schwarzchild metric (2.18), such that $g = -c^2 r^4 \sin^2\theta$. Then, the surface area element in our space is given by

$$dS = \bar{\partial}_t \bar{\partial}_r \sqrt{-g}\, d^4x \tag{3.86}$$

where

$$\bar{\partial}_t = \frac{1}{c(1-r_s/r)^{1/2}}\partial_t, \quad \bar{\partial}_r = (1-r_s/r)^{1/2}\partial_r \tag{3.87}$$

are the normalised tangent vectors in Schwarzchild geometry. This evaluates to

$$dS = r^2 \sin\theta d\theta d\phi \tag{3.88}$$

We could have immediately deduced this from the fact that the metric is spherically symmetric, but we have shown this method here as it is easily generalised. From this, it is clear that the area of the event horizon is given by

$$A = 4\pi r_s^2 \tag{3.89}$$

We must have that the final area is greater than the initial area, giving us the following constraint on the masses

$$M \geq \left(m_1^2 + m_2^2\right)^{1/2} \tag{3.90}$$

It follows that the upper limit on the energy that can be emitted as gravitational radiation in such a black hole merger is

$$\Delta E_{GW}/c^2 \leq m_1 + m_2 - \left(m_1^2 + m_2^2\right)^{1/2} \tag{3.91}$$

4. *Cosmology*

This chapter aims to give the reader a basic outline of cosmology in General Relativity, including:

- Homogeneity, Isotropy and Curved Spaces
- The Friedmann Equations
- Large Scale Dynamics
- Thermal History of the Universe

We will now pull back out focus from small perturbative distortions in local spacetime to consider the evolutionary dynamics of the universe as a whole. Cosmology is a subject that can only really be treated properly with some understanding of General Relativity, as it involves the way in which matter influences the curvature of the universe as a whole.

4.1 Homogeneity, Isotropy and Curved Spaces

Much of the study of cosmology is based on the *cosmological principle*; this assumes that the universe is

- Homogeneous - That, at any given time, the universe looks exactly the same at every point in space. This is evidently only valid when looking on large scales, but that is what we are concerned with in cosmology
- Isotropic - That the universe looks the same in any direction of observation. This is confirmed by the fact that the observed universe appears very regular, and the same regardless in what direction we look at it

Note that homogeneity and isotropy are related concepts, but they are indeed distinct. For example, a universe with is isotropic will be homogeneous, while a universe that is homogeneous may not be isotropic. A universe which is only isotropic around one point is not homogeneous. We believe that our universe is both.

These assumptions of the cosmological principle has important physical consequences. For instance, the metric cannot be a function of space - as otherwise this would break both homogeneity and isotropy - and can thus only depend on time. One possible solution is that the universe is an infinite, static, globally flat sheet, meaning that the Minkowski metric $\eta_{\mu\nu}$ holds everywhere. However, there is a simple way of showing that this cannot be the case. Consider the energy density of n_0 galaxies per unit volume of luminosity L:

$$du = \frac{dF}{c} = n_0 \times \frac{L}{4\pi r^2} \times 4\pi r^2 \, dr = \frac{L n_0 dr}{c} \tag{4.1}$$

Evidently, integrating this over all space will result in a divergent integral; if the universe was indeed globally flat and static, we would be blinded by the fire of trillions of suns. This is known as *Olber's Paradox*, which disqualifies the static Minkowski model.

4.1.1 Curved Spaces

The assumption of homogeneity and isotropy means that we are restricted to maximally symmetric spaces; these have the same number of symmetries as ordinary Euclidean space. More formally, they are the spaces for which there exist $n(n+1)/2$ independent Killing vectors, where n is the dimension of the space. It can be shown that the only possible spaces that we can consider are those with constant curvature; namely, flat euclidean space (though with time dependence), a positively curved space, and a negatively curved space.

Consider a four-dimensional space (w, x, y, z). The surface of a positively and negatively curved four-sphere is given by

$$\pm w^2 + x^2 + y^2 + z^2 = \pm R^2 = \text{constant} \tag{4.2}$$

From this, we can construct a three-space hypersurface by recognising that on this surface, a small change in w^2 is restricted to satisfy

$$\pm d(w^2) = -d(x^2 + y^2 + z^2) \equiv -d(r^2) \quad \longrightarrow \quad \pm dw = -\frac{r dr}{w} \tag{4.3}$$

Hence, we have that

$$(dw)^2 = \frac{r^2 (dr)^2}{w^2} = \frac{r^2 (dr)^2}{R^2 \mp r^2} \tag{4.4}$$

The line element in the four-space is then

$$ds^2 = dw^2 + dx^2 + dy^2 + dz^2 = dw^2 + dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \tag{4.5}$$

with $(dw)^2$ as given above. It follows immediately that the line element of the three-surface of a four-sphere is given by

$$ds^2 = \frac{dr^2}{1 \mp r^2/R^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2 \tag{4.6}$$

Our metric for the space is then given by $-c^2 d\tau^2 = -c^2 dt^2 + a(t)^2 ds^2$, such that

$$\boxed{-c^2 d\tau^2 = -c^2 dt^2 + a(t)^2 \left[\frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2\right]} \tag{4.7}$$

where we have defined $k = -1/R^2$. This is known as the *Friedmann-Robertson-Walker (FRW) metric*, which we shall make extensive use of throughout this chapter. $k = 0$ corresponds to a flat space, while $k = 1$ and $k = -1$ correspond to a positively and negatively curved space respectively.

4.1.2 Scale Factor

The coefficient $a(t)$ is known as the *scale factor*, which essentially determines how 'big' the three-surface is within our four-dimensional space. In cosmology, this is the central parameter that we will be working with.

Consider two distant objects that lie a given distances from one another. At a time t_1, they are at a distance r_1, while at a time t_2, they are at a distance r_2. During that time interval, the change between r_1 and r_2 is clearly given by

$$\frac{r_2}{r_1} = \frac{a(t_2)}{a(t_1)} \tag{4.8}$$

and, because of the cosmological principle, this holds true for any set of chosen points. It then makes sense to parametrise the distance between two points as

$$r(t) = \ell\, a(t) \tag{4.9}$$

where ℓ is completely independent of time. ℓ corresponds to the conformal coordinates (x, y, z) that remain unchanged during the evolution of the universe, while the physical coordinates are $a(t)(x, y, z)$ that do scale with this evolution. Let us calculate how quickly these two objects separate. From (4.9), their relative velocity is given by

$$\dot{r}(t) = \ell \dot{a}(t) = \frac{\dot{a}}{a} r(t) \equiv H(t) r(t) \tag{4.10}$$

We have thus derived *Hubble's law*:

$$\boxed{\dot{r} = H(t) r, \quad H(t) = \frac{\dot{a}(t)}{a(t)}} \tag{4.11}$$

$H(t)$ is known as the Hubble parameter, or - when evaluated at the current time $t = t_0$ - it is known as the Hubble constant $H_0 \approx 70\,\mathrm{kms^{-1}Mpc^{-1}}$.

Cosmological Redshift

We have already encountered gravitational redshift in the case of the Schwarzchild metric. Let us now consider redshift in our FRW geometry. Suppose that a cosmologically distance source emits to photons, the first one at a time t_e and the second one a short time later, at a time $t_e + \Delta t_e$. Now, we can orient our coordinate system such that these photons travel radially ($d\theta = d\phi = 0$). As photons travel along null geodesics, $d\tau = 0$, meaning that

$$cdt = \pm \frac{a(t)dr}{\sqrt{1-kr^2}} \tag{4.12}$$

where the positive and negative sign correspond to motion away from and towards the origin respectively. Usually, the observer is placed at the origin, such that one adopts the negative solution. The *co-moving distance* (associated with the conformal coordinates, and so-called as it moves with the expansion) to the observer is given by the integral

$$\ell_o = c\int_{t_e}^{t_o} \frac{dt}{a(t)} = c\int_{t_e+\Delta t_e}^{t_o+\Delta t_o} \frac{dt}{a(t)} \tag{4.13}$$

Note that ℓ_o does not depend on time, because we have taken out the explicit time dependence introduced by the scale factor $a(t)$, meaning that we can make the third step above. Then, it follows that

$$\int_{t_e+\Delta t_e}^{t_o+\Delta t_o} = \int_{t_e}^{t_o} + \int_{t_o}^{t_o+\Delta t_o} - \int_{t_e}^{t_e+\Delta t_e} \tag{4.14}$$

allowing us to write

$$\int_{t_o}^{t_o+\Delta t_o} \frac{dt}{a(t)} = \int_{t_e}^{t_e+\Delta t_e} \frac{dt}{a(t)} \tag{4.15}$$

For small time intervals, this becomes

$$\frac{\Delta t_o}{a(t_o)} = \frac{\Delta t_e}{a(t_e)} \tag{4.16}$$

Let us take Δt_e as the period of the photon, meaning that the wavelength is $\Delta t_e/c$. This means that we can write that

$$\boxed{\lambda_o = \lambda_e(1+z), \quad 1+z = \frac{a(t_o)}{a(t_e)}} \tag{4.17}$$

where we have defined the redshift parameter z. Note that this has nothing to do with a Doppler shift; the only thing which is moving (in conformal coordinates) is the photon, and that cannot be Doppler shifted.

