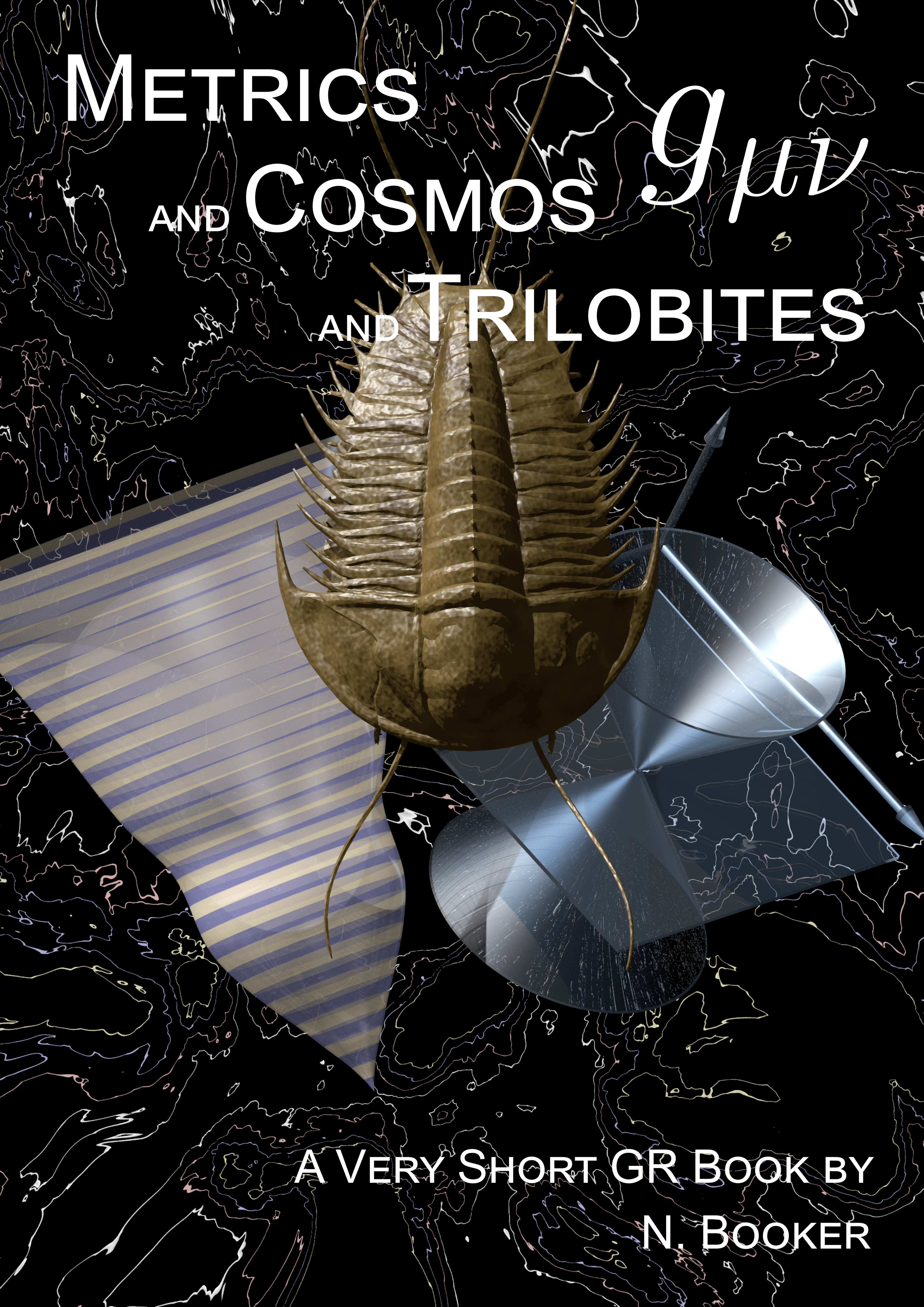
Metrics and Cosmos and Trilobites

$g_{\mu\nu}$



A Very Short GR Book by

N. Booker

To my parents

Contents

Chapter 1

Instead of a foreword

1.1 Acknowledgements

Metrics and Cosmos and Trilobites emerged from a series of lecture notes based on the general relativity and cosmology courses at University College London lectured by Prof. Christian G. Böhmer and Prof. Betti Hartmann respectively. Sources used for the book include:

- *A Course in Modern Mathematical Physics* by Peter Szekeres (University of Adelaide)
- *Cosmology* by Mitchell A. Berger (University of Exeter)
- *Cosmology* by Christian G. Böhmer (University College London)
- *Exact Space-Times in Einstein's General Relativity* by Jerry B. Griffiths (Loughborough University) and Jiří Podolský (Univerzita Karlova)
- *General Relativity* by Mitchell A. Berger (University of Exeter)
- *General Relativity* by Christian G. Böhmer (University College London)
- *General Relativity, Black Holes, and Cosmology* by Andrew J. S. Hamilton (University of Colorado Boulder)
- *Gravitation* by Charles W. Misner (University of Maryland, College Park), Kip S. Thorne (Caltech) and John Archibald Wheeler (University of Texas at Austin)
- *Introduction to 3+1 Numerical Relativity* by Miguel Alcubierre (Universidad Nacional Autónoma de México)
- *The General Theory of Relativity: A Mathematical Approach* by Farook Rahaman (Jadavpur University)

I want to extend my gratitude to the leaders of GR and cosmology at UCL over the past decades: Prof. Mitchell A. Berger lectured GR and cosmology up to 2007, when he left for Exeter, and responded to me when I e-mailed him about his 2004 lecture notes 20 years after their completion. Prof. Christian G. Böhmer lectured GR and cosmology from 2007 to 2024 and 2021 respectively (although with brief interruptions) and supervised my master's thesis. Prof. Betti Hartmann lectured GR and cosmology from 2024 and 2021 on respectively and supervised my final undergraduate internship.

I would like to thank A. Lukov, F. Silva and A. Vats for discussions and advice in improving this book, K. Batrakov, M. Henderson, A. Lukov, P. Schlachter and F. Silva for proofreading drafts of this book and E. Szymańska for being the first person to have used this proto-book for a GR exam. I am also eternally indebted to Abhijeet Vats, under whose guidance I was able to develop my LaTeX skills to a satisfactory level. Without them, this book would undoubtedly not have been in its current form.

Finally, this book would not have been possible without the encouragement and support of my colleagues who are specialising or have specialised in gravitation and cosmology - S. López, M. Mishra, D. Petrov, P. Schlachter and F. Silva - with whom I have made many fruitful exchanges in physics and beyond.

1.2 How to use the book

The nature of GR as a mathematically demanding field is well-known, which is why the chapters preceding the Einstein field equations (or simply the *field equations*) will almost be entirely concerned with mathematics. Conversely, the same cannot be said about rigour[1]. The main reason for this increase in mathematical sophistication lies in the nature of GR as an expansion of the previously known special relativity into curved space. The study SR without much differential geometry is only possible because the curvature is zero, and the effects associated with it vanish.
With this in mind, this book consists of three parts:

- Part I resembles a standard 'General Relativity I' course in most universities, but with a few useful additions. This part requires an understanding of physics at a penultimate undergrad year level.
- Part II consists of introductions to more advanced topics from which one can migrate to more specialised references. This part requires Part I but is independent of Part III. For this part, the reader is encouraged to refer to more sophisticated references.
- Part III resembles a standard course in cosmology and an introduction in physical cosmology. This part requires Part I but is independent of Part II.

For any comments, suggestions or typos, please e-mail `zcapxix(at)ucl(dot)ac(dot)uk`.

[1]For a mathematician's text on GR, see *General Relativity: A Mathematical Introduction* by Christian Bär (Universität Potsdam).

Part I

Preliminaries and introduction

Chapter 2

Geometry and tensors

> **Quote 2.1** There is no difference between Time and any of the three dimensions of Space except that our consciousness moves along it.
>
> *H. G. Wells,* The Time Machine

In this chapter, we will begin with an introduction to manifolds and the many objects defined on them. Then, we will discuss the significance of a tensor, how it transforms and how one takes derivatives of it.

2.1 Before tensors: covariance and contravariance

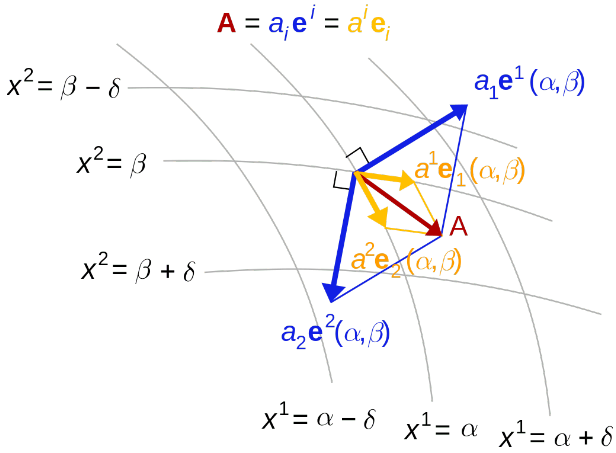


Figure 2.1: **Covariant** and **contravariant** components of the same vector A.

Before defining tensors, it is necessary to understand the *index notation* or the *Einstein summation convention*. At this point, you have probably seen a vector with an 'upstairs' upper index V^μ or with a 'downstairs'[1] lower index V_ν. The n^{th} component of these vectors are then V^n and V_n respectively. The positioning of the index is not arbitrary. An upper index denotes that the vector is *covariant*, and a lower index denoted that the vector is *contravariant*.

Definition 2.1 (Covariance and contravariance) Geometrically speaking, covariance and contravariance describe how the components of a vector are projected. When components are projected *perpendicular to* the axes, they are said to be covariant. When projected *along* the axes, they are said to be contravariant.

Remark 2.1 In flat space, the two projections yield identical results, which is why you have most likely not come across it beforehand.

[1]This expression is lovingly borrowed from Leonard Susskind's GR lectures.

Remark 2.2 Often, covariant and contravariant vectors are simply called a *dual vectors* and *vectors*.

Definition 2.2 (Index notation) As previously, seen, an arbitrary index is used as a stand-in for a range of indices. For n dimensions:

$$A^\mu B_\mu = \sum_{\mu=1}^{n} A^\mu B_\mu \tag{2.1}$$

One also have the following conventions:

- **Coordinate systems** are usually represented by capital letters X, Y, etc. Their axes are then represented by X^μ, X_μ, etc.
- **Partial derivatives** are often written in a shorthand that eliminates the symbol ∂.

$$\frac{\partial f}{\partial X^\mu} = \partial_\mu f = f_{,\mu} \tag{2.2}$$

We can, from the concept of a vector, generalise the concept of a *tensor*. Intuitively, it would be convenient to think of them in terms of *rank*.

Definition 2.3 (Rank) The rank of an object is the total number of its contravariant and covariant indices.

We can now consider some examples. A scalar is a single number without any indices, and thus has a rank of 0. A vector, which is a 1D array of numbers and has 1 index, has a rank of 1. A matrix, which is a 2D array and has 2 indices, has rank 2, and so on.

Note 2.1 A significant distinction must be made here. The rank, while corresponding to the 'dimensionality' of the array-like object, has no relation with the number of dimensions of the spacetime in which the object is encoded. As you will see later, a metric always has rank 2, and this is independent of the space it encodes. Instead, the dimension of spacetime is represented in the number of components each index holds. For example, a 2D metric has $2^2 = 4$ components: g_{00}, g_{01}, g_{10}, g_{11}, whereas a 4D metric have $4^2 = 16$ components from g_{00} to g_{33}.

Now imagine a 3D 'matrix', or an array of matrices. This object would thus have a rank of 3. The same applies for higher dimensions. It turns out that *many*[2] of these objects can be generalised as *tensors*. A scalar is thus a rank-0 tensor, a vector a rank-1 tensor, and so on.

Note 2.2 A common misconception is that tensors are essentially glorified matrices. However, it is worth noting that not all matrix-like objects are tensors. We will discuss this in detail in the sections.

2.2 Minimal introduction to manifolds

While of minimal[3] interest to physicists, it is nonetheless intuitively useful to understand manifolds. Again, the motivation stems from the fact that we deal with various curved spaces in GR, which no longer have convenient simplifications like flat spaces. We start with an open ball:

Definition 2.4 (Open ball) An *open ball* is the set of all points x in $\mathbf{R}^n$ such that $|x - y| < r$ for some fixed $y \in \mathbf{R}^n$ and $r \in \mathbf{R}$.

Remark 2.3 The open ball is the interior of an n-sphere of radius r centred at y.

Definition 2.5 (Open set) An *open set* is a set constructed from an arbitrary (maybe infinite) union of open balls. i.e. a set $V \subset \mathbf{R}^n$ is open if, for any $y \in V$, there is a open ball centred at y which is completely inside V.

Definition 2.6 (Chart) An *chart* or *coordinate system* consists of a subset U of a set M, along with a one-to-one map $\phi : U \to \mathbf{R}^n$ such that the image $\phi(U)$ is open in $\mathbf{R}$. We then can say that U is an open set in M.

[2]This is merely a sneak peek of tensors. We have not defined them here.

[3]Get it?

Definition 2.7 (Atlas) A C^∞ *atlas* is an indexed collection of charts $\{(U_\alpha, \phi_\alpha)\}$ which satisfies two conditions:

1. The union of the U_α is equal to M; that is, the U_α cover M.
2. The charts are smoothly sewn together. More precisely, if two charts overlap, $U_\alpha \cap U_\beta \neq \emptyset$, then the map $\left(\phi_\alpha \circ \phi_\beta^{-1}\right)$ takes points in $\phi_\beta\left(U_\alpha \cap U_\beta\right) \subset \mathbf{R}^n$ onto $\phi_\alpha\left(U_\alpha \cap U_\beta\right) \subset \mathbf{R}^n$, and all of these maps must be C^∞ where they are defined.

Remark 2.4 A chart is what we normally think of as a coordinate system on some open set, and an atlas is a system of charts which are smoothly related on their overlaps.

Definition 2.8 (Manifold) Finally, a *manifold* M generalises the idea of a surface or a space. It is a set M along with a 'maximal atlas', one that contains every possible compatible chart:

- M is a set of points which can be mapped into $\mathbb{R}, n \in N$, where n is called the dimension of the manifold.
- This mapping must be one-to-one.
- If two mappings overlap, one must be a differentiable function of the other.

We are now in a position to investigate, quoting Mitchell A. Berger, 'things that live on manifolds'.

Definition 2.9 (Scalar, vector and tensor fields) Scalar fields are functions which assign numbers to points on the manifold. Essentially, it is a map f from a manifold M to the set of real numbers:

$$f : M \to \mathbb{R} \tag{2.3}$$

Vector fields are functions which assign vectors to points on the manifold, and so on.

We then consider a *curve*, which essentially does the inverse, mapping a point on the manifold to a number[4]:

Definition 2.10 (Curve) A curve γ is defined by

$$\gamma : \mathbb{R} \to M \tag{2.4}$$

Remark 2.5 Notably, a curve does *not* transform as a tensor[5]. To derive its transformational properties and those of its derivatives, we need to investigate something that *does* transform tensorially.
We thus introduce *tangent vectors*:

Definition 2.11 (Tangent vector) The tangent vector to a curve X^μ represents the tangent direction on a specific point on a curve. It is given by

$$V^\mu = \frac{dX^\mu}{d\lambda} \tag{2.5}$$

2.3 General coordinate transformations

A topic of practical interest is *general coordinate transformations*, which is simply the transformation of objects from one coordinate system to another. As previously mentioned, we consider two charts $X^\mu = (X^1, \cdots, X^n)$ and $X'^\nu = (X'^1, \cdots, X'^m)$ and some vector V^μ in the coordinate system X^μ:

Theorem 2.1 (Vector transformations) A covariant vector undergoes general coordinate transformations as

$$V'^\nu = \frac{\partial X'^\nu}{\partial X^\mu} V^\mu \tag{2.6}$$

[4] This can be thought of as a 'constraint' that defines the curve.
[5] We will cover this shortly.

and a contravariant vector transforms as

$$V'_\nu = \frac{\partial X^\mu}{\partial X'^\nu} V_\mu \tag{2.7}$$

Remark 2.6 Note how the initial coordinate system's position in the fraction is opposite to that of its initial vector index.

For vectors, only one coordinate set is transformed, as seen above. Most tensors have more than one coordinate index, and coordinate transformations likewise rise in their sophistication. In fact, now that we have defined general coordinate transformations, we are in a position to finally define tensors.

Definition 2.12 (Tensor) A *tensor* is something that transforms as a tensor.

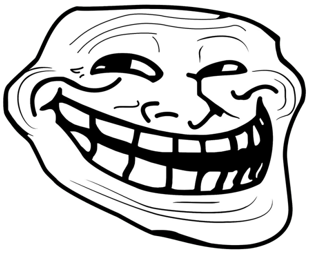


Figure 2.2: Ain't I a stinker?

Theorem 2.2 (Tensor, take 2) A tensor with rank $p+q$ undergoes general coordinate transformations as

$$T'^{\mu_1\cdots\mu_p}_{\nu_1\cdots\nu_q} = \frac{\partial X'^{\mu_1}}{\partial X^{\lambda_1}} \cdots \frac{\partial X'^{\mu_p}}{\partial X^{\lambda_p}} \frac{\partial X^{\sigma_1}}{\partial X'^{\nu_1}} \cdots \frac{\partial X^{\lambda_q}}{\partial X'^{\nu_q}} T^{\sigma_1\cdots\sigma_p}_{\lambda_1\cdots\lambda_q} \tag{2.8}$$

Remark 2.7 Here we see the significance of the ostensibly trollish Definition 2.12 and Note 2.2. If an object does not transform like the above, we can know for certain that it is *not* a tensor.

Exercise 2.1 Derive the tensor transformation properties from the vector transformation properties.

We have now arrived at a point where we can essentially derive the transformation properties of any arbitrary object, even if they do not transform as tensors.

- The general strategy is to first simplify the object such that its contents are transformed into as many tensors as possible. The end result should consist of tensors and maps, derivatives, operators, etc. on them.
- Then, instead of transforming the object as a whole (which may seem very alluring), we perform coordinate transformations to every tensor in it. The end result is then how the object transforms.

It is also worth remembering that with a parameter λ and some coordinate systems x^μ and X'^μ:

$$\frac{d}{d\lambda}\frac{\partial X'^\mu}{\partial X^\nu} = \frac{\partial^2 X'^\mu}{\partial X^\nu \partial X'^\sigma}\frac{dX^\sigma}{d\lambda} \tag{2.9}$$

In effect, we have 'invented' a new index σ to accommodate for the indexless total derivative with respect to the parameter. However, this is not a problem as the upper and lower σ 'cancel out', as we will see almost immediately.

Theorem 2.3 (Contraction) If a tensor has identical upper and lower indices, it can be *contracted*:

$$T^{\mu\nu\lambda\sigma}_{\alpha\beta\gamma\sigma} = T^{\mu\nu\lambda}_{\alpha\beta\gamma} \tag{2.10}$$

We note that, importantly, indices are always 'conserved' in a given equivalence. That is to say, after contraction, the indices on the LHS should be identical to those on the RHS. This means that some indices are arbitrary. They are known as *free indices*. Consider the following expression:

$$W^{\mu\nu} = T^{\mu\nu\lambda}_{\lambda} + T^{\mu\nu\sigma}_{\sigma} \tag{2.11}$$

The 'conserved' indices are clearly μ and ν. Hence, λ and σ are the free indices. Due to their arbitrary nature, one can relabel them to any symbol, including fixing one to another, which yields

$$W^{\mu\nu} = T^{\mu\nu\lambda}_{\lambda} + T^{\mu\nu\lambda}_{\lambda} = 2T^{\mu\nu} \tag{2.12}$$

While this may seem like cheating, it is legitimate, if not essential in many questions.
One important quality of tensors is symmetry or anti-symmetry, which we will now investigate.

Definition 2.13 (Symmetry and anti-symmetry) A tensor is *symmetric* if

$$T^{\mu\nu} = T^{\nu\mu} \tag{2.13}$$

and *anti-symmetric* or *skew-symmetric* if

$$T^{\mu\nu} = -T^{\nu\mu} \tag{2.14}$$

We can expand any tensor $T^{\mu\nu}$ into its symmetric and anti-symmetric parts:

$$T^{\mu\nu} = \underbrace{T^{(\mu\nu)}}_{\text{symmetric part}} + \underbrace{T^{[\mu\nu]}}_{\text{anti-symmetric parts}} \tag{2.15}$$

To prove symmetry, one merely needs to prove that $T^{[\mu\nu]} = 0$, and vice versa.
In much of GR literature, several peculiar shorthand notations, including () and [], are utilised. They are introduced here for the sake of completeness:

- **Antisymmetrisarion:** Also known as the *index commutator*, this operation returns an antisymmetric tensor:
$$g_{\nu[\lambda}R_{\sigma]\mu} = \frac{1}{2}(g_{\nu\lambda}R_{\sigma\mu} - g_{\nu\sigma}R_{\lambda\mu}) \tag{2.16}$$
- **Symmetrisation:** Also known as the *index anticommutator*, this operation returns a symmetric tensor:
$$g_{\nu(\lambda}R_{\sigma)\mu} = \frac{1}{2}(g_{\nu\lambda}R_{\sigma\mu} + g_{\nu\sigma}R_{\lambda\mu}) \tag{2.17}$$
- **Traceless part:** The traceless (or rather symmetric) part of a tensor $T_{\mu\nu}$ is expressed as $T_{\langle\mu\nu\rangle}$.

Note 2.3 In Equation 2.16 and Equation 2.17, the prefactors are not always $\frac{1}{2}$. While the numerator is always 1, the denominator corresponds to the the number of terms that symmetrisation/antisymmetrisation yields. For example:

$$M^{\{\lambda\mu\nu\}} = \frac{1}{6}\left(M^{\lambda\mu\nu} + M^{\lambda\nu\mu} + M^{\mu\lambda\nu} + M^{\mu\nu\lambda} + M^{\nu\lambda\mu} + M^{\nu\mu\lambda}\right) \tag{2.18}$$

2.4 Kronecker, Levi-Civita and the metric

We conclude the introduction to tensors with a few basic tensors which you may have seen before.

Definition 2.14 (Kronecker delta) The *Kronecker delta* is defined as

$$\delta^{\mu}_{\nu} = \begin{cases} 1 & \mu = \nu \\ 0 & \mu \neq \nu \end{cases} \tag{2.19}$$

It is invariant under transformations.

Fun fact 2.1 On Wikipedia, the title *Kronecker tensor* exists as a redirect of the Kronecker delta.

Note 2.4 The coordinate system and primes become important when we see that later on

$$\frac{\partial X'^{\mu}}{\partial X'^{\nu}} = \delta^{\mu}_{\nu} \quad \text{but} \quad \frac{\partial X^{\mu}}{\partial X'^{\nu}} \neq \delta^{\mu}_{\nu} \tag{2.20}$$

This is because in the latter, X^μ and X'^μ are in different coordinate systems, and we have no knowledge on whether they are orthogonal.

Definition 2.15 (Levi-Civita symbol) The *Levi-Civita symbol* is defined as

$$\epsilon_{\mu\nu\lambda} = \begin{cases} 0 & \mu = \lambda \text{ or } \nu = \lambda \text{ or } \lambda = \mu \\ +1 & \mu, \nu, \lambda \in \text{ permutation of } (1,2,3) \\ -1 & \mu, \nu, \lambda \notin \text{ permutation of } (1,2,3) \end{cases} \tag{2.21}$$

Remark 2.8 Physically, the Levi-Civita symbol defines the outer product. This is seen in the following:

$$(a \times b)_\mu = \epsilon_{\mu\nu\lambda} a_\nu b_\lambda \tag{2.22}$$

Definition 2.16 (Metric) The almighty *metric*[a] $g_{\mu\nu}$ is essential for inner products. Physically, it defines the line element:

$$ds^2 = g_{\mu\nu} dX^\mu dX^\nu \tag{2.23}$$

In n dimensions, a metric has n^2 components. However, only $\frac{n(n+1)}{2}$ components are independent due to symmetry.

[a] Or the *metric tensor* if you're boring at parties.

Remark 2.9 In a 4D space, a metric has 10 independent components, but one can always make 4 of them zero by choosing the correct coordinates.

Remark 2.10 In practice, the line element is merely one of the two ways of labelling metrics. See the 2-dimensional metric below as an example.

$$ds^2 = dx^2 + dy^2 \leftrightarrow g_{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{2.24}$$

Theorem 2.4 (Raising and lowering indices) One of the most common uses of metrics is to raise or lower indices:

$$T^\mu = g^{\mu\nu} T_\nu \quad T_\mu = g_{\mu\nu} T^\nu \tag{2.25}$$

But we note that $g^{\mu\nu}$, sometimes called the *inverse metric*, is undefined:

Definition 2.17 (Inverse metric) By definition, the inverse metric $g^{\mu\nu}$ satisfies

$$g_{\mu\nu} g^{\nu\lambda} = \delta^\lambda_\mu \quad g_{\mu\nu} g^{\mu\nu} = \delta^\mu_\mu = n \tag{2.26}$$

where n is the dimension of the metric.

Remark 2.11 As a metric has rank 2, one can solve for an inverse of any metric merely by solving for the inverse matrix. For diagonal metrics, we have $g^{\mu\nu} = 1/g_{\mu\nu}$ in every component.

Definition 2.18 (Singularity) A quantity becomes *singular* where it becomes poorly behaved (e.g. becomes infinite or zero). For a metric, conditions for singularity thus emerge when its determinant is zero. We have two types of singularities:

- **Curvature singularity:** The curvature becomes infinite (e.g. the centre of a black hole).
- **Coordinate singularity:** Singularities that can be eliminated by a change of coordinate. We will discuss this in detail in part II.

Remark 2.12 Mathematically, a zero determinant suggests that the line element ds^2 does not properly describe distances and times, causing the coordinate system to break down. Physically, the gravitational field becomes infinitely strong, leading to a breakdown of GR. See also initial singularity.

Definition 2.19 (Metric signature) The *signature* of a metric denotes its signs and has two common forms. Using the *Minkowski metric* in chapter 3 as an example:

- The number of positive and negative metric components. i.e. $(3, 1)$.

- The signs of the metric components. i.e. $(-,+,+,+)$

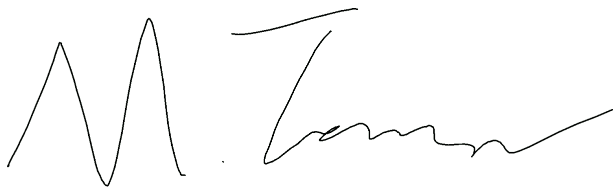


Figure 2.3: Signature of the metric.

Definition 2.20 (Norm) The *norm* of a tensor generalises the concept of *magnitude* in vectors. For a rank-1 tensor:

$$||V|| = \sqrt{g_{\mu\nu}V^{\mu}V^{\nu}} \tag{2.27}$$

For a rank-2 tensor:

$$||T|| = \sqrt{g_{\mu\lambda}g_{\nu\sigma}T^{\mu\nu}T^{\lambda\sigma}} \tag{2.28}$$

and so on.

2.5 Geodesics and classical mechanics

Definition 2.21 (Action) For a particle with 4-position x^{μ}, the *action* is defined as

$$S = \int \mathcal{L}(x^{\mu}, \dot{x}^{\mu})d^4x = \int L dx^0 = \int L d\lambda \tag{2.29}$$

where L is the Lagrangian, $\mathcal{L}$ is the *Lagrangian density*[a], and λ is the previously seen parameter which actually denotes proper time.

[a]Often also simply called the *Lagrangian*, although you will be able to tell the difference by looking at the notation.

Theorem 2.5 (Action principle) The *action principle* is simply another name for the *principle of stationary action*, which is itself often erroneously known as the *principle of least action*[a]. This simply means that the time derivative of the action of an isolated system is zero.

[a]This is because the principle states that instead of at a minimum, action tends to stay *stationary*, be it a maximum, a minimum or a saddle point.

Derivation 2.1 (Euler-Lagrange equations and the boundary term) To find the equations of motion via the action principle, we vary the action S given in Equation 2.29 with respect to x^{μ}, which involves integration by parts:

$$\delta S = \int d\lambda \left[\frac{\partial L}{\partial x^{\mu}}\delta x^{\mu} + \frac{\partial L}{\partial \dot{x}^{\mu}}\delta \dot{x}^{\mu}\right] \tag{2.30}$$

Noting that $\delta\dot{x}^{\mu} = \frac{d}{d\lambda}(\delta x^{\mu})$, we can write

$$\delta S = \int d\lambda \left[\frac{\partial L}{\partial x^{\mu}}\delta x^{\mu} + \frac{\partial L}{\partial(\frac{d}{d\lambda}x^{\mu})}\frac{d}{d\lambda}(\delta x^{\mu})\right] \tag{2.31}$$

We can apply integration by parts to the term involving $\partial\mu(\delta x^{\mu})$. The variation of the action is thus

$$\delta S = \int d\lambda \underbrace{\left(\frac{\partial L}{\partial x^{\mu}} - \frac{d}{d\lambda}\frac{\partial L}{\partial \dot{x}^{\mu}}\right)\delta x^{\mu}}_{①} + \int d\lambda \underbrace{\frac{d}{d\lambda}\left(\frac{\partial L}{\partial \dot{x}^{\mu}}\partial x^{\mu}\right)}_{②} \tag{2.32}$$

Through this process, we have exposed the so-called *boundary term* ②, which is a total derivative and does not contribute to the equations of motion. This is because it can be converted into a surface

integral over the boundary of the integration region using Gauss's law[a] In contrast, ① is known as the *bulk term*.
There are two scenarios in which the boundary term can be ignored:

- The position and its derivative vanishes at the boundary.
- The boundary extends into infinity.

Assuming the first point and applying the action principle leads to

Theorem 2.6 (Euler-Lagrange equations)

$$\frac{\partial L}{\partial x^\mu} - \frac{d}{d\lambda}\frac{\partial L}{\partial \dot{x}^\mu} = 0 \tag{2.33}$$

By recognising that $\frac{d}{d\lambda}$ is just the derivative over 4-coordinates or the *4-derivative*, we can see that this is the Euler-Lagrange equations that we have previously seen.