Resolving Olber's Paradox

We are now in a position to resolve Olber's paradox outlined at the start of this section. Adopt the FRW metric for a flat universe ($k = 0$), such that

$$-c^2d\tau^2 = -c^2dt^2 + a^2dr^2 + a^2r^2d\Omega^2 \tag{4.18}$$

where we have defined the solid angle element $d\Omega^2 = d\theta^2 + \sin^2\theta\phi^2$. We want to calculate volumes and surfaces within this space, and so we shall use the same technique as that used at the end of the last chapter. The invariant volume is

$$d^4V = \sqrt{-g}\, dtdrd\theta d\phi = r^2a^2\sin\theta\, dtdrd\theta d\phi \tag{4.19}$$

The volume and surface area element in this space is then given by

$$d^3V = \bar{\partial}_t d^4V = r^2 a^2 \sin\theta \, dr d\theta d\phi \tag{4.20}$$

$$dS = \bar{\partial}_t \bar{\partial}_r d^4V = r^2 a \sin\theta \, d\theta d\phi \tag{4.21}$$

where

$$\bar{\partial}_t = \partial_t, \quad \bar{\partial}_r = a^{-1}\partial_r \tag{4.22}$$

are once again the normalised tangent vectors. Then, the energy density of n_0 galaxies per unit volume of luminosity L at the current time $t = t_0$ is given by

$$du = \frac{1}{c}\left(n_0 \frac{L}{dS} d^3V\right) = \frac{1}{c} n_0 La \, dr \tag{4.23}$$

However, this is the energy density when the radiation was emitted. The energy density observed now can thus be calculated from

$$a(t_0) du_0 = du_0 = a(t) du \tag{4.24}$$

where we have set $a(t_0) = 1$. Recalling (4.12), it is clear that we write

$$du_0 = \frac{1}{c} n_0 La^2 \, dr = \frac{1}{c} n_0 La^2 \left(c\frac{dt}{a}\right) = n_0 La \, dt \tag{4.25}$$

meaning that the full energy density is

$$u = n_0 L \int_0^{t_0} dt \, a(t) \tag{4.26}$$

This integral is clearly finite, meaning that adopting a scale factor $a(t)$ allows us to resolve Olber's paradox.

4.2 The Friedmann Equations

Now that we have a metric to describe the universe, we can in principle solve the field equations (1.118), which we shall now do. Unlike with the Schwarzchild solution, we shall retain the cosmological constant Λ, as this is important on cosmological scales.

4.2.1 The Friedmann Solution

Recall that the components of the FRW metric are

$$g_{00} = -1, \quad g_{rr} = \frac{a^2}{1-kr^2}, \quad g_{\theta\theta} = a^2r^2, \quad g_{\phi\phi} = a^2r^2\sin^2\theta \tag{4.27}$$

Adopting the notation $\dot{a} = da/dt$, the non-zero components of the affine connection are

$$\begin{aligned}
\Gamma^0_{rr} &= -\frac{1}{2}g^{00}\partial_0 g_{rr} = \frac{a\dot{a}}{1-kr^2} \\
\Gamma^0_{\theta\theta} &= -\frac{1}{2}g^{00}\partial_0 g_{\theta\theta} = a\dot{a}r^2 \\
\Gamma^0_{\phi\phi} &= -\frac{1}{2}g^{00}\partial_0 g_{\phi\phi} = a\dot{a}r^2\sin^2\theta \\
\\
\Gamma^r_{\theta\theta} &= -\frac{1}{2}g^{rr}\partial_r g_{\theta\theta} = -r(1-kr^2) \\
\Gamma^r_{\phi\phi} &= -\frac{1}{2}g^{rr}\partial_r g_{\phi\phi} = -r(1-kr^2)\sin^2\theta \\
\\
\Gamma^\theta_{\phi\phi} &= -\frac{1}{2}g^{\theta\theta}\partial_\theta g_{\phi\phi} = -\sin\theta\cos\theta \\
\Gamma^\phi_{\theta\phi} &= \frac{1}{2}g^{\phi\phi}\partial_\theta g_{\phi\phi} = \cot\theta \\
\\
\Gamma^r_{0r} &= \Gamma^\theta_{0\theta} = \Gamma^\phi_{0\phi} = \frac{\dot{a}}{a} \\
\Gamma^\theta_{r\theta} &= \Gamma^\phi_{r\phi} = \frac{1}{r}
\end{aligned}$$

Then, using (2.7), the components of the Ricci tensor are

$$R_{00} = -3\frac{\ddot{a}}{a}, \quad R_{ij} = (\ddot{a}a + 2\dot{a}^2 + 2k)g_{ij} \tag{4.28}$$

with the Ricci scalar being

$$R = \frac{6}{a^2}(\ddot{a}a + \dot{a}^2 + k) \tag{4.29}$$

We note that the curvature along $t =$ constant slices in the space is $R = 6k/a^2$, which is consistent with either flat, positively or negatively curved space for $k = 0$, $k = 1$, and $k = -1$ respectively.

Fluid Conservation

The universe is evidently not empty (or else how would this author be sitting here writing this?), so we are not interested in vacuum solutions to the fields equations. The assumptions that we have made about isotropy and homogeneity tells us that the universe must be filled with a perfect fluid with no overall velocity, meaning that the source term is given by

$$\mathbb{T}_{\mu\nu} = p g_{\mu\nu} + \left(\rho + \frac{p}{c^2}\right) \mathsf{U}_\mu \mathsf{U}_\nu, \quad \mathsf{U}_\mu = (c, \mathbf{0}) \tag{4.30}$$

Note that we can also write the stress energy tensor as

$$\mathbb{T}_{\mu\nu} = \left(\begin{array}{c|ccc} \rho c^2 & 0 & 0 & 0 \\ \hline 0 & & & \\ 0 & & g_{ij}p & \\ 0 & & & \end{array}\right), \quad \mathbb{T}^\mu{}_\nu = \mathrm{diag}(-\rho c^2, p, p, p) \tag{4.31}$$

Before substituting everything into the field equations, it is useful to consider the zeroth component of the conservation equation:

$$0 = \nabla_\mu \mathbb{T}^\mu{}_0 = \partial_\mu \mathbb{T}^\mu{}_0 + \Gamma^\mu_{\mu 0}\mathbb{T}^0{}_0 - \Gamma^\lambda_{\mu 0}\mathbb{T}^\mu{}_\lambda = -\partial_0 \rho c^2 - 3\frac{\dot{a}}{a}(\rho c^2 + p) \tag{4.32}$$

We prose that the equation of state for polytropic fluids, namely that

$$p = w\rho c^2 \tag{4.33}$$

Then, the conservation of energy equation becomes

$$\frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{a}}{a} \quad \longrightarrow \quad \rho \propto a^{-3(1+w)} \tag{4.34}$$

We have three different 'cosmological fluids' to consider:

- Dust ($w = 0$, $\rho \propto a^{-3}$) -Dust is collisionless, non-relativistic matter for which $w = 0$. Examples include ordinary stars and galaxies, for which pressure is negligible in comparison with energy density. Dust is often referred to as 'matter', and universes whose energy density is mostly due to dust are known as *matter-dominated.* The density of matter falls off as

  $$\rho \propto a^{-3} \tag{4.35}$$

  which can simply be interpreted as the decrease in the number density of the particles as the universe expands according to a

- Radiation ($w = 1/3$, $\rho \propto a^{-4}$) - Radiation may be used to describe either actual electromagnetic radiation, or massive particles moving at relative velocities sufficiently close to c that they become indistinguishable from photons, at least as far as their equation of state is concerned. A universe in which most of the density is in the form of radiation is known as *radiation-dominated.* As $w = 1/3$, radiation density falls off as

  $$\rho \propto a^{-4} \tag{4.36}$$

  The energy density of radiation thus falls off more quickly than matter. This is because the number density of photons decreases in the same way as the number density of non-relativistic particles, but individual photons also lose energy as a^{-1} due to the cosmological redshift.