[a] As such, the boundary term is also called the *surface term*.

Before proceeding, we make a few comments on the bulk and boundary terms:

- The bulk term is so-called because it integrates over the entire volume of spacetime. When an action principle is imposed, the integrand of the bulk term vanishes, as seen in Equation 2.33.
- Hence, the bulk term reflects how the action responds to changes in the position x^μ in the 'bulk' or the 'interior' (i.e. everything minus the boundary) of the spacetime we consider. Under an action principle, the bulk term defines the allowed configurations for x^μ via the Euler-Lagrange equations.
- The boundary term reflects the influence of boundary conditions for the action. By imposing an action principle, we have also assumed that $\delta x^\mu = 0$ on the boundary - a boundary condition.
- This is why boundary conditions (like Dirichlet or Neumann conditions) are usually imposed in variational problems to ensure well-defined dynamics in the bulk term.

Now we consider *geodesics*. A geodesic is a curve representing in some sense the shortest path/arc between two points in a surface. In flat geometry, this is naturally a straight line. In simpler words, it is the 'straightest possible line' of a geometry. More rigorously:

Definition 2.22 (Geodesic) Particles that travel on geodesics satisfy the action principle.

Definition 2.23 (Geodesic types) In GR, geodesics can be classified into one of 3 types:

- **Spacelike**: This represents objects travelling faster than light. It satisfies

$$ds^2 \begin{cases} > 0 & \text{in the signature } (-,+,+,+) \\ < 0 & \text{in the signature } (+,-,-,-) \end{cases} \tag{2.34}$$

  For obvious reasons, spacelike geodesics do not represent any physical objects.

- **Timelike**: This represents objects travelling slower than light. It satisfies

$$ds^2 \begin{cases} < 0 & \text{in the signature } (-,+,+,+) \\ > 0 & \text{in the signature } (+,-,-,-) \end{cases} \tag{2.35}$$

  Timelike geodesics represent all objects travelling below the speed of light.

- **Null**: Also known as *lightlike*, this represents objects travelling exactly at the speed of light. It satisfies

$$ds^2 = 0 \tag{2.36}$$

  Lightlike geodesics represent photons and hypothetical gravitons, among other things.

At this point, we should discuss why the three geodesic types are so-called. The case for null or *lightlike* geodesics is intuitive as they travel with the speed of light. But what of the other two?

- *Spacelike geodesics* are so-called because the primary component in the separation along a timelike geodesic is spatial rather than temporal. i.e. the spatial component dominates.
- *Timelike geodesics* are so-called because the primary component in the separation along a timelike geodesic is temporal rather than spatial. i.e. the temporal component dominates.

Often geodesics are discussed in the context of a *light cone*, which provides a graphic representation of their physical nature.

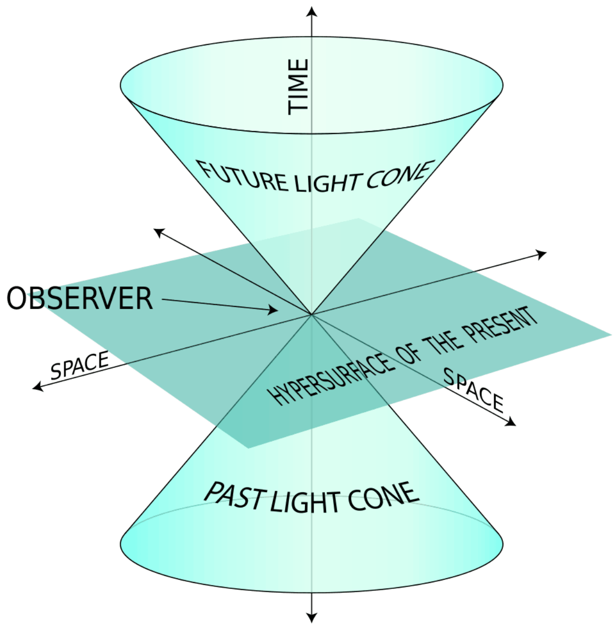


Figure 2.4: Visualisation of a light cone.

Remark 2.13 Spacelike geodesics are inside the light cone; timelike geodesics are outside the light cone; while lightlike geodesics are exactly on the light cone.

Remark 2.14 Here we see why the temporal and spatial components of the metric have different signs. This is to ensure that the ds^2 correctly distinguishes between time-like, space-like, and null intervals.

Derivation 2.2 (Geodesic equation) The action for a free particle is given by the integral of the spacetime interval (or proper time) along the path:

$$S = \int \sqrt{-g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu} d\lambda \tag{2.37}$$

where the square root ensures that the action corresponds to proper time (or proper length for spacelike paths). The corresponding Lagrangian is

$$L = \sqrt{g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}} \tag{2.38}$$

One can simplify this by replacing the action with:

$$S' = \int g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu d\lambda \tag{2.39}$$

which squares the proper time element. This also simplifies the Lagrangian to

$$L' = g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} \tag{2.40}$$

L' is proportional to the square of the proper time interval. Thus S' is proportional to S. The square root is a monotonic function, and its extremum coincides with the extremum of the squared term. Thus, both actions describe the same dynamics. One can verify this by proving that the Euler-Lagrange equations derived from S and S' are identical[a]. For the same reason that we will get the same Euler-Lagrange equations, we also invert the sign for simplicity.
An implication of this is that we can derive the motions of equation using this simplified scheme. We first transform the metric into the simplified Lagrangian in Equation 2.40
Remark 2.15 For example:

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 \leftrightarrow g_{\mu\nu} = L = -1\dot{t}^2 + 1\dot{x}^2 + 1\dot{y}^2 + 1\dot{z}^2 \tag{2.41}$$

By inserting Equation 2.40 into the Euler-Lagrange equations, we can derive the following expression

Theorem 2.7 (Geodesic equation) The *geodesic equation*, which some geodesic X^μ travels on, is

$$\frac{d^2X^\mu}{d\lambda^2} + \Gamma^\mu_{\nu\sigma}\frac{dX^\nu}{d\lambda}\frac{dX^\sigma}{d\lambda} = 0 \tag{2.42}$$

where $\Gamma^\mu_{\nu\sigma}$ is a cute little quantity known as the *Christoffel symbol*.

[a]This should be expected, as the proportionality does not affect the stationarity of the action.

Remark 2.16 The geodesic equation reflects the maximisation of proper time in GR.

2.6 Christoffel symbols, the qt 3.14s

Quote 2.2 Γ^a_{bc} is called Christoffel symbol and is of paramount interest in general relativity. The Christoffel symbol is called symbol because it is NOT a tensor. It does NOT transform like a tensor under general coordinate transformations.

Christian G. Böhmer, 2008

In the next section, we will discover that performing a derivative of a tensor is not as simple as merely performing a partial derivative on it. Instead, we perform *covarinat derivatives*, where our good friend, the Christoffel symbol, comes into play.

Definition 2.24 (Christoffel symbol)

$$\Gamma^\lambda_{\mu\nu} = \frac{1}{2}g^{\lambda\sigma}\left(g_{\sigma\mu,\nu} + g_{\nu\sigma,\mu} - g_{\mu\nu,\sigma}\right) \tag{2.43}$$

Sometimes, they are called *Christoffel connections* due to their nature as an *affine connection* to the metric in differential geometry.

Remark 2.17 Aren't you a little qt 3.14? Yes you are! Yes you are!

Quote 2.3 This little thing is quite a beast. (...) In an exam, I give you g, and I ask you to find Γ.

Christian G. Böhmer, on the Christoffel symbol, 2023

Exercise 2.2 Show that the Christoffel symbol transforms under general coordinate transformations as

$$\Gamma'^\mu_{\nu\lambda} = \frac{\partial X'^\mu}{\partial X^\lambda}\frac{\partial X^\sigma}{\partial X'^\nu}\frac{\partial X^\rho}{\partial X'^\lambda}\Gamma^\lambda_{\sigma\rho} + \frac{\partial^2 X^\rho}{\partial X'^\nu \partial X'^\lambda}\frac{\partial X'^\mu}{\partial X^\rho} \tag{2.44}$$

Fun fact 2.2 The Christoffel symbol we use is the *Christoffel symbols of the second kind.* There also exists a *Christoffel symbols of the first kind*, which is $\Gamma_{\mu\nu\lambda}$ and has full lower indices. Rather anticlimactically, they are created by simply applying the metric:

$$\Gamma_{\mu\nu\lambda} = g_{\mu\sigma}\Gamma^\sigma_{\mu\nu} \tag{2.45}$$

2.7 Covariant derivatives

> **Quote 2.4** Tell you what: forget covariant; let's just call it the shmovariant derivative.
>
> *Leonard Susskind, to Andre Cabannes, in* General Relativity: The Theoretical Minimum

There are many examples in physics where derivatives need to be extended:

- For example, we recall the infamous *Navier-Stokes equation*[6]:

$$\frac{D\vec{V}}{Dt} = \nabla p + v\nabla^2\vec{V} \tag{2.46}$$

  where $\frac{D}{Dt} = \frac{\partial}{\partial t} + \vec{V}\cdot\nabla$ is the material derivative.

- In quantum mechanics, the Laplacian ∇ in the Schrödinger equation is replaced by the *gauge covariant derivative* $\left(\nabla - \frac{iq}{\hbar c}\vec{A}\right)^2$ where $\vec{A}$ is the vector potential.

Likewise, in GR, you will find that the partial derivative of a tensor does not transform as a tensor. In fact, the whole definition of a partial derivative is problematic.

Derivation 2.3 (Transformation of partial derivatives) We take a simple example and transform the partial derivative of a simple vector A^m:

$$\begin{aligned}\frac{\partial A^\mu}{\partial X^\mu} &= \frac{\partial X^\lambda}{\partial X'^\nu}\frac{\partial}{\partial X^\lambda}\left(\frac{\partial X^\mu}{\partial X^\sigma}A^\sigma\right)\\ &= \frac{\partial X^\lambda}{\partial X'^\nu}\left(\frac{\partial^2 X'^\mu}{\partial X^\sigma\partial X^\lambda}A^\sigma + \frac{\partial X'^\mu}{\partial X^\sigma}\frac{\partial A^\sigma}{\partial X^\lambda}\right)\\ &= \frac{\partial X^\lambda}{\partial X'^\nu}\frac{\partial^2 X'^\mu}{\partial X^\sigma\partial X^\lambda}A^\sigma + \frac{\partial X^\lambda}{\partial X'^\nu}\frac{\partial X'^\mu}{\partial X^\sigma}\frac{\partial A^\sigma}{\partial X^\lambda}\end{aligned} \tag{2.47}$$

We note that the first term does not transform tensorially. As such, the partial derivative of a tensor is *not* a tensor.

This motivates a derivative that returns a tensor when given one. Thus, we need an extension for the derivative to recover the correct transformation. The result is the *covariant derivative*:

Definition 2.25 (Covariant derivative) The covariant derivative ∇ is defined as

$$\nabla_\mu T^\nu = \partial_\mu T^\nu + \Gamma^\nu_{\mu\lambda}T^\lambda \tag{2.48}$$

And here we see the significance of Christoffel symbols: to perform covariant derivatives.

Exercise 2.3 Show that, in Euclidian space, this just reduces to the good ol' partial derivative.

Theorem 2.8 (Covariant derivative properties) The covariant derivative has the following properties:

1. Linearity: for all $\alpha, \beta \in \mathbb{R}$
$$\nabla_\mu(\alpha A + \beta B) = \alpha\nabla_\mu A + \beta\nabla_\mu B \tag{2.49}$$
2. Leibnitz rule:
$$\nabla_\mu(AB) = B\nabla_\mu A + A\nabla_\mu B \tag{2.50}$$
3. Commutativity with contraction
$$\nabla_\mu A^{\nu_1\cdots\lambda\cdots\nu_m}_{\sigma_1\cdots\lambda\cdots\sigma_n} = \nabla_\mu A^{\nu_1\cdots\nu_m}_{\sigma_1\cdots\sigma_n} \tag{2.51}$$
4. Torsion free: for all smooth functions $f \in C^\infty(M)$ (for which case ∇ reduces to ∂)
$$\nabla_\mu\nabla_\nu f = \nabla_\nu\nabla_\mu f \tag{2.52}$$

[6] The solution of which, by the time of writing, is still worth $1m.

By using these properties and the covariant derivative's nature, we can derive the covariant derivatives of higher-rank tensors:

Derivation 2.4 (Covariant derivatives of higher-rank tensors) We observe, for example, $\nabla_a T_b$:

$$\nabla_\mu T_\nu = \partial_\mu T_\nu - \Gamma^\lambda_{\mu\nu} T_\lambda \tag{2.53}$$

Notice how the arbitrary index c went upwards? We therefore observe the almost slavish loyalty of the poor Christoffel symbol: for each index of the target tensor T which would become an arbitrary index, the Christoffel symbol sacrifices one of its indices, but in the opposite direction.
Hence, for contravariant rank-2 tensors, we have

$$\nabla_\lambda T^{\mu\nu} = \partial_\lambda T^{\mu\nu} + \Gamma^\mu_{\lambda\sigma} T^{\sigma\nu} + \Gamma^\nu_{\lambda\sigma} T^{\mu\sigma} \tag{2.54}$$

The other two forms we need to know then become trivial:

$$\nabla_\lambda T^\mu_\nu = \partial_\lambda T^\mu_\nu + \Gamma^\mu_{\lambda\sigma} T^\sigma_\nu - \Gamma^\sigma_{\lambda\mu} T^\mu_\sigma \tag{2.55}$$

$$\nabla_\lambda T_{\mu\nu} = \partial_\lambda T_{\mu\nu} - \Gamma^\sigma_{\lambda\mu} T_{\sigma\nu} - \Gamma^\sigma_{\lambda\nu} T_{\mu\sigma} \tag{2.56}$$

All of this is expected, given the summation indices created by the Christoffel symbols should cancel each other out. Note also that terms which have arbitrary indices assuming assuming contravariant positions *on the Christoffel* are negative.

Definition 2.26 (Compatatibility) In GR, the covariant derivative ∇ is always *metric compatible.* This means that

$$\nabla_\lambda g_{\mu\nu} = 0 \tag{2.57}$$

This ensures the length of a vector does not change as it is transported along a curve in the space.

Note 2.5 The covariant derivatives of the Kronecker delta and the Levi-Civita symbol are always zero.

Chapter 3

Special relativity revisited

Quote 3.1 You shouldn't be able to pump solids, but you can pump peanut butter.

Christian G. Böhmer, 22 November 2023

Before we continue differential geometry concerning GR, we need to wrap up a few loose ends in classical physics and SR. SR courses in most undergrad degrees are structured in a way that resembles the their historical derivation. While this makes sense in as introductions, we will find it far more intuitive to explain many underlying mechanisms using previously introduced concepts in tensors.

3.1 Electromagnetism

To begin with, let us dial the clock back a little bit, back to the days of James Clerk Maxwell. At this point, you are probably familiar with the scalar and vector potentials. The *magnetic vector potential* A is defined from the magnetic field B as

$$B_\mu = \nabla \times A_\mu \tag{3.1}$$

whereas the *electric scalar potential* ϕ is simply a relabelling of the voltage V, defined from the electric field E and the magnetic vector potential

$$E_\mu = -\nabla\phi - \partial_t A_\mu \tag{3.2}$$

Intuitively, this gives rise to a 4-vector called the *4-potential*:

Definition 3.1 (4-potential)

$$a^\mu = (\phi/c, A_i) = (\phi/c, A_x, A_y, A_z) \tag{3.3}$$

Another 4-vector quantity is the so-called *4-current density vector* or simply the *4-current.* We are already familiar with the *3-current* $J_i = (J_x, J_y, J_z)$. Combined with the charge density, this gives the 4-current

Definition 3.2 (4-current)

$$j^\mu = (c\rho, J_i) = (c\rho, J_x, J_y, J_z) \tag{3.4}$$

A tensorial representation of electromagnetic fields is the *Faraday tensor*, which we can use to derive Maxwell's equations.

Definition 3.3 (Faraday tensor) The *Faraday tensor* or the *electromagnetic tensor* is:

$$F^{\mu\nu} = \partial^\mu a^\nu - \partial_\nu a^\mu = \begin{pmatrix} 0 & E_x/c & E_y/c & E_z/c \\ -E_x/c & 0 & B_z & -B_y \\ -E_y/c & -B_z & 0 & B_x \\ -E_z/c & B_y & -B_x & 0 \end{pmatrix} \tag{3.5}$$

Derivation 3.1 (Maxwell's equations) At this point, it should come as no surprised that Maxwell's equations are not derived. Instead, they emerge naturally. We consider the more familiar version of Maxwell's equations:

- Structure equations:
$$\nabla \cdot \mathbf{B} = 0 \quad \nabla \times \mathbf{E} + \partial_t \mathbf{B} = 0 \tag{3.6}$$

- Source equations:
$$\nabla \cdot \mathbf{E} = 4\pi\rho \quad \nabla \times \mathbf{B} - \partial_t \mathbf{E} = \mathbf{J} \tag{3.7}$$

While we will not perform the derivation here, it is possible, by inspecting the structure of the Faraday tensor, to convert them to the following form:

Theorem 3.1 (Maxwell's equations) In tensorial form, Maxwell's equations are

- Structure equations:
$$\partial_\mu F_{\nu\lambda} + \partial_\lambda F_{\mu\nu} + \partial_\nu F_{\lambda\mu} = 0 \tag{3.8}$$

- Source equations:
$$\partial_\nu F^{\mu\nu} = -4\pi J^\mu \tag{3.9}$$

Exercise 3.1 Derive Equation 3.6 and Equation 3.7 by plugging the Faraday tensor into Equation 3.8 and Equation 3.7.

In SR conditions, the so-called *Lorenz gauge* is always satisfied. This is to eliminate an extra degree of freedom[1]:

Theorem 3.2 (Lorenz gauge)
$$\underbrace{\partial_\mu a^\mu = 0}_{\text{SR}} \quad \underbrace{\nabla_\mu a^\mu = 0}_{\text{GR}} \tag{3.10}$$

In the Lorenz gauge, we can then simplify the source equations as

$$\Box a^\mu = -\nu_0 j^\mu \tag{3.11}$$

where we have the *d'Alembertian* (or the *d'Alembert operator*):

Definition 3.4 (D'Alembertian)

$$\Box = \partial^a \partial_a = \eta^{\mu\nu} \partial_b \partial_a = \frac{\partial^2}{\partial t^2} - \frac{\partial^2}{\partial x^2} - \frac{\partial^2}{\partial y^2} - \frac{\partial^2}{\partial z^2} \tag{3.12}$$

Fun fact 3.1 The reason why it is denoted as a square remains a historical mystery.

3.2 Stress-energy tensor

A key element is the *stress-energy tensor* (or the *energy-momentum tensor*). It represents the matter and enegry content in GR:

[1] We will investigate this further in *Electron's Destiny*.

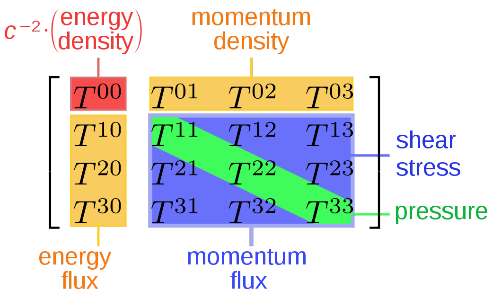


Figure 3.1: Components of the stress-energy tensor.

Remark 3.1 This density and flux of energy and momentum are the sources of the gravitational field in the field equations in GR, just as mass density is the source of such a field in Newtonian gravity.
If the energy content consists solely of an electromagnetic field, we can derive the stress-energy tensor an via the Faraday tensor:

Definition 3.5 (Stress-energy tensor in an EM field)

$$T_{\mu\nu} = \frac{1}{\mu_0}\left(F_{\mu\lambda}F^{\lambda}_{\nu} - \frac{1}{4}\eta_{\mu\nu}F_{\sigma\rho}F^{\sigma\rho}\right) \tag{3.13}$$

where, like c, μ_0 is generally regarded as 1. $\eta_{\mu\nu}$ is the metric in flat spacetime seen in Equation 3.20.

Remark 3.2 Physically, an EM field-dominated stress-energy tensor corresponds to conserved neutral currents in the 4-directions.

Exercise 3.2 Show that the trace of the stress-energy tensor of the electromagnetic field is zero.

If the matter-energy content is instead represented by a scalar field ϕ, the stress-energy tensor is instead

Definition 3.6 (Stress-energy tensor in a scalar field)

$$T_{\mu\nu} = -\partial_\mu\phi\partial_\nu\phi + \eta_{\mu\nu}\frac{1}{2}\left(\eta^{\lambda\sigma}\partial_\lambda\phi\partial_\sigma\phi - m^2\phi^2\right) \tag{3.14}$$

Remark 3.3 So far, aside from the Higgs field, which is a complex scalar field, no scalar fields have been observed in nature.
With this in mind, we can investigate perfect fluids, which will come in handy in cosmology.
One can begin by looking at a more realistic fluid. Assuming an incompressible, viscous fluid, for which we have the following stress-energy tensor:

Theorem 3.3 (Stress-energy tensor for incompressible, viscous fluids)

$$T^{\mu\nu} = \rho u^\mu u^\nu - (p - \xi\theta)(\eta^{\mu\nu} - u^\mu u^\nu) - 2\eta\sigma^{\mu\nu} \tag{3.15}$$

where we define the *shear tensor* $\sigma^{\mu\nu}$ as

$$\sigma^{\mu\nu} = \frac{1}{2}(u_{\mu,\lambda}h^{\lambda}_{\nu} + u_\nu h^{\lambda}_{\mu}) - \frac{1}{3}\theta h_{\mu\nu}) \tag{3.16}$$

u_a is a unit time-like vector representing the 4-velocity of the fluid, ρ is the energy density of the fluid, p is its pressure as measured in its rest frame, η is *shear viscosity coefficient*[a] and ξ is the *bulk viscosity coefficient*.

[a]Not to be confused with the determinant of the Minkowski metric.

Remark 3.4 Often we define the *projection tensor* $h_{\mu\nu}$ to simplify the expression:

$$h_{\mu\nu} = \eta_{\mu\nu} - u_\mu u_\nu \tag{3.17}$$

Definition 3.7 (Perfect fluids) Perfect fluids are idealised fluids with no viscosity or heat conduction.

Remark 3.5 As they have no viscosity, perfect fluids do not have surface tension or shear stresses either, although they may be compressible *or* incompressible.
Remark 3.6 Despite the lack of heat conduction, perfect fluids still have temperature. This temperature can change in various ways, such as adiabatic Processes, work done/gravitational effects on the perfect fluid (which can still compress or expand), interaction with external fields, radiation (which they can still emit or absorb), etc.
Assuming perfect fluids, we can simplify the stress-energy tensor to

Theorem 3.4 (Stress-energy tensor of perfect fluids) A perfect fluid has the following stress-energy tensor:

$$T_{\mu\nu} = \rho u_\mu u_\nu - p(\eta_{\mu\nu} - u_\mu u_\nu) \tag{3.18}$$

Theorem 3.5 (Local conservation law for perfect fluids) A perfect fluid satisfies the following equations of motion:

$$\partial^a T_{\mu\nu} = 0 \tag{3.19}$$

Remark 3.7 This will yield both the continuity equation and the relativistic Euler equation, which describes the conservation of momentum. You will see this in greater detail in Part III.

3.3 Special relativity and Lorentz boosts

In SR, we operate in Minkowski space with the line element $ds^2 = -dt^2 + dx^2 + dy^2 + dz^2$. Hence, we define the Minkowski metric:

Definition 3.8 (Minkowski metric)

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.20}$$

> **Quote 3.2** There are many, many, many physicists who like the metric to be $(1, 1, 1, -1)$. I use $(-1, 1, 1, 1)$. Why? I don't know. I guess it's because I am a *maverick*. (students laugh) And I warn you that almost everybody uses the other notation. I think Einstein used the same as I did, but I don't remember.
>
> *Leonard Susskind, on the metric, 17 November 2008*

Fun fact 3.2 In fact, in some notations, t is the 4^{th} coordinate instead.
Remark 3.8 For s to be proper time τ, the norm of the 4-position x^μ must observe $||X|| = 1$.

Definition 3.9 (Lorentz transformation) The *Lorentz transformation* or *Lorentz boost* Λ^ν_μ is defined as

$$x'^\nu = \Lambda^\nu_\mu x^\mu \tag{3.21}$$

Theorem 3.6 (Preservation of the Minkowski metric) The defining relation of the Lorentz transformation is the preservation of the Minkowski metric:

$$\Lambda^\mu_\nu \Lambda^\lambda_\sigma \eta_{\mu\lambda} = \eta_{\nu\sigma} \tag{3.22}$$

which simply returns to $\eta_{\mu\lambda}$ due to the indices being arbitrary.

Derivation 3.2 (Transformation matrix) Assume a spaceship with velocity v in the x-direction with respect to the Earth. We have the ship's rest frame S, Earth's rest frame E, and that their origins coincide at an event p (or rather, some *4-position* x^μ) p. We then have the transformation

$$\begin{pmatrix} t \\ \vec{x} \end{pmatrix}_S = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} t \\ \vec{x} \end{pmatrix}_E \tag{3.23}$$

Given that $c = 1$ in both frames, we consider a moving photon in both frames:

$$\begin{pmatrix} t \\ \vec{x} \end{pmatrix} = \begin{cases} \begin{pmatrix} t \\ t \end{pmatrix} & \text{moving to the right} \\ \begin{pmatrix} t \\ -t \end{pmatrix} & \text{moving to the left} \end{cases} \tag{3.24}$$

Now we assign the event $\begin{pmatrix} t_E \\ t_E \end{pmatrix}$ (for the right-moving photon) or $\begin{pmatrix} t_E \\ -t_E \end{pmatrix}$ (for the left-moving photon) in the Earth's rest frame. Plugging in these coordinates in the transformation, and recalling that *the event coincides in both frames* yield

$$\alpha + \beta = \gamma + \delta \tag{3.25}$$

for the right-moving photon and

$$\alpha + \beta = \gamma - \delta \tag{3.26}$$

for the left-moving photon. Combining the equations yields

$$\alpha = \delta \quad \beta = \gamma \to \begin{pmatrix} t \\ \vec{x} \end{pmatrix}_S = \begin{pmatrix} \gamma & \delta \\ \delta & \gamma \end{pmatrix} \begin{pmatrix} t \\ \vec{x} \end{pmatrix}_E \tag{3.27}$$

However, we still need to determine what the values of γ and δ are. Again we consider both reference frames. In the spaceship, we have

$$\begin{pmatrix} t \\ \vec{x} \end{pmatrix}_S = \begin{pmatrix} t \\ 0 \end{pmatrix}_S \tag{3.28}$$

Quote 3.3 But, Earthlings see this move at speed V[a]

Mitchell A. Berger, confirming that he is an alien, 2004

[a]Denoted here by $\vec{v}$.

As such we have

$$\begin{pmatrix} t \\ \vec{x} \end{pmatrix}_E = \begin{pmatrix} t \\ \vec{v}t \end{pmatrix}_E \tag{3.29}$$

where $\vec{v}$ is the 3-velocity. Plugging both in, and we get

$$\gamma = -\delta\vec{v} \tag{3.30}$$

Thus, we have

$$\begin{pmatrix} t \\ \vec{x} \end{pmatrix}_S = \begin{pmatrix} \gamma & -\vec{v}\gamma \\ -\vec{v}\gamma & \gamma \end{pmatrix} \begin{pmatrix} t \\ \vec{x} \end{pmatrix}_E \tag{3.31}$$

and the transformation matrix is

Definition 3.10 (Lorentz transformation in matrix form)

$$\lambda^\mu_\nu = \begin{pmatrix} \gamma & -\vec{v}\gamma \\ -\vec{v}\gamma & \gamma \end{pmatrix} \tag{3.32}$$

We can thus find the so-called *inverse Lorentz transform* matrix:

$$(\lambda^\mu_\nu)^{-1} = \begin{pmatrix} \gamma & \vec{v}\gamma \\ \vec{v}\gamma & \gamma \end{pmatrix} \tag{3.33}$$

But what is γ? We know from the invariance of the proper distance that the determinant of the boost matrix is 1. Solving for the determinant equation yields

Definition 3.11 (Lorentz factor)

$$\gamma = \frac{1}{\sqrt{1 - \frac{\vec{v}^2}{c^2}}} \tag{3.34}$$

The Lorentz factor is the source of many SR effects we know and love, such as tile dilation and length contraction. We conclude with a comment on the latter:

> **Quote 3.4** Consider a metre stick at rest on the spaceship; The space travellers measure the position of the ends of the stick simultaneously at P, R. Earthlings see P, Q as simultaneous events corresponding to the ends of the stick at $t_E = 0$.
>
> *Mitchell A. Berger, once again betraying his alien loyalties, 2004*

3.4 Relativistic dynamics

In SR kinematics, energy and momentum conservation are still observed, but their definition changes drastically.

Definition 3.12 (Relativistic energy) Taking into account mass-energy, the *relativistic energy* observes

$$E^2 = \underbrace{p^2c^2}_{\text{kinetic energy}} + \underbrace{m^2c^4}_{\text{rest energy}} \tag{3.35}$$

Remark 3.9 Despite $E = mc^2$, the energy is nonetheless conserved because we now take into account the *rest energy*.
Relativistic dynamics is a very practical branch of SR. For example, high energy and accelerator physics in general often use them to investigate particle dynamics. We thus usually take into account the speed of light. From this, we can recover the speed of light from the Lorentz transformation matrix and thus the 'recovered' the Lorentz transformation matrix:

$$\Lambda^\mu_\nu = \begin{pmatrix} \gamma & -\beta\gamma \\ -\beta\gamma & \gamma \end{pmatrix} \tag{3.36}$$

where $\beta = \frac{v}{c}$. The recovered inverse Lorentz transformation matrix follows.

Definition 3.13 (4-momentum) The *4-momentum* is

$$P^\mu = (E, p^i) = (E/c, p_x, p_y, p_z) \tag{3.37}$$

where p^i is the so-called *3-momentum*.