- Vacuum Energy ($w = -1$, $\rho \not\propto a$) -Let us consider the right-hand side of (1.118). From (4.30) above, we can write that

  $$\frac{8\pi G}{c^4}\mathbb{T}_{\mu\nu} - \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}\left[\bar{p} g_{\mu\nu} + \left(\bar{\rho} + \frac{\bar{p}}{c^2}\right)\mathsf{U}_\mu \mathsf{U}_\nu\right] \tag{4.37}$$

where we have defined

$$\bar{p} = p - \frac{c^4\Lambda}{8\pi G}, \bar{\rho} = \rho + \frac{c^2\Lambda}{8\pi G} \tag{4.38}$$

The effect of the cosmological constant is thus to increase the spatial pressure, but increase the energy density; this corresponds to the energy density of the vacuum. This vacuum energy has $w = -1$, meaning its density is independent of a, which is what we would expect. Since the energy density in matter and radiation decreases as the universe expands, if there is non-zero vacuum energy, it tends to win out in the long term. If this is the case, we say that the universe becomes *vacuum-dominated.*

Equations of the Universe

Now that we have established that the cosmological constant leads to an effective density, we shall no-longer worry about it as a separate term, and instead include it within our density ρ, which now includes that due to dust, radiation, and vacuum energy. As such, the field equations can be written in the form

$$R_{\mu\nu} = \frac{8\pi G}{c^4}\left(\mathbb{T}_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\mathbb{T}^{\mu}{}_{\mu}\right) \tag{4.39}$$

The $\mu\nu = 00$ equation is

$$-3\frac{\ddot{a}}{a} = 4\pi G\left(\rho + \frac{3p}{c^2}\right) \tag{4.40}$$

while, due to isotropy, all components $\mu\nu = ij$ give

$$\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 + 2\frac{k}{a^2} = 4\pi G\left(\rho - \frac{p}{c^2}\right) \tag{4.41}$$

We can use (4.40) to get rid of the second derivatives in (4.41), such that we arrive at

$$\boxed{\begin{aligned} H^2 = \left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2} + \frac{\Lambda c^2}{3} && (4.42)\\ \dot{H} + H^2 = \frac{\ddot{a}}{a} &= -\frac{4\pi G}{3}\left(\rho + \frac{3p}{c^2}\right) + \frac{\Lambda c^2}{3} && (4.43)\end{aligned}}$$

where we have included the cosmological constant Λ explicitly. These are the *Friedmann equations* that govern the expansion of space in homogeneous and isotropic models of the universe; that is, under the FRW metric.

4.2.2 The Critical Density

A useful parameter to define is the *critical density*

$$\boxed{\rho_c = \frac{3H^2}{8\pi G}} \tag{4.44}$$

which corresponds to the density at which there is no curvature ($k = 0$). We can also define the *density ratio*

$$\Omega = \frac{\rho}{\rho_c} \tag{4.45}$$

such that the first Friedmann equation (4.42) can be written as

$$\Omega - 1 = \frac{k}{H^2a^2} \tag{4.46}$$

The sign of k, and thus the curvature, is determined by whether Ω is greater than, equal to, or less than one. We have that

$$\begin{array}{ccccccc} \rho < \rho_c & \leftrightarrow & \Omega < 1 & \leftrightarrow & k = -1 & \leftrightarrow & \text{open} \\ \rho = \rho_c & \leftrightarrow & \Omega = 1 & \leftrightarrow & k = 0 & \leftrightarrow & \text{flat} \\ \rho > \rho_c & \leftrightarrow & \Omega > 1 & \leftrightarrow & k = +1 & \leftrightarrow & \text{closed} \end{array}$$

Included within our total density ρ, we have matter (ρ_m), radiation (ρ_γ) and vacuum energy (ρ_Λ). This means that we can define the density ratios

$$\Omega_m = \frac{\rho_m}{\rho_c}, \quad \Omega_\Lambda = \frac{\rho_\Lambda}{\rho_c}, \quad \Omega_\gamma = \frac{\rho_\gamma}{\rho_c}, \quad \Omega_k = -\frac{k}{a^2 H^2} \tag{4.47}$$

Now, let us assume that ρ_c and all our densities are evaluated at the current time $t = t_0$, such that

$$\rho_m = \rho_m^0 \left(\frac{a_0}{a}\right)^3 \tag{4.48}$$

$$\rho_\gamma = \rho_\gamma^0 \left(\frac{a_0}{a}\right)^4 \tag{4.49}$$

$$\rho_\Lambda = \rho_\Lambda^0 \tag{4.50}$$

as previously discussed. Remarking that

$$\frac{8\pi G}{3} = \frac{H_0^2}{\rho_c}, \quad -\frac{k}{a_0^2} = H_0^2 \Omega_k \tag{4.51}$$

we can write the first Friedmann equation (4.42) as

$$\boxed{\left(\frac{H}{H_0}\right)^2 = \left[\Omega_\gamma \left(\frac{a_0}{a}\right)^4 + \Omega_m \left(\frac{a_0}{a}\right)^3 + \Omega_k \left(\frac{a_0}{a}\right)^2 + \Omega_\Lambda\right]} \tag{4.52}$$

where all the density ratios are evaluated at the current time. This is usually the form of the first Friedmann equation that is used because it essentially contains all the information that we need. We can measure H_0, and all the density ratios at the current time, from which we can find a solution for a. We shall do this in section (4.3).

4.2.3 Distances in Cosmology

In an expanding universe, it is not trivial to define distances. We shall define several different (useful) kinds of distance here that will enter into our later discussions.

Comoving Distance

We have already met the idea of comoving distance; it is the time-independent distance $\ell \equiv D_c$ that is associated with the conformal coordinates, given by:

$$\boxed{D_C = c\int_t^{t_0} \frac{dt}{a(t)}} \tag{4.53}$$

For values of $a \approx 1$, or small values of z, we can expand the comoving distance to second order in z as:

$$D_C \approx \frac{c}{H_0}\left[z - \frac{z^2}{2}(1 + q_0) + \dots\right] \tag{4.54}$$

The quantity $D_H = c/H_0$ is sometimes referred to as the *Hubble distance*.

Proper Distance

The proper distance is the distance measured by the time it takes for light from that object to reach us. It is given by the radial integral obtained from (4.12), assuming that the observer is at $r = 0$:

$$\int_0^{D_M} \frac{dr}{\sqrt{1-kr^2}} = \begin{cases} \frac{D_H}{\sqrt{\Omega_k}} \sinh^{-1}[\sqrt{\Omega_k} D_M/D_H] & \text{for } \Omega_k > 0 \\ D_M & \text{for } \Omega_k = 0 \\ \frac{D_H}{\sqrt{|\Omega_k|}} \sin^{-1}[\sqrt{\Omega_k} D_M/D_H] & \text{for } \Omega_k < 0 \end{cases} \tag{4.55}$$

where we have used the fact that $k = -\Omega_k H_0^2/c^2 = -\Omega_k/D_H^2$ from (4.51). This means that the proper distance is given by

$$D_M = \begin{cases} \frac{D_H}{\sqrt{\Omega_k}} \sinh[\sqrt{\Omega_k} D_C/D_H] & \text{for } \Omega_k > 0 \\ D_C & \text{for } \Omega_k = 0 \\ \frac{D_H}{\sqrt{|\Omega_k|}} \sin[\sqrt{\Omega_k} D_C/D_H] & \text{for } \Omega_k < 0 \end{cases} \tag{4.56}$$

As the curvature takes the opposite sign to Ω_k, it is useful to remark that we re-obtain the hyperbolic and trigonometric functions for open and closed universe geometries respectively.

Angular Distance

Suppose now that we look at an object of a finite size which is transverse to our line of sight, and lies a certain distance from us. If we divide the physical transverse size of the object by the angle that the object subtends in the sky (the angular size), we obtain the angular diameter distance:

$$D_A = \frac{D_M}{1+z} \tag{4.57}$$

Hence, if we know the size of an object and its redshift, we can calculate, for a given universe, D_A.

Luminosity Distance

Alternatively, we may know the brightness or luminosity of an object at a given distance. We know that the flux of that object at a distance D_L should be given by an expression of the form

$$F = \frac{L}{4\pi D_L^2} \tag{4.58}$$

D_L is known as the luminosity distance, and is related to the other distances through

$$D_L = (1+z)D_M = (1+z)^2 D_A \tag{4.59}$$

Note that for small $z \ll 1$, all these distances converge to $zD_H = zc/H_0$.

4.3 Large Scale Dynamics

Thus far in this chapter, we have simply introduced many definitions and equations that are used in the study of cosmology. In this section, we shall apply them to examine the behaviour of the universe under certain conditions. We know that, for a given curvature, the behaviour of the universe is entirely described by the scale factor $a(t)$, and so our aim is generally to solve (4.52) for $a(t)$, giving the integral

$$\boxed{\int_0^t dt = \frac{1}{H_0}\int_0^a \frac{da}{a\left[\Omega_\gamma a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda\right]^{1/2}}} \tag{4.60}$$

Evidently, this integral is not (easily) analytically solvable in the above form, and so we shall consider particular cases. Throughout these calculations, we shall set $a_0 = 1$, as is conventional to do so.

4.3.1 The Deceleration Parameter

Let us expand the scale factor $a(t)$ around the current time $t = t_0$:

$$a(t) = a(t_0) + \dot{a}(t_0)\,(t-t_0) + \frac{1}{2}\ddot{a}(t_0)(t-t_0)^2 + \dots \tag{4.61}$$

If we divide throughout by $a(t_0)$, and recall the definition of the Hubble parameter (4.11), we can write

$$\frac{1}{1+z} = \frac{a(t)}{a(t_0)} = 1 + H_0(t-t_0) - \frac{q_0}{2}H_0^2(t-t_0)^2 + \dots \tag{4.62}$$

which defines the *deceleration parameter* q_0 as

$$\boxed{q_0 = -\frac{a_0\ddot{a}_0}{\dot{a}_0^2}} \tag{4.63}$$

Note that for $q_0 < 0$, the expansion of the universe is accelerating. We can write the second Freidmann equation (4.43) as

$$q_0 = \frac{1}{2\rho_c}\left(\rho + \frac{3p}{c^2}\right) = \frac{1}{2\rho_c}\sum_i(1+3w_i)\rho_i = \frac{1}{2}\sum_i(1+3w_i)\Omega_i \tag{4.64}$$

where all mass ratios are evaluated at the current time. This means that q_0 is directly related to the density ratios for the matter of the given universe that we are concerned with.

4.3.2 Universe Types

Let us examine some particular cases of the various combination of the parameters $k, \Omega_\gamma, \Omega_m, \Omega_k$ and Ω_Λ that we could have to describe a universe. The density ratios shall be assumed zero unless indicated. We shall be using (4.60) in all cases, and integrals shall be from $t = 0$ and $a = 0$ to some arbitrary t and a. These are generally grouped into three different categories, corresponding to flat, open and closed universes.

Flat Universes

Flat universes are those with $k = 0, \Omega_k = 0$, sometimes referred to as the *Einstein-deSitter universe.* We shall consider four special cases:

- Matter Dominated ($\Omega_m = 1$):

$$\int dt = \frac{1}{H_0}\int da\, a^{1/2} \quad \longrightarrow \quad a(t) = \left(\frac{3}{2}H_0 t\right)^{2/3} \tag{4.65}$$

By this model, the age of the universe is given by $t_0 = 2/3H_0 \approx 10^9$ years.

- Radiation Dominated ($\Omega_\gamma = 1$):

$$\int dt = \frac{1}{H_0}\int da\, a \quad \longrightarrow a(t) = (2H_0 t)^{1/2} \tag{4.66}$$

This leads to slightly slower growth than in the matter dominated case.

- Vacuum Dominated ($\Omega_\Lambda = 1$):

$$\int dt = \frac{1}{H_0}\int da\, \frac{1}{a} \quad \longrightarrow \quad a(t) = e^{H_0 t} \tag{4.67}$$

Thus, a vacuum dominated flat universe expands exponentially.

- Radiation and Matter (Ω_γ, Ω_m):

$$\int dt = \frac{1}{H_0}\int da\, \frac{1}{a\left[\Omega_\gamma a^{-4} + \Omega_m a^{-3}\right]^{1/2}} \tag{4.68}$$

Defining the *inferno ratio* $I_{\gamma m} = \Omega_\gamma/\Omega_m$, we can write this integral as

$$\Omega_m^{1/2} H_0 t = \int da\, \frac{a}{\left[a + I_{\gamma m}\right]^{1/2}} \tag{4.69}$$

which has solution

$$\frac{3}{2}\Omega_m^{1/2} H_0 t = (a + I_{\gamma m})^{3/2} - 3cI_{\gamma m}(a + I_{\gamma m})^{1/2} + 2I_{\gamma m}^{3/2} \tag{4.70}$$

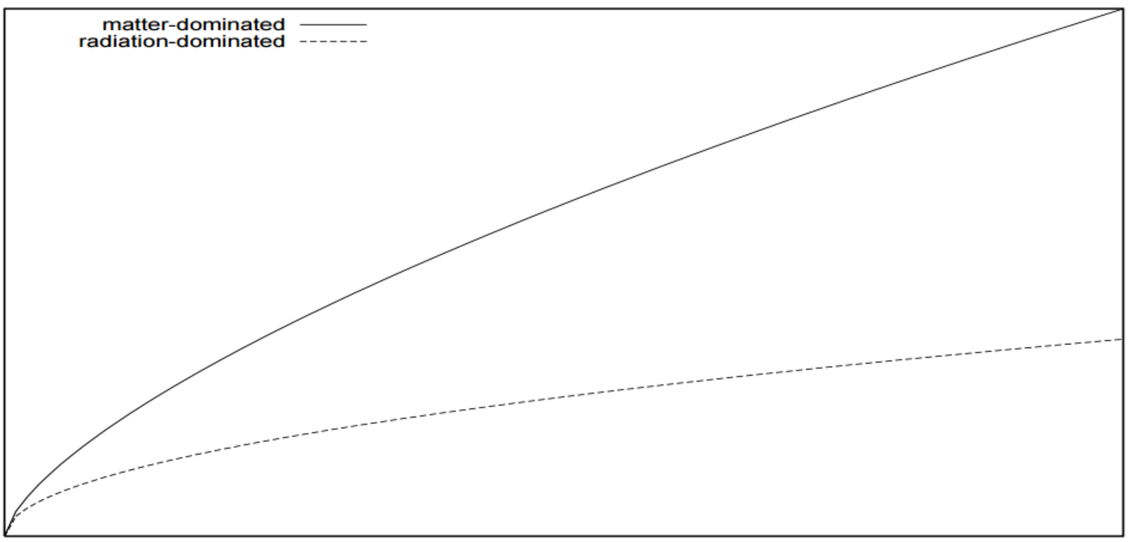

Figure 4.1: A plot of the scale factor (vertical axis) against time (horizontal axis) for the matter dominated and radiation dominated cases of a closed universe

Open Universes

For open universes, $k = -1, \Omega_k > 0$. When dealing with such systems, it is often useful to introduce the *conformal time* $d\eta = dt/a$. Recall that $\Omega_k = -k/H_0^2$ (in units of $a_0 = 1$). We consider two special cases:

- Matter dominated (Ω_m):

$$\int dt = \frac{1}{H_0}\int da\,\frac{1}{a\left[\Omega_m a^{-3} + \Omega_k a^{-2}\right]^{1/2}} = \int da\,\frac{1}{\left[I_{mk}a^{-1} + 1\right]^{1/2}} \tag{4.71}$$

where we have introduced the ratio $I_{mk} = \Omega_m/\Omega_k$. If we write the time integral in terms of the conformal time η, these integrals become

$$\int d\eta = \int da\,\frac{1}{\left[I_{mk}a + a^2\right]^{1/2}} \quad\longrightarrow\quad \eta = \cosh^{-1}\left(\frac{2a}{I_{mk}} + 1\right) \tag{4.72}$$

This can be re-arranged to give

$$a(t) = \frac{1}{2}I_{mk}\left(\cosh\eta - 1\right) \tag{4.73}$$

Now, $dt = ad\eta$, meaning that

$$t = \int d\eta\, a = \frac{1}{2}I_{mk}\int d\eta\,(\cosh\eta - 1) = \frac{1}{2}I_{mk}(\sinh\eta - \eta) \tag{4.74}$$

This means that the open, matter dominated universe is described by parametric equations in the conformal time:

$$a(t) = \frac{1}{2}I_{mk}\left(\cosh\eta - 1\right), \quad t = \frac{1}{2}I_{mk}(\sinh\eta - \eta), \quad I_{mk} = \frac{\Omega_m}{\Omega_k} \tag{4.75}$$

- Radiation dominated (Ω_γ):

$$\int dt = \frac{1}{H_0}\int da\,\frac{1}{a\left[\Omega_\gamma a^{-4} + \Omega_k a^{-2}\right]^{1/2}} = \int da\,\frac{1}{\left[I_{\gamma k}a^{-2} + 1\right]^{1/2}} \tag{4.76}$$

where we have introduced the ratio $I_{\gamma k} = \Omega_\gamma/\Omega_k$. Again making use of conformal time,

$$\int d\eta = \int da\,\frac{1}{\left[I_{\gamma k} + a^2\right]^{1/2}} \quad\longrightarrow\quad \eta = \sinh^{-1}\left(\frac{a}{I_{\gamma k}^{1/2}}\right) \tag{4.77}$$