In the last section, the Lorentz transformation's actions on the 4-position were considered. But:

- The 4-velocity is the proper time-derivative of the 4-position. i.e. a tangent vector $\frac{dx^\mu}{d\tau}$
- The 4-momentum is the 4-velocity multiplied by mass. i.e. $\frac{d^2x^\mu}{d\tau^2}$

It would thus be intuitive that 4-momentum also observes Lorentz transformations. And indeed, it does. This yields the boosted 4-momentum:

$$P' = (\gamma mc, \beta\gamma mc) \tag{3.38}$$

By inspecting the 4-momenta before and after, it can be found that

Theorem 3.7 (4-momentum conservation) The inner product of a four-momentum with itself $P \cdot P$ is always conserved under a change of reference frame.

This corresponds to the invariance of proper time.

Chapter 4

Curvature

Quote 4.1 You can spend an hour staring at them trying to figure out these pictures.

Christian G. Böhmer, on Circle Limit III*, 15 November 2023*

Quote 4.2 those fishies are doing such a great job at tiling the hyperbolic plane

Paulina Schlachter, on Circle Limit III*, 3 October 2024*

In this chapter, the concept of curvature is finally introduced. We will investigate various curvature tensors and their properties, which will allow us to understand the geometrical mechanisms of curved spacetime.

4.1 Parallel transport

Before we introduce curvature, it would be intuitive to first discuss parallel transport.

Definition 4.1 (Parallel transport) A vector V^μ at each point on some curve γ is said to be parallelly transported along the curve if

$$T^\mu \nabla_\mu V^\nu = 0 \tag{4.1}$$

where T^μ is the tangent vector.

Note 4.1 (Deriving tangent vectors) Consider latitude curves on the 2-sphere θ, ϕ with two points on it (θ_1, ϕ), (θ_2, ϕ):

$$X^\mu(\lambda) = (\theta_1 + \lambda(\theta_2 - \theta_1), \phi) \tag{4.2}$$

where λ is some parameter. The tangent vector is thus

$$T^\nu = (\theta_2 - \theta_1, 0) \tag{4.3}$$

Physically, parallel transport denoted the transport of a vector along a curve with respect to its tangent vector. It is a generalisation of translation in 3D flat space. Unlike in 3D flat space, however, we find that the vector has 'tilted' after parallel transport.

Remark 4.1 Parallel transport of tangent vectors is *autoparallel*. This gives rise to geodesics, on which tangent vectors are always parallely transported.

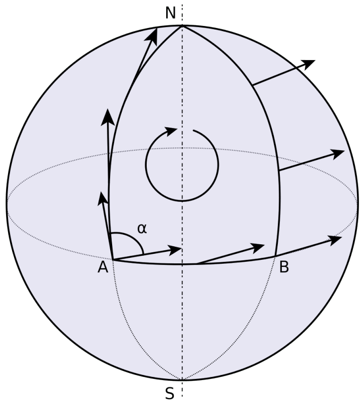


Figure 4.1: Tilting of the vector after parallel transport.

It is then necessary to introduce the concept of curvature.

4.2 Nonmetricity, torsion, or the lack of them in GR

Before actually going into GR, we might find it convenient in defining what GR is not.

Definition 4.2 (Nonmetricity tensor) The *nonmetricity tensor* is the covariant derivative of the metric tensor:

$$Q_{\mu\nu\lambda} = \nabla_\mu g_{\nu\lambda} \tag{4.4}$$

It measures the rate of change of the components of the metric tensor along the flow of a given vector field.

Sadly, it vanishes in Riemannian geometry and can be used to study non-Riemannian spacetimes.

Note 4.2 As such, the covariant derivative of Riemannian metrics is zero!

Previously, we mentioned that covariant derivatives are torsion-free. But what does it mean?

Definition 4.3 (Torsion tensor) The *torsion tensor* can be associated with any given connection:

$$T^\lambda_{\mu\nu} = \Gamma^\lambda_{\mu\nu} - \Gamma^\lambda_{\nu\mu} = 2\Gamma^\lambda_{[\mu\nu]} \tag{4.5}$$

which is antisymmetric in its lower indices.

Definition 4.4 (Torsion-free) A connection which is symmetric in its lower indices is known as 'torsion-free'.

Remark 4.2 In GR, we assume manifolds to be torsion-free. This explains why all Christoffel symbols we see in GR are symmetric in their lower indices. Although in some modified gravity theories, such as the Einstein-Cartan theory, this assumption is dropped[1].
As such, we regrettably have to say goodbye to our dear friends, the nonmetricity and torsion tensors.

4.3 Riemann tensor, Ricci tensor and scalar

Now we can finally investigate the curvature tensors, which lie in the heart of GR. Among them, the *Riemann tensor* holds the entirety of the information on curvature, and all other curvature tensors are 'reduced' versions of it. It is defined by a series of Christoffel symbols:

Definition 4.5 (Riemann tensor)

$$R^\sigma_{\mu\nu\lambda} = \Gamma^\sigma_{\mu\lambda,\nu} - \Gamma^\sigma_{\nu\lambda,\mu} + \Gamma^\sigma_{\nu\sigma}\Gamma^\rho_{\rho\nu} - \Gamma^\rho_{\mu\sigma}\Gamma^\sigma_{\rho\mu} \tag{4.6}$$

[1] In fact, nonmetricity and torsion are both staples of modified gravity theories.

In n dimensions, a Riemann tensor has n^4 components. However, only $\frac{n^2(n^2-1)}{12}$ components are independent due to symmetry.

Theorem 4.1 (Riemann tensor properties) The Riemann tensor has the following properties:

1. **Antisymmetry I:**
$$R_{\mu\nu\lambda\sigma} = -R_{\nu\mu\lambda\sigma} \tag{4.7}$$

2. **Antisymmetry II:**
$$R_{\mu\nu\lambda\sigma} = -R_{\mu\nu\sigma\lambda} \tag{4.8}$$

3. **Cyclic identity:**
$$R_{\sigma\mu\nu\lambda} + R_{\lambda\sigma\mu\nu} + R_{\sigma\nu\lambda\mu} = 0 \tag{4.9}$$
This is the (in)famous *first Bianchi identity*.

4. **Symmetry:**
$$R_{\mu\nu\lambda\sigma} = R_{\lambda\sigma\mu\nu} \tag{4.10}$$

5. **Kretschmann scalar:** An invariant quality is the *Kretschmann scalar* K, defined by
$$K = R^{\mu\nu\lambda\sigma}R_{\mu\nu\lambda\sigma} \tag{4.11}$$

Remark 4.3 From Equation 4.7, Equation 4.8 and Equation 4.9, we can prove Equation 4.10 by permuting the indices of Equation 4.9 four times and adding them together.

Exercise 4.1 Do it.

Remark 4.4 The Riemann tensor can be thought of as a symmetric matrix of bivectors, where bivectors are antisymmetric rank-2 tensors (forms that can be written as $\omega_{\mu\nu} = -\omega_{\nu\mu}$).
While the Riemann tensor holds the most complete information on curvature, it is arguably not the most useful one. It does not appear in the field equations, instead, a reduced version of it known as the *Ricci tensor* does. The Ricci tensor is acquired by contracting two indices in the Riemann tensor, and thus contains roughly half the information encoded in the Riemann tensor.

Definition 4.6 (Ricci tensor)
$$R_{\mu\nu} = R^{\lambda}_{\mu\lambda\nu} \tag{4.12}$$
Just like the metric[a], in n dimensions, a Ricci tensor has n^2 components. However, only $\frac{n(n+1)}{2}$ components are independent due to symmetry.

[a] This is expected, as both have rank 2.

Exercise 4.2 Determine the number of independent components of the Riemann tensor and the Ricci tensor in good ol' 4 dimensions.

The *Ricci scalar* is a even further reduced form of the Riemann tensor, and thus contains the least amount of information:

Definition 4.7 (Ricci scalar)
$$R = R^{\mu}_{\mu} \tag{4.13}$$
It is also known as the *scalar curvature*, which provides a better picture of its physical significance. Again, its significance in GR is seen in its use in the field equations.

Remark 4.5 In practice, we derive the Ricci scalar via applying the almighty inverse metric on the Ricci tensor:
$$R = g^{\mu\nu}R_{\mu\nu} \tag{4.14}$$

Again we consider the significance of these tensors. How do they link to the concept of parallel transport? Previously, we have established that curvature is the 'tilting' of a vector under parallel transport. Hence, we cal also interpret the Riemann tensor, the Ricci tensor and the Ricci scalar in the same vein. They represent the failure of parallelism in a curved space.

Figure 4.2: Barclay and Einstein discuss '10 [independent] components of the curvature tensor' in 'The N$^{\text{th}}$ Degree'. Note also '$g_{\mu\nu}$ = metric tensor' and '$U_{ru} + S_e + I = Y_a + T_s + U_r + A$'.

Derivation 4.1 (2D Riemann and Ricci tensors) In a 2-dimensional space, an index can only assume 2 values - 0 or 1. Thus the Riemann tensor has only one independent component R^1_{010} or R_{0101}. Every other component is related to this component through the Riemann tensor properties. The composition of the Riemann tensor can be simplified:

$$R_{\mu\nu\lambda\sigma} = K(g_{\mu\lambda}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\lambda}) \tag{4.15}$$

Remark 4.6 Likewise, the Ricci tensor has only one independent component. *This turns out to be a scalar quantity which is the Ricci scalar.*
In this case, the Ricci scalar is equivalent to the Gaussian curvature K, and the Ricci tensor can be represented as

$$R_{\mu\nu} = \frac{1}{2}Kg_{\mu\nu} \tag{4.16}$$

Remark 4.7 The Gaussian curvature essentially measures how much the surface bends locally. In classical differential geometry, it is the product of the principal curvatures. By applying $g^{\mu\nu}$ on both side, one can conclude that the Gaussian curvature is R.

And thus completes our long trek to curvature and GR.

4.4 Weyl tensor, its demise and conformal curvature

Finally, with a heavy heart, we discuss the Weyl tensor. The Riemann tensor can be broken up into 3 rank-4 tensors. This is known as the *Ricci decomposition*:

Theorem 4.2 (Ricci decomposition)

$$R_{\mu\nu\lambda\sigma} = S_{\mu\nu\lambda\sigma} + E_{\mu\nu\lambda\sigma} + C_{\mu\nu\lambda\sigma} \tag{4.17}$$

The first 2 terms are objects with no name defined as

$$S_{\mu\nu\lambda\sigma} = \frac{R}{n(n-1)}(g_{\mu\sigma}g_{\nu\lambda} - g_{\mu\lambda}g_{\nu\sigma}) \tag{4.18}$$

$$E_{\mu\nu\lambda\sigma} = \frac{1}{n-2}(Z_{\mu\sigma}g_{\nu\lambda} - Z_{\nu\sigma}g_{\mu\lambda} - Z_{\mu\lambda}g_{\nu\sigma} + Z_{\nu\lambda}g_{\mu\sigma}) \tag{4.19}$$

where, for a manifold of dimension n:

$$Z_{\mu\nu} = R_{\mu\nu} - \frac{1}{n}Rg_{\mu\nu} \tag{4.20}$$

The final term $C_{\mu\nu\lambda\sigma}$ is the ill-fated *Weyl tensor*, which is defined as

Definition 4.8 (Weyl tensor)

$$C_{\mu\nu\lambda\sigma} = R_{\mu\nu\lambda\sigma} + \frac{1}{n-2}\left(R_{\mu\sigma}g_{\nu\lambda} - R_{\mu\lambda}g_{\nu\sigma} + R_{\nu\lambda}g_{\mu\sigma} - R_{\nu\sigma}g_{\mu\lambda}\right) + \frac{1}{(n-1)(n-2)}R\left(g_{\mu\lambda}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\lambda}\right) \tag{4.21}$$

where n is the dimension of the manifold. One can simplify this using the index commutator seen in Equation 2.16:

$$C_{\mu\nu\lambda\sigma} = R_{\mu\nu\lambda\sigma} - \frac{2}{n-2}\left(g_{\mu[\lambda}R_{\sigma]\nu} - g_{\nu[\lambda}R_{\sigma]\mu}\right) + \frac{2}{(n-1)(n-2)}Rg_{\mu[\lambda}g_{\sigma]\nu} \tag{4.22}$$

Just like the Riemann tensor, in n dimensions, a Weyl tensor has n^4 components. However, only $\frac{n(n+1)(n+2)(n-3)}{12}$ components are independent due to symmetry.

The Weyl tensor is also called the *conformal tensor* because it is invariant under conformal transformations of the metric.
With respect to the Weyl tensor, we note the following points:

- It is what remains when the Ricci part, associated with matter content, is 'subtracted' from the Riemann tensor. i.e. It is essentially the trace-free/anti-symmetric part of the Riemann tensor.

- It encodes the free gravitational field independent of matter and determines 10 components of the Riemann tensor - those not directly determined by matter and energy content (via the field equations).

- In other words, while the Riemann tensor represents the total curvature, and the Ricci tensor represents curvature generated by matter, the Weyl tensor represents curvature generated by non-matter sources: gravitational waves and tidal forces. We will see this in greater detail later.

- As a result, in vacuum, the Weyl tensor is equal to the Riemann tensor. This is significant in that (as we will later see) for the vacuum solution, the Weyl tensor *is* curvature.

Theorem 4.3 (Weyl tensor contractions) Every contraction between indices in Weyl tensors gives 0.

Unfortunately, despite being introduced to UCL GR course in 2007, the Weyl tensor was removed from the GR curriculum at some point around 2008.

> **Quote 4.3** The Weyl tensor is a bit tangential. It is nice to work with. (...) If I had 10 more hours in my lectures, I would have covered it.
>
> *Christian G. Böhmer, 23 February 2024*

Remark 4.8 The rotation and distortion of the Weyl tensor (or rather the *Weyl fields*) bring about and characterise gravitational waves. In linearised gravity, which assumes vacuum field equations, the components of the Weyl tensor describe the tidal forces that would be felt by a ring of test particles as the wave passes through.

Figure 4.3: Demise of Weyl Tensor.

Finally, we can see quite easily that the Weyl tensor is the *conformally invariant* equivalent of the Riemann tensor. From this we consider the question: is there also a tensor that acts as the conformally invariant form of the Ricci tensor? As it turns out, the answer is yes. For the Ricci tensor, there exists a *Bach tensor*[2] which fulfils its role in conformally invariant cases.
To derive it, we introduce an intermediary quantity called the *Schouten tensor*:

Definition 4.9 (Schouten tensor)

$$P_{\mu\nu} = \frac{1}{n-2}\left(R_{\mu\nu} - \frac{R}{2(n-1)}g_{\mu\nu}\right) \tag{4.23}$$

where n is again the number of dimensions.

The tensor itself is not very meaningful[3], but it does help us derive other interesting qualities. As it turns out, the Bach tensor can be represented in terms of the Schouten tensor.

Definition 4.10 (Bach tensor)

$$B_{\mu\nu} = P_{\lambda\sigma}W^{\lambda\sigma}_{\mu\nu} + \nabla^{\lambda}\nabla_{\lambda}P_{\mu\nu} - \nabla^{\lambda}\nabla_{\mu}P_{\nu\lambda} \tag{4.24}$$

This is essentially a measure of the deviation from conformal flatness. i.e. conformal curvature.