Transforming back to normal time coordinates, we have that

$$a(t) = I_{\gamma k}^{1/2}\sinh\eta, \quad t - t_0 = I_{\gamma k}^{1/2}\cosh\eta \tag{4.78}$$

for some constant t_0. The fact that $\eta = 0$ at $t = 0$ means that $t_0 = -I_{\gamma k}^{1/2}$. Thus, the open, radiation dominated universe is described by the parametric equations

$$a(t) = I_{\gamma k}^{1/2}\sinh\eta, \quad t = I_{\gamma k}^{1/2}\left(\cosh\eta - 1\right), \quad I_{\gamma k} = \frac{\Omega_\gamma}{\Omega_k} \tag{4.79}$$

At early times, $\eta \ll 1$, meaning that we can expand the trigonometric and hyperbolic functions in η to approximate behaviour during the early universe:

$$\sin\eta = \eta - \frac{1}{3!}\eta^3, \quad \cos\eta = 1 - \frac{1}{2!}\eta^2$$
$$\sinh\eta = \eta + \frac{1}{3!}\eta^3, \quad \cosh\eta = 1 + \frac{1}{2!}\eta^2$$

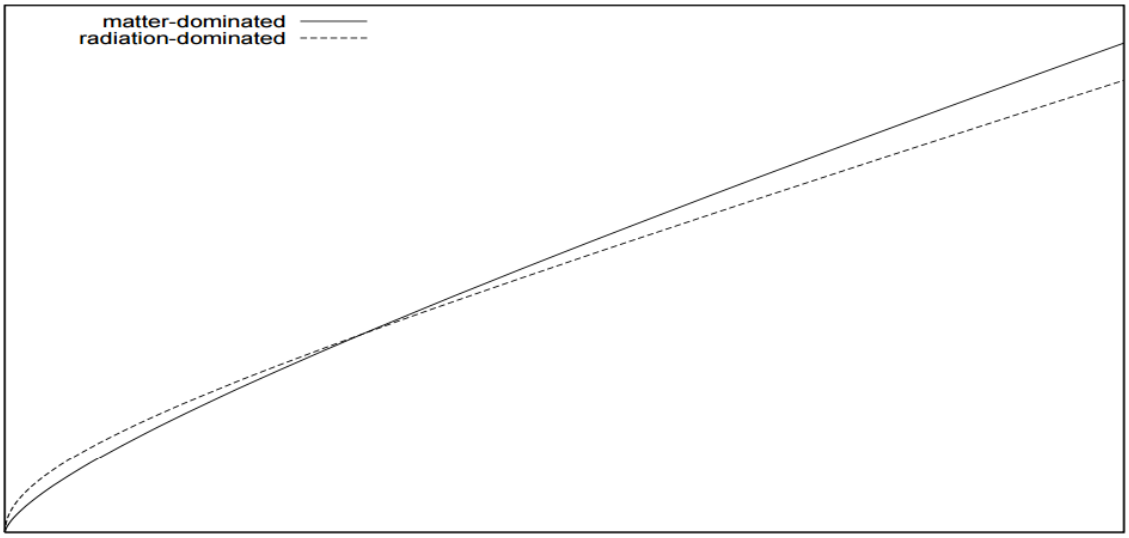

Figure 4.2: A plot of the scale factor (vertical axis) against time (horizontal axis) for the matter dominated and radiation dominated cases of an open universe

Closed Universes

For closed universes, $k = 1, \Omega_k < 0$. We again consider two special cases:

- Matter dominated (Ω_m):

$$\int dt = \frac{1}{H_0}\int da \, \frac{1}{a\left[\Omega_m a^{-3} + \Omega_k a^{-2}\right]^{1/2}} = \int da \, \frac{1}{\left[|I_{mk}|\, a^{-1} - 1\right]^{1/2}} \tag{4.80}$$

where I_{mk} is defined as before. We have included the absolute value, as by definition Ω_k is negative for closed universes. Using conformal time:

$$\int d\eta = \int da \, \frac{1}{\left[|I_{mk}|\, a - a^2\right]^{1/2}} \quad \longrightarrow \quad \eta = \sin^{-1}\left(\frac{2a - |I_{mk}|}{|I_{mk}|}\right) + \frac{\pi}{2} \tag{4.81}$$

such that

$$\frac{2a - |I_{mk}|}{|I_{mk}|} = \sin\left(\eta - \frac{\pi}{2}\right) = -\cos\eta \quad \longrightarrow \quad a(t) = \frac{1}{2}\,|I_{mk}|\,(1 - \cos\eta) \tag{4.82}$$

Using the fact that $dt = a d\eta$,

$$t = \int d\eta \, a = \frac{1}{2}\,|I_{mk}| \int d\eta \,(1 - \cos\eta) = \frac{1}{2}\,|I_{mk}|\,(\eta - \sin\eta) \tag{4.83}$$

This means that the scale factor for a closed, matter dominated universe is given by the parametric equations

$$a(t) = \frac{1}{2}\,|I_{mk}|\,(1 - \cos\eta), \quad t = \frac{1}{2}\,|I_{mk}|\,(\eta - \sin\eta), \quad I_{mk} = \frac{\Omega_m}{\Omega_k} \tag{4.84}$$

- Radiation dominated (Ω_γ):

$$\int dt = \frac{1}{H_0}\int da \, \frac{1}{a\left[\Omega_\gamma a^{-4} + \Omega_k a^{-2}\right]^{1/2}} = \int da \, \frac{1}{\left[|I_{\gamma k}|\, a^{-2} - 1\right]^{1/2}} \tag{4.85}$$

where $I_{\gamma k}$ is defined as before. Using the conformal time, we have that

$$\int d\eta = \int da \, \frac{1}{\left[|I_{\gamma k}| - a^2\right]^{1/2}} \quad \longrightarrow \quad \eta = \sin^{-1}\left(\frac{a}{|I_{\gamma k}|}\right) \tag{4.86}$$

As before, transforming back to normal time coordinates yields

$$a(t) = |I_{\gamma k}|^{1/2} \sin\eta, \quad t - t_0 = -|I_{\gamma k}|^{1/2} \cos\eta \tag{4.87}$$

for some constant t_0. The fact that $\eta = 0$ at $t = 0$ means that $t_0 = |I_{\gamma k}|^{1/2}$, meaning that we finally have

$$a(t) = |I_{\gamma k}|^{1/2} \sin\eta, \quad t = |I_{\gamma k}|^{1/2} (1 - \cos\eta), \quad I_{\gamma k} = \frac{\Omega_\gamma}{\Omega_k} \tag{4.88}$$

We note that in both cases, the evolution of $a(t)$ is cycloidal; the scale factor grows at an ever decreasing rate, until it reaches a point at which the expansion is halted and reversed. It then starts to shrink, before finally collapsing again.

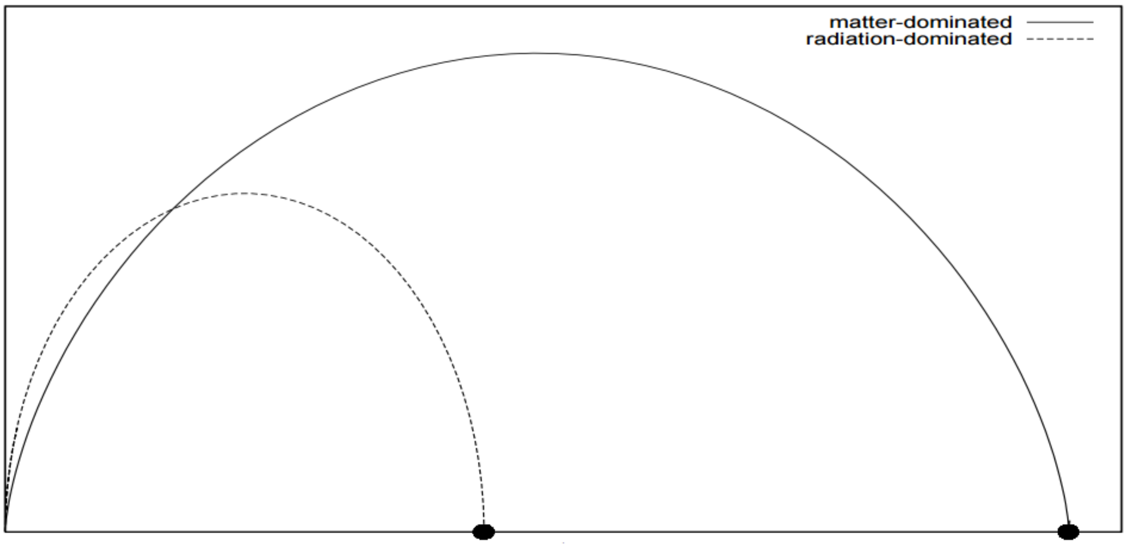

Figure 4.3: A plot of the scale factor (vertical axis) against time (horizontal axis) for the matter dominated and radiation dominated cases of an closed universe