4.5 Physical significance of curvatures

In innocent flat spaces, there is no curvature, and geodesics go about their lives quite happily. However, with curvature, we need to consider how geodesics develop differently. Hence, we introduce the *geodesic deviation equation.*

Derivation 4.2 (Geodesic deviation equation) Our object of interest is a two-parameter *family of geodesics* $X^\mu(\lambda, s)$, where the good ol' tangent vector to a geodesic remains $T^\mu = \frac{dX^\mu}{d\lambda}$.
The the so-called *deviation vector* or *displacement vector*, which illustrated the displacement of one geodesic with respect to a nearby geodesic, is defined as

$$N^\mu = \frac{dX^\mu}{ds} \tag{4.25}$$

Now we can attempt to associate T^μ and N^μ. Using

$$\frac{d^2X^\mu}{d\lambda ds} = \frac{d^2X^\mu}{dsd\lambda} \tag{4.26}$$

[2]Named after the little-known Rudolf Bach.
[3]That is to say, it arises naturally. Why it is named at all is a mystery of itself.

we can show that

$$N^\mu \frac{\partial T^\mu}{\partial x^\mu} = T^\mu \frac{\partial N^\mu}{\partial x^\mu} \tag{4.27}$$

and

$$N^\mu \nabla_\mu T^\nu = T^\mu \nabla_\mu N^\nu \tag{4.28}$$

Physically, $T^\mu \nabla_\mu N^a$ is the relative velocity of a nearby geodesic, as it gives the rate of change along a geodesic of the displacement to a nearby geodesic.
We can apply $T^\lambda \nabla_\lambda$ to both sides of Equation 4.28:

$$T^\lambda \nabla_\lambda \left(T^\mu \nabla_\mu N^\nu\right) = T^\lambda \nabla_\lambda \left(N^\mu \nabla_\mu T^\nu\right) \tag{4.29}$$

The LHS denotes the relative acceleration of a nearby geodesic. It is this quantity which will be related to curvature.
We can further reduce this equation:

$$\begin{aligned} T^\lambda \nabla_\lambda \left(T^\mu \nabla_\mu N^\nu\right) &= T^\lambda \left(\nabla_\lambda N^\mu\right) \nabla_\mu T^\nu + T^\lambda N^\mu \nabla_\lambda \nabla_\mu T^\nu \\ &= \underbrace{T^\lambda \left(\nabla_\lambda N^\mu\right) \nabla_\mu T^\nu + T^\lambda N^\mu \nabla_\mu \nabla_\lambda T^\mu}_{①} - T^\lambda N^\mu R^\lambda_{\nu\mu\sigma} T^\sigma \end{aligned} \tag{4.30}$$

We can show that ① vanishes:

$$\begin{aligned} T^\lambda \left(\nabla_\lambda N^\mu\right) \nabla_\mu T^\nu + T^\lambda N^\mu \nabla_\mu \nabla_\lambda T^\nu &= N^\lambda \left(\nabla_\lambda T^\mu\right) \nabla_\mu T^\mu + T^\lambda N^\mu \nabla_\mu \nabla_\lambda T^\nu \\ &= N^\lambda \left(\nabla_\lambda T^\mu\right) \nabla_\mu T^\nu + T^\mu N^\lambda \nabla_\lambda \nabla_\mu T^\nu \\ &= N^\lambda \left(\nabla_\lambda T^\mu \nabla_\mu T^\nu + T^\mu \nabla_\lambda \nabla_\mu T^a\right) \\ &= N^\lambda \nabla_\lambda \left(T^\mu \nabla_\mu T^\nu\right) \\ &= 0 \end{aligned} \tag{4.31}$$

which yields the final result

$$T^\lambda \nabla_\lambda \left(T^\mu \nabla_\mu N^\nu\right) = -T^\lambda N^\mu R^\nu_{\lambda\mu\sigma} T^\sigma = \left(R^\nu_{\mu\lambda\sigma} T^\lambda T^\sigma\right) N^\mu \tag{4.32}$$

By using the notation $\frac{D}{D\lambda} = T^\mu \nabla_\mu$, which represents *covariant derivative along the curve parameterised by* λ, one can write down the geodesic deviation equation as it is commonly seen:

Theorem 4.4 (Geodesic deviation equation) For the deviation vector N and tangent vector T of a geodesic

$$\frac{D^2 N^\mu}{D\lambda^2} = \left(R^\nu_{\mu\lambda\sigma} T^\lambda T^\sigma\right) N^\mu \tag{4.33}$$

where, importantly, T observes

$$T_\mu T^\mu = 1 \quad g_{\mu\nu} T^\mu N^\nu = 0 \tag{4.34}$$

Remark 4.9 $T_\mu T^\mu = 1$ implies that T^μ is a unit vector with respect to the given metric. $g_{\mu\nu} T^\mu N^\nu = 0$ implies that T^μ and N^ν are orthogonal. N^ν can be derived from T^μ using the inner product.
This all looks a bit bulky. So we stop for a moment and look at the physical significance of this:

- The deviation vector describes how the separation between two nearby geodesics changes under the effects of curvature as they move along their paths.
- Consider a curved spacetime, and one will conclude that it inevitably leads to the existence of a Riemann tensor. This gives rise to the so-called *tidal forces*.
- Expectedly, in a vacuum solution, the Weyl tensor is solely responsible for encoding the tidal forces.

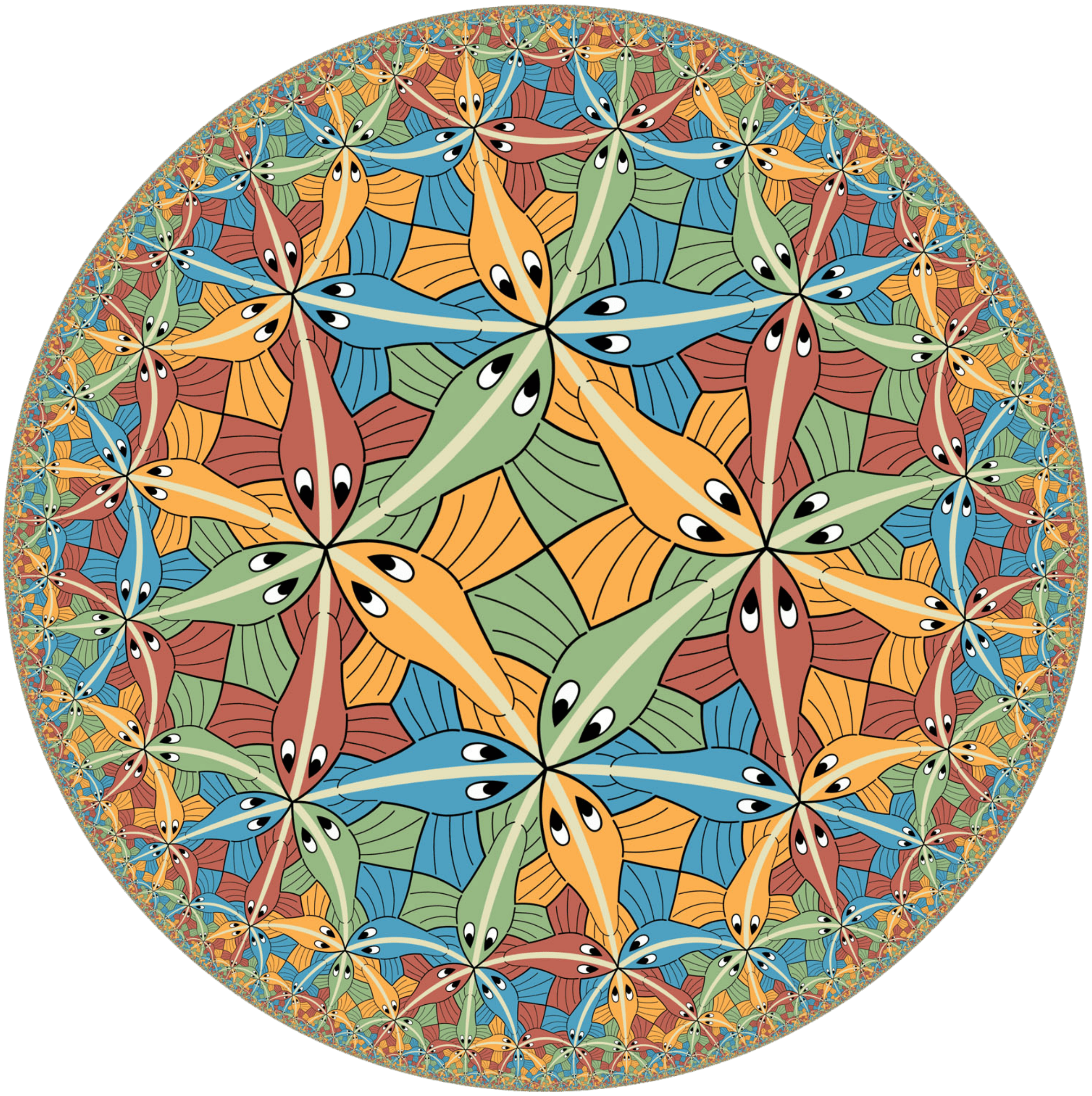


Figure 4.4: *Circle Limit III* (M. C. Escher, 1959)

Chapter 5

Field equations and the Schwarzschild solution

Quote 5.1 Space tells matter how to move
Matter tells space how to curve

John Archibald Wheeler, in Gravitation*, 1973*

In this chapter, we finally look at the field equations, the centrepiece of GR. In most introductory GR literature, the field equations are simply given as-is. In this book, we will instead derive it from the Einstein-Hilbert action like real relativists. Afterwards, we will investigate in depth the Schwarzschild solution, which is one of the simplest exact solutions of the field equations.

5.1 Einstein-Hilbert action

We recall the action in terms of the Lagrangian density in Equation 2.29. In GR, one can decompose the action into two terms:

Definition 5.1 (Einstein-Hilbert action)

$$S = \int_{\mathcal{V}} \mathcal{L} d^4x = S_H + S_M \tag{5.1}$$

where we have integrated over a region $\mathcal{V}$ of the manifold.

Remark 5.1 An important plot twist, already foreshadowed, reveals itself here: Equation 2.38 is a very limiting case (applying to only free particles).
The simplest gravitational action was proposed by Hilbert and Einstein in 1915:

Definition 5.2 (Hilbert term) The *Hilbert term* S_H is defined as

$$S_H = \frac{1}{16\pi} \int_{\mathcal{V}} R\sqrt{-g} d^4x \tag{5.2}$$

where g is the (negative) determinant[a] of the metric g_{ij} and $\sqrt{-g}d^4x$ is the proper volume element. The $1/16\pi$ term will become significant soon.

[a] *Not* the trace!

The Hilbert Lagrangian (density) is expectedly

$$\mathcal{L}_H = R\sqrt{-g} \tag{5.3}$$

The only other contribution is the matter field contribution:

Definition 5.3 (Matter action) The *matter action* S_M is defined as

$$S_M = \int_{\mathcal{V}} \mathcal{L}_M(\phi, \partial_\mu \phi, g_{\mu\nu})\sqrt{-g} d^4x \tag{5.4}$$

where ϕ is the matter field, which is a scalar field.

Remark 5.2 Importantly, the term $\sqrt{-g}$ is included in $\mathcal{L}_H$ but left out of $\mathcal{L}_M$. This appears to be a result of convention.

With the Einstein-Hilbert action, we can ultimately recover the field equations. The general strategy is to have S be stationary under any change $\delta\psi$ in the scalar field ψ_0. i.e. for

$$\delta\psi = \left.\frac{d\psi_\lambda}{d\lambda}\right|_{\lambda=0} \tag{5.5}$$

where λ is a parameter, and we demand that

$$\delta\psi|_{\partial\mathcal{V}} = 0 \tag{5.6}$$

Due to the action principle, ψ_0 is a solution to the field equations.

Derivation 5.1 (Proto-field equations) We begin with the Hilbert term. The variation of the Hilbert Lagrangian is

$$16\pi\delta\mathcal{L}_H = -\frac{\delta g}{2\sqrt{-g}} g^{\mu\nu}R_{\mu\nu} + (\delta g^{\mu\nu}R_{\mu\nu} + g^{\mu\nu}\delta R_{\mu\nu})\sqrt{-g} \tag{5.7}$$

We can use *Jacobi's formula*:

Theorem 5.1 (Jacobi's formula) For a matrix A,

$$\frac{d}{dt}\det A(t) = \operatorname{tr}\left(\operatorname{adj}(A(t))\frac{dA(t)}{dt}\right) = (\det A(t)) \cdot \operatorname{tr}\left(A(t)^{-1} \cdots \frac{dA(t)}{dt}\right) \tag{5.8}$$

By using this on the metric, one yields

$$\delta g = g g^{\mu\nu}\delta g_{\mu\nu} = -g g_{\mu\nu}\delta g^{\mu\nu} \tag{5.9}$$

Plugging this into the Hilbert term:

$$16\pi\delta\mathcal{L}_H = \left[\left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R\right)\delta g^{\mu\nu} + g^{\mu\nu}\delta R_{\mu\nu}\right]\sqrt{-g} \tag{5.10}$$

By virtue of the so-called *Paltini identity*, it is understood that

Theorem 5.2 (Paltini identity)

$$\delta R_{\mu\nu} = \nabla_\rho \delta\Gamma^\rho_{\nu\mu} - \nabla_\nu \delta\Gamma^\rho_{\rho\mu} \tag{5.11}$$

We then define a contravariant vector

$$\delta V^\rho \doteq g^{\mu\nu}\delta\Gamma^\rho_{\mu\nu} - g^{\rho\nu}\delta\Gamma^\mu_{\mu\nu} \tag{5.12}$$

The formula for $16\pi\delta\mathcal{L}_H$ can be hence rewritten as

$$\sqrt{-g}g^{\mu\nu}\delta R_{\mu\nu} = \partial_\rho(\sqrt{-g}V^\rho) \tag{5.13}$$

This can then be integrated to solve for the action. From Stokes' theorem:

$$\delta S_H = \frac{1}{16\pi}\int_\Sigma \left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R\right)\sqrt{-g}\delta g^{\mu\nu}d^4x + \frac{1}{16\pi}\oint_{\partial\nu} V^\mu d\sigma_\mu d^3x \tag{5.14}$$

where $d\sigma_\mu$ is the oriented volume element of the hypersurface $\partial\Sigma$.

The variation of the matter action is much easier:

$$\delta S_M = \int_\Sigma \left[\frac{\partial\mathcal{L}_M}{\delta g^{\mu\nu}} - \frac{1}{2}\mathcal{L}_M g_{\mu\nu}\right]\sqrt{-g}\delta g^{\mu\nu}d^4x \tag{5.15}$$

The stress-energy tensor is defined via the matter action as

$$T_{\mu\nu} \doteq -2\frac{\delta\mathcal{L}_M}{\delta g^{\mu\nu}} + \mathcal{L}_M g_{\mu\nu} \tag{5.16}$$

The total variation of the action hence becomes

$$\delta S = \frac{1}{16\pi}\int_\sigma \left[R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi T_{\mu\nu}\right]\sqrt{-g}\delta g^{\mu\nu}d^4x \tag{5.17}$$

From the action principle, $\delta S = 0$. We hence find what will be later determined as a proto-form of the field equations

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi T_{\mu\nu} = \kappa T_{\mu\nu} \tag{5.18}$$

where we have defined the *coupling constant* $\kappa = 8\pi G$[a].

[a]For simplicity, we have previously regarded G as 1. Now we add it back for completeness.

We review the physical significance of the terms:

- $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$ represents the curvature (and with that, the geometry) of the spacetime.
- $T_{\mu\nu}$ represents the matter-energy content of the spacetime.

5.2 Einstein tensor: Gravity is geometry

Before moving on from the proto-field equations, we are tempted to consider the famous *equivalence principle*, which states that the effects of gravitation and acceleration cannot be distinguished. But what is the mathematical representation of this?
Consider, for example, classical two-body gravity and electromagnetism. as determined by the inverse square law:

$$F = \begin{cases} \frac{GMm}{r^2} & \text{electromagnetism} \\ \frac{Q_1Q_2}{kr^2} & \text{gravitation} \end{cases} \tag{5.19}$$

For simplicity, we now consider only one of the two bodies to be in motion. One can determine the acceleration of the moving body by dividing both sides of the two equations by m, the mass of the moving body:

$$a = \begin{cases} \frac{GM}{r^2} & \text{electromagnetism} \\ \frac{Q_1Q_2}{kmr^2} & \text{gravitation} \end{cases} \tag{5.20}$$

Unlike in electromagnetism, the acceleration in gravitation has no dependence on the mass of the moving body - any object, regardless of mass, experiences the same acceleration. It can then be seen that gravitational mass is identical to inertial mass, and that gravitation is not a force in the traditional sense. From this, Einstein hypothesised that perhaps *gravity is geometry*, which eventually led to his field equations.

Quote 5.2 Gravity is geometry

Various GR contributors

But what does this actually mean? Despite having established that gravitation is not a 'traditional' force, the observation that gravitation stems from mass still holds. As such, 'gravity' is then equivalent to matter content. i.e. the stress-energy tensor. Geometry is represented by the curvature tensor terms $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$.
At this point, it is clear that that $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$ represents the part of the curvature information *that corresponds to the matter-energy content.* This correspondence suggests that $R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R$ is a physically meaningful object in itself. This is indeed the case, and the object is known as the *Einstein tensor* $G_{\mu\nu}$.

Definition 5.4 (Trace-reverse) We introduce the *trace-reverse* of a rank-2 tensor

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}g_{\mu\nu}h \tag{5.21}$$

where we predictably have $h = g^{\mu\nu} h_{\mu\nu}$.

Theorem 5.3 (Trace reverse property)

$$g^{\mu\nu}\bar{h}_{\mu\nu} = \bar{h} = h - 2h = -h \tag{5.22}$$

We can then define the Einstein tensor as the trace-reverse of the Ricci tensor:

Definition 5.5 (Einstein tensor)

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \tag{5.23}$$

Quote 5.3 You can think of the Einstein tensor as the trace reverse of the Ricci tensor. Nobody says it like that but strictly speaking this is how it is defined.

Christian G. Böhmer, 23 November 2023

Theorem 5.4 (Einstein tensor properties) The Einstein tensor satisfies the following properties:

- **Symmetry:**

$$G_{\mu\nu} = G_{\nu\mu} \tag{5.24}$$

- **Contracted Bianchi identity:** The Einstein tensor has zero covariant divergence

$$\nabla^{\mu} G_{\mu\nu} = 0 \tag{5.25}$$

- **Trace:**

$$G = g^{\mu\nu} G_{\mu\nu} = -R \tag{5.26}$$

  This follows from Equation 5.22.

We can now finally assemble the field equations.

Fun fact 5.1 (A bit of history) Einstein's original guess of the field equations in 1915 was

Theorem 5.5 (Einstein field equations)

$$G_{\mu\nu} = \kappa T_{\mu\nu} \tag{5.27}$$

Remark 5.3 What could possibly go wrong?

5.3 Variants of the field equations

In developing the Einstein static universe solution (a universe without any dynamics), Einstein discovered that his 1915 guess does not allow this. Soon, however, Einstein corrected his mistake and proposed a 'correct'[1] field equation:

Theorem 5.6 (Einstein field equations with the cosmological constant)

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu} \tag{5.28}$$

where Λ is the cosmological constant.

Remark 5.4 We can likewise derive this from the Einstein-Hilbert action, whose Hilbert term is amended as

$$S_H = \frac{1}{16\pi} \int_{\mathcal{V}} (R - 2\Lambda)\sqrt{-g}\, d^4x \tag{5.29}$$

Note 5.1 Fortunately, beyond cosmology, the term $\Lambda g_{\mu\nu}$ is usually ignored for simplicity.

[1] Read: currently well-accepted

Another variant of the field equations assumes zero matter content. As we have effectively assumed the space to be a vacuum, this is known as the *vacuum field equations*, where both $T_{\mu\nu}$ and R are zero:

Theorem 5.7 (Vacuum field equations)

$$R_{\mu\nu} = 0 \tag{5.30}$$

We can also consider the *Einstein-Maxwell equations.* Despite their name, the equations have very little to do with the unification of forces. Instead, it simply assumes that the matter component of the Einstein field equations (i.e. the stress-energy tensor) is dominated by an EM field. We recall the SR stress-energy tensor for an EM field from Equation 3.13 and replace the Minkowski metric $\eta_{\mu\nu}$ with a general metric $g_{\mu\nu}$. The stress-energy tensor then becomes

$$T_{\mu\nu} = F_{\mu\lambda}F^{\lambda}_{\nu} - \frac{1}{4}g_{\mu\nu}F_{\sigma\rho}F^{\sigma\rho} \tag{5.31}$$

The field equations hence become

Theorem 5.8 (Einstein-Maxwell equation)

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \kappa\left(F_{\mu\lambda}F^{\lambda}_{\nu} - \frac{1}{4}g_{\mu\nu}F_{\sigma\rho}F^{\sigma\rho}\right) \tag{5.32}$$

5.4 Linearised gravity

It is well-known that the field equations are highly non-linear. While a large number of exact solutions are known, the generation of exact solutions is nonetheless considered to be hard. We therefore attempt to convert the field equations into a linear form by utilising a technique known as *linearised gravity* or the *weak field approximation*, with the latter being so-called due to the fact that the gravitational fields involved in the generation of gravitational waves are *smol.*

In the mathematical formulation of gravitational waves, we thus consider a *smol* perturbation $h_{\mu\nu}$ to Minkowski spacetime

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} + \mathcal{O}((h_{\mu\nu})^2) \quad \text{where} \quad |h_{\mu\nu}| \ll 1 \tag{5.33}$$

This process is called *linearised gravity.* An alternate name appears as the *tensor-field theory of gravity in flat spacetime* in *MTW.*

Remark 5.5 As both the metric $g_{\mu\nu}$ and the Minkowski metric $\eta_{\mu\nu}$ are symmetric, $h_{\mu\nu}$ is also symmetric.

Remark 5.6 One can also derive the inverse metric in linearised gravity as

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu} \tag{5.34}$$

Derivation 5.2 (Linearised field equations) We begin by writing out the Christoffels under the linearised gravity regime.

$$\Gamma^{\mu}_{\alpha\beta} = \frac{1}{2}\eta^{\mu\nu}(h_{\alpha\mu,\beta} + h_{\beta\nu,\alpha} - h_{\alpha\beta,\nu}) = \frac{1}{2}(h^{\mu}_{\alpha,\beta} + h^{\mu}_{\beta,\alpha} - h_{\alpha\beta}{}^{,\mu}) \tag{5.35}$$

Remark 5.7 As we have $\eta_{\mu\nu} \approx g_{\mu\nu}$, we have forgoed $g_{\mu\nu}$ for the Minkowski metric in our derivations.

From definition, the so-called *Ricci tensor* is hence

$$R_{\mu\nu} = \frac{1}{2}(h^{\alpha}_{\mu,\nu\alpha} + h^{\alpha}_{\nu,\mu\alpha} - h_{\mu\nu}{}^{,\alpha}_{,\alpha} - h_{,\mu\nu}) \tag{5.36}$$

The last term is effectively a 'scalar perturbation' $h = \eta^{\alpha\beta}h_{\alpha\beta}$ differentiated with respect to the indices μ and ν. From our newfound tradition of using the Minkowski metric, it is not surprising that the trace-reversed perturbation is

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h \tag{5.37}$$

Here we impose a gauge condition. Recall the *Lorenz gauge* in special relativity which is $A^{\alpha}_{,\alpha} = 0$. The tensorial equivalent here is

$$\bar{h}^{\mu\alpha}{}_{,\alpha} = 0 \tag{5.38}$$

Inserting the Ricci tensor and the tensorial Lorenz gauge yields the linearised Einstein field equations.

Theorem 5.9 (Linearised field equations)

$$G_{\mu\nu} = -\frac{1}{2}\Box\bar{h}_{\mu\nu} = -\frac{1}{2}\bar{h}_{\mu\nu ,\alpha}{}^{,\alpha} = \kappa T_{\mu\nu} \tag{5.39}$$

where $\Box$ is the d'Alembertian.

Derivation 5.3 (Plane wave solutions) We now return to the linearised field equations as seen in Theorem 5.9. As gravitational waves traverse through regions effectively devoid of matter, we also set the stress-energy tensor (i.e. matter content) as zero.

Theorem 5.10 (Linearised vacuum field equations)

$$\bar{h}_{\mu\nu ,\alpha}{}^{,\alpha} = 0 \tag{5.40}$$

Noting that this is effectively a wave equation, we can reasonably set up a trial solution

$$\bar{h}_{\mu\nu} = A_{\mu\nu}\exp(ik_\alpha x^\alpha) \tag{5.41}$$

where k_α is the wavevector, x^α is the 4-position and $A_{\mu\nu}$ is some tensorial amplitude. Much like in electromagnetic waves, only the real part of the phase term is physical. Using Euler's identity, we find the physically meaningful part of the equation

$$\bar{h}_{\mu\nu} = A_{\mu\nu}\cos(ik_\alpha x^\alpha) \tag{5.42}$$

By inserting this $\bar{h}_{\mu\nu}$ into the gauge $\bar{h}^{\mu\alpha}_{,\alpha} = 0$, we can find the following constraints on k_α:

$$\underbrace{k_\alpha k^\alpha = 0}_{k_\alpha \text{ is a null vector}} \qquad \underbrace{A_{\mu\alpha}k^\alpha = 0}_{A_{\mu\alpha} \text{ is orthogonal to } k_\alpha} \tag{5.43}$$

Here the physical significance is clearly seen. Expectedly, k_α corresponds to a frequency ω. For convenience, we reduce the number of spatial dimensions to one. The *gravitational wave* is then represented by

Definition 5.6 (Gravitational wave)

$$\bar{h}_{\mu\nu} = A_{\mu\nu}\cos(\omega x - \omega t) \tag{5.44}$$

We can now impose further gauge conditions by adjusting the initial data for the Lorenz gauge equations.

Derivation 5.4 (Transverse-traceless (TT) gauge) For a given 4-velocity u_ν, we impose the following gauge conditions

$$\underbrace{A^{0\nu} = 0 \rightarrow A^{\mu\nu}u_\nu = 0}_{\text{transverse wave}} \qquad \underbrace{A^\mu_\mu = 0}_{\text{traceless wave amplitude}} \tag{5.45}$$

This is the so-called *transverse-traceless gauge* or the *TT gauge*.
From $A^{0\nu} = 0$, we can see that the first row and the first column vanishes. As $A^{\mu\nu}$ is established to be traceless, we also have

$$A^{11} + A^{22} + A^{33} = 0 \tag{5.46}$$

Considering also that $A^{\mu\nu}$ is symmetric, the most general matrix that satisfies these conditions leaves only two independent wave amplitudes out of the original 10:

$$A^{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & A_+ & A_\times & 0 \\ 0 & A_\times & -A_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{5.47}$$

$A_+ = 0$ and $A_\times = 0$ represent two different polarisations of gravitational waves:

- The A_+ mode or the *plus polarisation* describes stretching and squeezing along axes aligned with the x- and y-axes. When a gravitational wave in this mode passes through, it elongates spacetime along one axis (say, the x-axis) while contracting along the perpendicular axis (y-axis), and then alternates this behaviour.

- The $A_\times$ mode or the *cross polarisation* describes stretching and squeezing along axes rotated by 45 degrees relative to the x- and y-axes. i.e., along lines like

$$x' = (x+y)/\sqrt{2} \quad y' = (x-y)/\sqrt{2} \tag{5.48}$$

  The deformation pattern is the same as the A_+ mode, but the axes of elongation and contraction are rotated by 45 degrees.

Remark 5.8 This is analogous to polarisations in EM waves, which are separated by 90°. The angles are different as EM waves correspond to oscillations of EM fields, which are vector fields in orthogonal directions, whereas gravitational waves correspond to tensorial deformations of spacetime that are rotations of each other by 45 degrees in the transverse plane.

5.5 Deriving the Schwarzschild solution

One solution of particular interest is the *Schwarzschild solution*.

Figure 5.1: Faking his death after discovering the eternal tensor of youth $Y_{\mu\nu}$, Schwarzschild mangled his name and became a psychiatrist.

Definition 5.7 (Schwarzschild solution) The Schwarzschild solution describes the gravitational field outside a spherical mass, on the assumption that:

- The electric charge, angular momentum, and cosmological constant are all zero.
- It is spherically symmetric, static and vacuum[a].

[a] Even though a central mass is present, we restrict the manifold to the region outside that central mass. For inside the central mass, we have the so-called *Schwarzschild interior solution* which we will soon investigate.

Remark 5.9 The Schwarzschild metric describes a non-rotating, uncharged black hole. We will discuss the other three cases in the next chapter.

Derivation 5.5 (Schwarzschild solution) In the Minkowski space $X^\mu = (t, r, \theta, \phi)$, we first assume such a Lagrangian (and thus the metric):

$$L = -e^{\nu}\dot{t}^2 + e^{\lambda}\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\dot{\phi}^2 \tag{5.49}$$

where ν and λ are current unknowns and functions of r.
We have the Christoffel symbols

$$\begin{aligned}
&\Gamma^c_{tc} = 0 \\
&\Gamma^c_{rc} = \Gamma^t_{rt} + \Gamma^r_{rr} + \Gamma^\theta_{r\theta} + \Gamma^\phi_{r\phi} = \frac{2}{r} + \frac{1}{2}\left(\nu' + \lambda'\right) \\
&\Gamma^c_{\theta c} = \cot\theta \\
&\Gamma^c_{\phi c} = 0
\end{aligned} \tag{5.50}$$

and by that, the Ricci tensor components

$$\begin{aligned}
R_{tt} &= e^{\nu-\lambda}\left[\frac{1}{2}\nu'' + \frac{1}{4}\left(\nu'\right)^2 + \frac{1}{r}\nu' - \frac{1}{4}\nu'\lambda'\right] \\
R_{rr} &= -\frac{1}{2}\nu'' - \frac{1}{4}\left(\nu'\right)^2 + \frac{1}{4}\nu'\lambda' + \frac{1}{r}\lambda' \\
R_{\theta\theta} &= 1 - e^{-\lambda} + \frac{1}{2}r\lambda' e^{-\lambda} - \frac{1}{2}r\nu' e^{-\lambda} \\
R_{\phi\phi} &= \sin^2\theta R_{\theta\theta}
\end{aligned} \tag{5.51}$$

Due to the vacuum nature of Schwarzschild solutions, we can consider the vacuum field equations, as seen in Equation 5.30. By simply plugging in the Ricci tensor, we find that

$$e^{\nu} = e^{-\lambda} = 1 - \frac{\mathcal{C}}{r} \tag{5.52}$$

Remark 5.10 The gravitational equations imply that the Ricci tensor reduces to zero. This is also called the *source-free Newtonian field equations*.
The metric becomes

$$ds^2 = -\left(1 - \frac{\mathcal{C}}{r}\right)dt^2 + \left(1 - \frac{\mathcal{C}}{r}\right)^{-1}dr^2 + r^2 d\Omega^2 \tag{5.53}$$

where we have the common simplified notation for the angular terms

$$d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2 \tag{5.54}$$

Finally, we note the limit $c^2 \to \infty$ or $1/c^2 \to 0$. Only one of the non-vanishing Christoffel symbol components contains the constant $\mathcal{C}$, namely

$$\lim_{c^2\to\infty}\Gamma^r_{tt} = \frac{G\mathcal{C}}{2r^2} \tag{5.55}$$

Compare this with the Newtonian gravitational potential

$$\nabla\Phi = \frac{GM}{r^2} \tag{5.56}$$

and we can comfortably assume that $\mathcal{C} = 2M$.