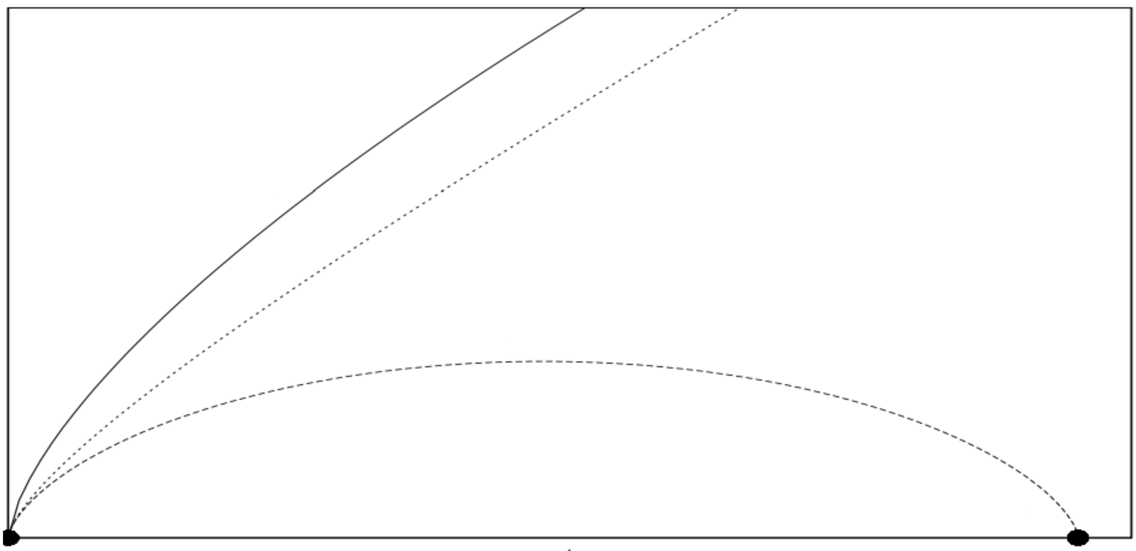

Figure 4.4: A plot of the scale factor (vertical axis) against time (horizontal axis) for the matter dominated models in the flat (dots), open (solid) and closed (dashes) universes

4.3.3 Our Universe

Looking at all these special cases is all very well and good, but how does this apply to the universe that we actually inhabit? From observation, we can estimate the values for the various relevant parameters at the current time t_0:

$$\begin{aligned} H_0 &= 70.4 \pm 1.4 \ \text{kms}^{-1}\text{Mpc}^{-1} \\ \rho_c &= 9.3 \times 10^{-30} \ \text{gcm}^{-3} \\ \Omega_m &= 0.273 \pm 0.014 \\ \Omega_\gamma &= (8.44 \pm 0.5) \times 10^{-5} \\ \Omega_k &= 0 \pm 0.02 \\ \Omega_\Lambda &= 0.728 \pm 0.016 \end{aligned}$$

The current paradigm is thus that the universe is flat $\Omega_k \approx 0$, but contains matter, radiation, and has a non-zero contribution from vacuum energy. From the data, it is clear that the dominant contribution is from the vacuum energy Λ, meaning that we are currently in a period of exponential growth. But before $z \approx 5$, the universe was dominated by cold matter, and $z \approx 3200$ the universe was dominated by radiation.

We already have analytic solutions for both the radiation dominated era (4.66) and the matter dominated era (4.65), which we shall repeat explicitly here

$$\text{Radiation dominated era :} \quad a(t) \approx \left(2\Omega_\gamma^{1/2} H_0 t\right)^{1/2} \tag{4.89}$$

$$\text{Matter dominated era :} \quad a(t) \approx \left(\frac{3}{2}\Omega_m^{1/2} H_0 t\right)^{2/3} \tag{4.90}$$

where we have re-introduced the relevant density ratios, which are evaluated at the current time. The transition from the radiation dominated era to the matter dominated era occurs when $\Omega_m a^{-3} = \Omega_\gamma a^{-4}$, at which $a = 0.00031$. Inserting this into (4.89), this occurred at a time roughly 7×10^4 years after the Big Bang at $t = 0$.

The late universe (small z), in which both matter and vacuum energy are important but radiation is unimportant, can be integrated analytically. This is fortunate, as it is the era that we can most readily observe. Let $0 < \Omega_m < 1$ and $\Omega_\Lambda = 1 - \Omega_m$, and set $\Omega_\gamma = \Omega_k = 0$ in (4.60). Then

$$\int dt = \frac{1}{H_0}\int da \, \frac{1}{a\left[\Omega_m a^{-3} + \Omega_\Lambda\right]^{1/2}} = \frac{1}{H_0}\int da \, \frac{a^{1/2}}{\left[\Omega_m + \Omega_\Lambda a^3\right]^{1/2}} \tag{4.91}$$

Using the substitution $x = (\Omega_\Lambda/\Omega_m)^{1/2} a^{3/2}$, this integral can be evaluated as

$$t = \frac{2}{3H_0}\frac{1}{\Omega_\Lambda^{1/2}}\int dx \, \frac{1}{(1+x^2)^{1/2}} = \frac{2}{3H_0\sqrt{1-\Omega_m}} \sinh^{-1}\left(\sqrt{\frac{1-\Omega_m}{\Omega_m}} a^{3/2}\right) \tag{4.92}$$

It is easy to verify that for small a, this approximates (4.90), and for $a \gg 1$, this formula describes an exponentially expanding universe. The above formula is very accurate for all redshifts up to $z \approx 1000$. This means that we can use it to accurately determine the age of the universe; let $a = 1$ in (4.92), to obtain an age of $t_{\text{universe}} \approx 13.7\,\text{Gy}$.

Horizons

The universe evidently has a definite age. This means that there is a maximum distance that light could have travelled during this time. More importantly, this light could only have travelled a finite comoving distance in that time; any light emitted in the part of the universe beyond this maximum distance will not have reached us. This is known as a *horizon*. For the early universe, $t \to 0$, and $a \to 0$, with the only surviving term in (**??**) being the radiation term. Let us consider $\Omega_\gamma a^{-n}$ instead of $\Omega_\gamma a^{-4}$. We thus obtain

$$D_C(a \to 0) = \frac{c}{H_0\Omega_0^{1/2}} \int_0^1 \frac{da}{a^{2-n/2}} \stackrel{?}{=} \frac{c}{H_0\Omega_0^{1/2}} \left(\frac{n}{2} - 1\right)^{-1} \tag{4.93}$$

For $n > 2$, this converges to a finite value (as above), which is the aforementioned horizon. For $n < 2$, the integral diverges, meaning that there is no horizon. As it is governed by the propagation of photons, the horizon expands at the speed of light, meaning that more of the universe becomes visible as time passes. Unless, however, the universe expands so fast that it can outrun the expansion of the horizon. Let us consider the case of being far in the future, corresponding to $a \to \infty$. We obtain the same integral as above, but with different limits:

$$D_C(a \to \infty) = \frac{c}{H_0\Omega_0^{1/2}} \int_1^\infty \frac{da}{a^{2-n/2}} \stackrel{?}{=} \frac{c}{H_0\Omega_0^{1/2}} \left(1 - \frac{n}{2}\right)^{-1} \tag{4.94}$$

Note that this is negative, as it is a distance into the future. For $n < 2$, this converges to a finite number, as shown above. This means that light sent out today will only ever be seen by a finite portion of the universe. This is what is sometimes referred to as the *horizon problem*. For an exponentially expanding universe, the distance for this horizon remains constant in time, even though the universe expands. This means that galaxies that are now inside the horizon will eventually pass through it, and be lost, assuming that the exponential expansion continues.

4.3.4 The Cosmological Distance Ladder

The measurements at the start of the previous section have quite large error bars - well, in comparison to results in other areas of physics - because measuring scales on a cosmological or intergalactic scale is actually quite difficult. We have a few ways of measuring distances, depending on distance range in which we are measuring. These are as follows:

- Parallax (≈ 100 parsecs) - By observing the apparent angular displacement of objects as the Earth orbits the Sun, we can determine the distance to them. Suppose that a star is a distance D from earth, and the Earth moves a distance $2d$ from its original position. Then, the half-angle subtended is $\alpha = d/D$ in the same angle approximation. If we let d be the distance of the Earth from the sun, then $D = \alpha^{-1}$, where α is measured in arcseconds, while D is measured in parsecs (3.09×10^{13} km)

- Spectroscopic Parallax (≈ 10 kpc) - This involves measuring luminosities, in order to determine luminosity distances. Stars emit roughly as blackbodies, meaning that we can calculate the temperature of the star from the peak in its blackbody radiation curve. Assuming that it obeys the Stefan-Boltzmann law $P = \sigma T^4$, then the luminosity of the star is $L = 4\pi R^2 \sigma T^4$. We can calculate the radius R of the star from the Doppler broadening of the source (broader lines correspond to smaller stars, as there is a greater variation in velocity). Comparing the observed flux on Earth to the luminosity gives us D_L

- Cepheid Variable Stars ($\approx 30\,\text{Mpc}$) - These stars have a time-variable luminosity that occurs on a period of days, with a magnitude of about 100-1000 $L_\odot$, meaning that they can be observed at great distances. The variation in luminosity occurs as a result of the steady-state expansion and contraction of the surface of the star. Enough measurements of the distances of these stars from others builds up a reference database from which we can determine distances

- Galaxy Rotation Curves ($\approx 100\,\text{Mpc}$) - Galaxies have an intrinsic rotation curve, which is related to the mass and type of the galaxy. This causes Doppler broadening of certain spectral emission lines, which can be related to the mass of the galaxy, and then the luminosity assuming that we have some information about its density profile

- Type 1a Supernovae ($\approx 1000\,\text{Mpc}$) - Supernovae are some of the most (optically) luminous events in the universe, outputting luminosities of order $10^9 L_\odot$. If we observe these, we can measure their brightness, and infer distances from their luminosity. In particular, type 1a supernovae have the similar behaviour that the rate at which the luminosity fades is related to the luminosity at the time of explosion; we can use this to determine its luminosity. This type of supernovae occur when a white dwarf gravitationally accrues just enough material to pass the threshold for supernovae, and as such are quite rare

A combination of these techniques are used to probe the observable universe. From what we learnt in chapter 3, there may also be another possible method; detecting gravitational radiation from massive objects. This would have the added benefit that the radiation is so weakly interacting that this technique would have a very long reach.

4.4 Thermal History of the Universe

We have shown, in quite painstaking detail, that the universe is currently expanding, and has been expanding for approximately 13.7 billion years. As such, there would have been a time close to $t = 0$ when it would have been very hot and dense. However, we know that the average temperature of the universe is about 2.73 K, meaning that it must have undergone a period of cooling. We shall have a look at the effect of this cooling on the matter within the universe.

4.4.1 Equilibrium

Let us make the simplifying assumption that the universe is in thermal equilibrium with itself, and therefore behaves as a blackbody. We can model the radiation as a quantum gas of photons with two possible polarisations, meaning that the density of states can be written as

$$g(k)d^3\mathbf{k} = \underbrace{2}_{\text{polarisation states}} \frac{V}{(2\pi)^3} 4\pi k^2 dk = \frac{Vk^2}{\pi^2} dk \tag{4.95}$$

Using the photon dispersion relation $\omega = ck$, this can be written as

$$g(\omega)d\omega = \frac{V}{\pi^2 c^3}\omega^2 d\omega \tag{4.96}$$

As the number density of the photons is quite large in thermal equilibrium, we can estimate them as having a continuous spectrum, such that we adopt Bose-Einstein statistics to write the mean occupation number as

$$\bar{n}_i = \frac{1}{e^{\beta\hbar\omega} - 1} \tag{4.97}$$

where we have introduced $\beta = (k_B T)^{-1}$. The spectral energy density is thus given by

$$\boxed{\rho(\omega) = \bar{n}_i g(\omega)\hbar\omega = \frac{\hbar}{\pi^2 c^3}\frac{\omega^3}{e^{\beta\hbar\omega} - 1}} \tag{4.98}$$

This is known as the *Plank distribution*. Integrating over all frequencies:

$$\rho_\gamma c^2 = \int d\omega\, \rho(\omega) = \frac{\hbar}{\pi^2 c^3}\int_0^\infty d\omega\, \frac{\omega^3}{e^{\beta\hbar\omega} - 1} = \frac{\hbar}{\pi^2 c^3}\frac{1}{(\hbar\beta)^4}\underbrace{\int_0^\infty dx\, \frac{x^3}{e^x - 1}}_{\pi^4/15} \tag{4.99}$$

where we have made the substitution that $x = \beta\hbar\omega$. Thus, the energy density of the photons is given by

$$\rho_\gamma c^2 = \frac{\pi^2}{15}(k_B T)\left(\frac{k_B T}{\hbar c}\right)^3 \propto T^4 \tag{4.100}$$

However, we know that $\rho_\gamma \propto a^{-4}$, which implies that the temperature of the universe scales as $T \propto a^{-1}$. This is valid if all matter in the universe interacts with photons (at sufficiently high temperatures, everything interacts strongly), and if radiation dominates over the other forms of matter in the universe. On this last point, we know that $\rho_\gamma \propto a^{-4}$, while $\rho_m \propto a^{-3}$, meaning that one would assume that $\rho_\gamma \ll \rho_m$ after sufficient evolution. However, if we define the number density of photons n_γ, we know that this scales in the same way as the number of baryons (non-relativistic particles, like protons and neutrons). Defining the *baryon to entropy ratio*

$$\eta_B = \frac{n_B}{n_\gamma} \approx 10^{-10} \tag{4.101}$$

it is clear that the number density of photons is dominant. This means that the temperature of photons effectively determines the temperature of the universe, or rather, the temperature of the universe scales as the inverse of the scale factor.

Matter in Equilibrium

We can obtain an idea of how matter behaves during the expansion by considering how its energy density evolves. For a particle of mass m, energy E and chemical potential μ, the associated energy level occupation numbers are given by

$$\bar{n} = \frac{1}{e^{\beta(E-\mu)} \pm 1} \tag{4.102}$$

where the positive sign corresponds to Fermi-Dirac statistics, while the negative sign corresponds to Bose-Einstein statistics. We shall assume that the matter obeys at least one of these distributions. We can examine two limits analytically:

- Early times - In the early period of the universe, $k_B T \gg mc^2$ as a is very small. The energy for a (massive) relativistic particle is given by $E^2 = (mc^2)^2 + (\hbar kc)^2$, so in this limit we can approximate that $E \approx \hbar kc$. This means that the density of states is given by
$$g(E)dE = \frac{(2s+1)V}{2\pi^2(\hbar c)^3}E^2 \tag{4.103}$$
where $(2s+1)$ is the spin degeneracy. Let us examine the energy density, given by
$$\rho_m c^2 = \frac{1}{V}\int_0^\infty dE\, Eg(E)\bar{n} = \frac{(2s+1)}{2\pi^2(\hbar c)^3}\int_0^\infty dE\, \frac{E^3}{e^{\beta(E-\mu)} \pm 1} \tag{4.104}$$
We can non-dimensionalise this integral using the substitution $x = \beta E$, such that we obtain
$$\rho_m c^2 = \frac{(2s+1)}{2\pi^2}(k_B T)\left(\frac{k_B T}{\hbar c}\right)^3 \int_0^\infty dx\, \frac{x^2}{e^{x-\beta\mu} \pm 1} \tag{4.105}$$
For massive particles, the chemical potential scales as $\mu \backsim -mc^2$, meaning that we can essentially ignore it it in this limit. Then, the integral expression above is simply a number. This means that $\rho_m c^2 \propto T^4$, which is the same scaling as radiation

- Later times - We have already shown that the universe cools as it expands. This means that at some point, we shall be at the limit where $k_B T \ll mc^2$. In this case, the energy is simply the kinetic energy $E = \hbar^2 k^2/2m$, meaning that the density of states becomes
$$g(E)dE = \frac{(2s+1)Vm^{3/2}}{\sqrt{2}\pi^2\hbar^3}E^{1/2} \tag{4.106}$$
Then, the number density is given by
$$n_m = \frac{1}{V}\int_0^\infty dE\, g(E)\bar{n} = \frac{(2s+1)m^{3/2}}{\sqrt{2}\pi^2\hbar^3}\int_0^\infty dE\, \frac{E^{1/2}}{e^{\beta(E-\mu)} \pm 1} \tag{4.107}$$
Once again, we non-dimensionalise the integral using $x = \beta E$, such that we obtain
$$n_m = (2s+1)\left(\frac{mk_B T}{2\pi\hbar^2}\right)^{3/2} \frac{2}{\pi^{1/2}}\int_0^\infty dx\, \frac{x^{1/2}}{e^{\beta(E-\mu)} \pm 1} \tag{4.108}$$
In this limit, we have that $\beta\mu \gg 1$, for a hot $(T \to \infty)$, dilute $(n \to 0)$ gas. Then, the integral can be written as
$$\frac{2}{\pi^{1/2}}\int_0^\infty dx\, \frac{x^{1/2}}{e^{\beta(E-\mu)} \pm 1} \approx \frac{2}{\pi^{1/2}}\int_0^\infty dx\, x^{1/2}e^{-x}e^{\beta\mu} = e^{\beta\mu} \tag{4.109}$$

This means that we can write our number density as

$$\boxed{n_m = (2s+1)\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{3/2} e^{\beta(\mu - mc^2)}} \tag{4.110}$$

where we have rescaled the chemical potential to as to exclude the energy cost associated with the rest mass of the particle. This is clearly the classical result, meaning that the associated energy density is $\rho_m c^2 = n_m mc^2$, and pressure $p = n_m k_B T$. Note that $p \ll \rho c^2$, meaning that pressure is negligible for non-relativistic matter

What we have shown here is that at sufficiently early times in the expansion, matter behaves like radiation. As it cools to temperatures below the mass thresholds, the number of particles behaving relativistically decreases significantly, until we arrive at the present day, where the vast majority of matter behaves non-relativistically.

4.4.2 Recombination

The composition of the early universe is thus a highly ionised plasma, consisting of a relativistic mixture of photons, electrons and nuclei. At high enough temperatures, electrons will not be able to kind to nuclei, as the ambient thermal energy is simply too high. However, we can imagine that the universe cools, electrons will in fact begin to become bound to nuclei. Let us consider the case of hydrogen, which has an ionisation energy of $\mathcal{R} = 13.6\,\text{eV}$. We can imagine that around $k_BT = \mathcal{R}$, there is a transition from ionised to bound hydrogen.

The Saha Equation

To investigate this, we consider the chemical process

$$p^+ + e^- \leftrightarrow H + \gamma \tag{4.111}$$

By the law of mass action, the chemical potentials on either side of the reaction must balance, meaning that

$$\mu_p + \mu_e = \mu_H + \cancel{\mu_\gamma} \tag{4.112}$$

However, we know that the chemical potential of a photon is zero, so we can ignore this (as indicated). We shall model the other baryonic species using (4.110), meaning that we can re-arrange for the chemical potentials

$$\mu = -k_BT\log\left(\frac{n_Q}{n}\right) + mc^2, \quad n_Q = (2s+1)\left(\frac{mk_BT}{2\pi\hbar^2}\right)^{3/2} \tag{4.113}$$

where we have introduced the *quantum concentration* n_Q. We shall ignore the spin degeneracy factor, as must globally cancel in the reaction. Then, we have that

$$-k_BT\left[\log\left(\frac{n_Q^p}{n_p}\right) + \log\left(\frac{n_Q^e}{n_e}\right) - \log\left(\frac{n_Q^H}{n_H}\right)\right] + m_pc^2 + m_ec^2 - m_Hc^2 = 0 \tag{4.114}$$

Recognising that $m_pc^2 + m_ec^2 - m_Hc^2 = \mathcal{R}$, and assuming that $m_p \approx m_h$ such that $n_Q^P \approx n_Q^H$, it follows that

$$\frac{n_pn_e}{n_H} = n_Q^e e^{-\beta\mathcal{R}} \tag{4.115}$$

Then, the total number of baryons is given by $n_B = n_p + n_H = n_e + n_H$, where the second expression follows from the assumption of charge neutrality. Let us define the *ionisation ratio*

$$\chi = \frac{n_e}{n_p + n_H} = \frac{n_e}{n_B} \tag{4.116}$$

It follows that

$$\boxed{\frac{1-\chi}{\chi^2} = n_B \left(\frac{m_e k_B T}{2\pi\hbar^2}\right)^{-3/2} e^{\beta\mathcal{R}}} \tag{4.117}$$

This is the *Saha equation*, which allows us to calculate the fraction of ionised hydrogen atoms given knowledge of the temperature ($T = 2.728(1+z)$ K) and the baryon number density ($n_B = 1.6(1+z)^3\,\mathrm{m}^{-3}$). At early times, $\chi \approx 1$. The step-function like transition occurs at $T \approx 3570$ K at a redshift of $z \approx 1100$. This is known as *recombination.*

Cosmic Microwave Background

At very early times, before recombination, the photon radiation will have been in thermal equilibrium with itself, and would satisfy a Plank distribution (4.98), with a very low mean-free path. After re-combination, the photons no longer interact with matter, and their mean-free path becomes infinite. This means that their distribution would remain a Plank distribution, only with the peak shifting with $T \propto a^{-1}$. This is what we observe as the *Cosmic Microwave Background (CMB)*; a homogeneous, isotropic distribution of photons that have propagated to us from recombination. We can calculate the temperature of the CMB from the fact that

$$a(t)T = \text{constant} \quad \longrightarrow \quad \frac{T_0}{T_{\text{rec}}} = \frac{a(t_{\text{rec}})}{a(t_0)} \tag{4.118}$$

where T_0 is the temperature of the CMB, while T_{rec} and a_{rec} are the temperature and scale factor at recombination respectively. Using the information above, we find that the temperature of the CMB is roughly $T_0 \approx 2.728$ K. All the photons that we observe at this temperature were emitted from the *surface of last scattering*; the spacetime hypersurface corresponding to the time of recombination.

4.4.3 Nucleosynthesis

We shall now examine the behaviour of the universe at $k_B T \approx 1$ MeV. At this temperature, we can no-longer assume thermal equilibrium, as nuclear processes being to play an important role. Let us consider protons and neutrons. These can combine to form the nuclei of the elements, such as

$$\begin{aligned} 1p &\to \text{Hydrogen nucleus} \\ 1p + 1n &\to \text{Deuterium Nucleus} \\ 1p + 2n &\to \text{Tritium Nucleus} \end{aligned}$$

and so on. The neutrons and protons can also convert into one another through weak electromagnetic interactions. Let us initially consider the equilibrium. Assuming that $\mu_p = \mu_n$, we can use (4.113) to show that in equilibrium,

$$\frac{n_n^{\text{eq}}}{n_p^{\text{eq}}} = e^{-\beta\Delta E}, \quad \Delta E = m_n c^2 - m_p c^2 \tag{4.119}$$

At high temperatures, $n_n \approx n_p$. However, as temperatures decrease, the number of neutrons falls exponentially, and there would barely be any neutrons at the present day $T_0 = 2.728$ K. This means that the equilibrium approach cannot be correct.

Beyond Equilibrium

Let use consider what is occurring when these two species are converting into one another. The reaction can be categorised by some rate Γ, and it must compete against the expansion of the universe, with which we can associated the rate $H = \dot{a}/a$. The relative sizes of Γ and H dictate how important the reactions are in keeping the neutrons and protons equilibrated. One can write down a Boltzmann distribution for the comoving neutron number N_n, which is simply the number of neutrons within a comoving volume. This is given by

$$\boxed{\frac{d\log N_n}{d\log a} = -\frac{\Gamma}{H}\left[1 - \left(\frac{N_n^{\text{eq}}}{N_n}\right)^2\right]} \tag{4.120}$$

where $N_n^{\text{eq}} = a^3 n_n^{\text{eq}}$ is the equilibrium expression above. If $\Gamma \gg H$, then we have that $N_n \approx N_n^{\text{eq}}$, meaning that we can indeed use the equilibrium value predicted by (4.119). However, if $\Gamma \ll H$, then the expansion of the universe will dominate, and inhibit the depletion or creation of neutrons by that reaction. The equation is then

$$\frac{d\log N_n}{d\log a} \approx 0 \tag{4.121}$$

meaning that the comoving neutron number is *frozen out* (and the number density will decay as a^{-3}). Evidently, the transition between the regimes occurs at $\Gamma \approx H$, and will depend on how Γ depends on temperature and masses. It turns out that for this reaction, $k_B T_f \approx 0.8$ MeV. This means that the relative number density of neutrons to protons will be frozen in at $n_n^{\text{eq}}/n_p^{\text{eq}} \approx 1/6$. However, we observe something closer to $1/7$ due to the finite decay time of the neutron.

We can use a very simple argument to find the fraction of helium verses hydrogen in our universe. We initially have $7/8$ in protons and $1/8$ in neutron. Then we need to pair up the protons and neutrons, reducing the number of unpaired protons to $6/8 \approx 75\%$. We thus expect to have roughly 25% of the mass in helium, and 75% in hydrogen.