Definition 5.8 (Schwarzschild metric) As such, we arrive at the *Schwarzschild metric*:

$$ds^2 = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1}dr^2 + r^2 d\Omega^2 \tag{5.57}$$

Remark 5.11 Two interesting extreme points exist for the Schwarzschild metric. In the limit $r \to \infty$ it approaches Minkowski spacetime in spherical polar coordinates. When $r = 2M$, the metric is singular and corresponds to the event horizon of a hypothetical black hole.

5.6 Orbits and test particles

For the Schwarzschild solution, there are two most commonly studied orbits:

- **Radial fall**: Since θ and ϕ are constants, the $d\Omega^2$ term reduces to zero and the equations simplify significantly. One can thus directly study how the radial coordinate r changes with time.
- **Orbital motion**: We can study the effective potential to understand the conditions for stable and unstable orbits.

Following up from where we left off, we consider the geodesic equations in Schwarzschild coordinates.

Derivation 5.6 (Schwarzschild geodesic equations) Starting with the equation for θ:

$$\frac{d}{d\lambda}(2r^2\dot{\theta}) = 2r^2\ddot{\theta} + 4r\dot{r}\dot{\theta} = 2r^2\sin\theta\cos\theta\dot{\phi}^2 \tag{5.58}$$

We can easily verify that $\theta = \pi/2$ solves this equation. Due to symmetry, the orbit must be planar, and the plane of the orbit is the equatorial plane.
We then consider the t equation. As L is time-independent:

$$\frac{d}{d\lambda}(-2f(r)\dot{t}) = 0 \to f(r)\dot{t} = E \tag{5.59}$$

where $f(r) = 1 - \frac{2M}{r}$ and E is related to energy. We then consider the ϕ equation. As L is ϕ-independent:

$$\frac{d}{d\lambda}(2r^2\sin^2\theta\dot{\phi}) = 0 \to r^2\dot{\phi} = 2\ell \tag{5.60}$$

where ℓ is related to angular momentum.
We plug $\phi = \pi/2$, $E = f(r)\dot{t}$ and $\ell = r^2\dot{\phi}$ into the Lagrangian:

$$Lf(r) = -E^2 + \dot{r}^2 + \frac{\ell^2}{r^2}f(r) \tag{5.61}$$

Hence

$$E^2 = \dot{r}^2 + \left(1 - \frac{2m}{r}\right)\left(\frac{\ell^2}{r^2} - L\right) \tag{5.62}$$

Dividing both sides by 2, we find that

$$\frac{1}{2}E^2 = \underbrace{\frac{1}{2}\dot{r}^2}_{\text{kinetic energy}} + \underbrace{\frac{1}{2}\left(1 - \frac{2M}{r}\right)\left(\frac{\ell^2}{r^2} - L\right)}_{\text{potential energy}} \tag{5.63}$$

We then recall from the last chapter that we can study orbital motions using the effective potential, which we define as the second term of the last equation:

Definition 5.9 (Effective potential)

$$V_{\text{eff}}(r) = \frac{1}{2}\left(1 - \frac{2M}{r}\right)\left(\frac{\ell^2}{r^2} - L\right) \tag{5.64}$$

Expanding yields

Theorem 5.11 (Schwarzschild geodesic equation)

$$V_{\text{eff}}(r) = \underbrace{\frac{LM}{r}}_{\text{Newtonian term}} + \underbrace{\frac{1}{2}\frac{\ell^2}{r^2}}_{\text{centrifugal barrier term}} - \underbrace{\frac{M\ell^2}{r^3}}_{\text{GR correction}} - \underbrace{\frac{1}{2}L}_{\text{insignificant constant}} \tag{5.65}$$

Remark 5.12 In small distances, the GR term dominates over the barrier term. We will see later that this leads to perihelion precession. By studying $V_{\text{eff}}(r)$, we will find that Newtonian mechanics is merely an extreme-case approximation of GR.

> **Quote 5.4** We got to the Moon with Newton. (...) Newton is perfectly fine to go to the Moon.
>
> *Christian G. Böhmer, on Newtonian mechanics, 6 December 2023*

Exercise 5.1 Noting the conservation of mass and energy in relation to the geodesic equations, derive that

$$\frac{dt}{ds} = \frac{E}{M}\left(1 - \frac{2M}{r}\right)^{-1} \qquad \frac{d\phi}{dr} = \frac{\ell/M}{r^2 \sin^2\theta} \tag{5.66}$$

In actuality, we are not particularly interested in the physical significance of E and ℓ. Instead, they simplify the Lagrangian so that we may conduct analyses with the following:

Theorem 5.12 (Schwarzschild Lagrangians) The world-line of a massive particle is assumed to be a timelike geodesic, and that of a photon is a null geodesic. For these cases:

$$L = \begin{cases} -1 & \text{for massive particles} \\ 0 & \text{for massless particles (photons, gravitons, etc.)} \end{cases} \tag{5.67}$$

Remark 5.13 For a massive particle, the Lagrangian is a very specific -1 dut to the tangent vector u^μ and m being normalised such that $L = -m\sqrt{g_{\mu\nu}u^\mu u^\nu} = -1$. For a photon, the absence of mass naturally gives $L = -m\sqrt{g_{\mu\nu}u^\mu u^\nu} = 0$. Note also that for a metric of signature $(+,-,-,-)$, we have $L = 1$ for a massive particle.

Remark 5.14 We can further consider the physical meaning of the two cases. For a massive particle, the tangent vector along a geodesic is the 4-velocity. For a massless particle, the tangent vector lives right on a light cone and is called a *lightlike vector* or *null vector*.

Definition 5.10 (Schwarzschild orbits) Orbits emerge on the radial stationary points of the effective potential. i.e. when

$$\frac{dV_{\text{eff}}}{dr} = 0 \to Lr^2 + \frac{\ell^2}{M}r - 3\ell^2 = 0 \tag{5.68}$$

Theorem 5.13 (Stability of orbits) By investigating the nature of stationary points, the stability of their corresponding orbits can be found:

- **Maximum**: Any *smol* perturbations will cause the particle to either fall towards the black hole or escape away from it. As such, the orbit is *unstable*.
- **Minimum**: Any *smol* perturbations will lead to oscillations about the orbit, with the particle ultimately returning to its orbit. As such, the orbit is *stable*.
- **Saddle point**: The point behaves like a maximum in one direction and like a minimum in another, and perturbations in the direction of the maximum will lead to *instability*.

Derivation 5.7 (Schwarzschild orbit equation) Returning to the orbital plane, we can again simplify the metric by considering the case of $\theta = \pi/2$ due to spherical symmetry. The resultant *equatorial Schwarzschild metric* is hence

$$ds^2 = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1} dr^2 + r^2 d\phi^2 \tag{5.69}$$

Dividing the metric by ds^2 and utilising the solutions of the last exercise, we find that

$$\frac{dr}{ds} = \sqrt{\frac{E^2}{M^2} - \frac{1}{M^2}\left(\frac{\ell^2}{r^2M^2} - L\right)\left(1 - \frac{2M}{r}\right)} \tag{5.70}$$

Noting that we are trying to derive a *orbit equation*, we replace s with ϕ to equate the radial and orbital components:

$$\frac{dr}{ds} = \frac{dr}{d\phi}\frac{\ell}{r^2 M} \tag{5.71}$$

This then gives

Definition 5.11 (Orbit equation)

$$\left(\frac{dr}{d\phi}\right)^2 = \frac{E^2 - (1 - \frac{2M}{r})\left(\frac{\ell^2}{r^2} - L\right)}{\ell^2/r^4} \tag{5.72}$$

A commonly used form of this has r replaced by $u = \frac{1}{r}$ for clarity[a]. It then follows that $\frac{dr}{d\phi} = -\frac{1}{u^2}\frac{du}{d\phi}$ and

$$\left(\frac{du}{d\phi}\right)^2 = \frac{E^2 + L}{\ell^2} - \frac{2ML}{\ell^2}u - u^2 + 2Mu^3 \tag{5.73}$$

Remark 5.15 Ignoring the last term yields the Newtonian orbit equation.
Differentiating with respect to ϕ, we can simplify the equations even more.

Definition 5.12 (Simplified orbit equation)

$$\frac{d^2u}{d\phi^2} = -\frac{ML}{\ell^2} - u + 3Mu^2 \tag{5.74}$$

Remark 5.16 Again, ignoring the last term yields the Newtonian orbit equation.

[a]You will almost immediately see why.

Exercise 5.2 Use the Einstein and Newton orbit equation to investigate a planet close to the Sun. Prove that the relativistic correction term results in very small deviations from the Newtonian predictions.

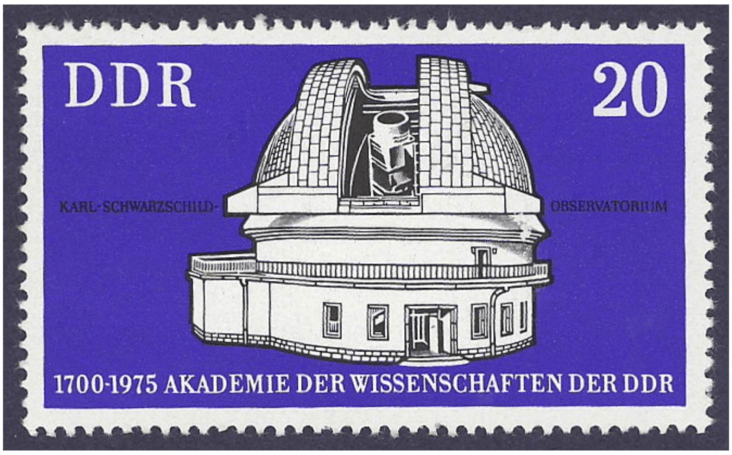


Figure 5.2: The GDR Academy of Sciences celebrates Schwarzschild.

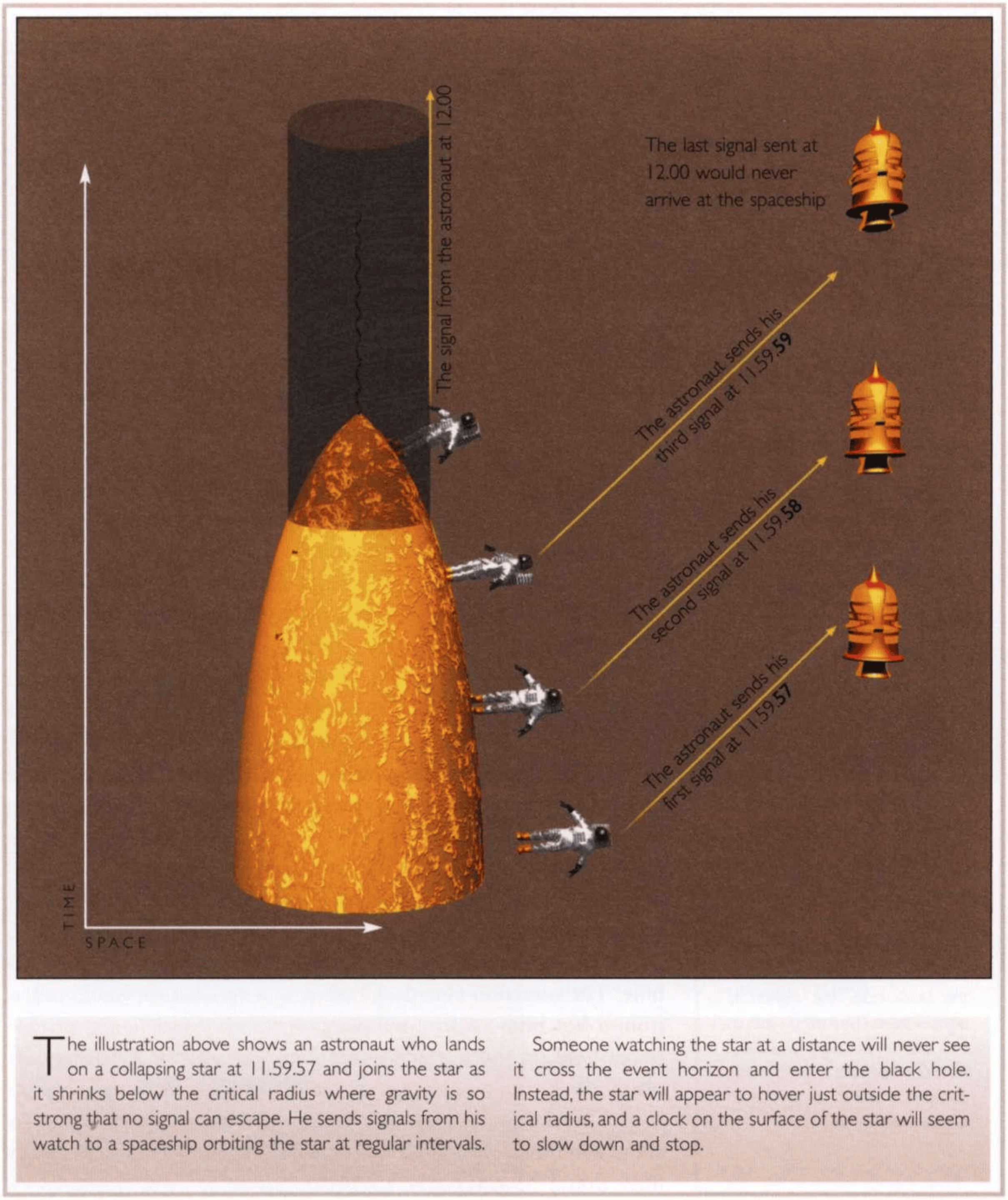


Figure 5.3: Astronaut entering the event horizon (*The Universe in a Nutshell*)

Chapter 6

Other coordinates and black holes

Quote 6.1 Where we're going, we won't need eyes to see.

William Weir, in Event Horizon

There is a book 'Exact Solutions of Einstein's Field Equations' which contains thousands of solutions to the field equations. But we are only concerned with a solution if it is physically relevant: For example, we are interested in an exact solution of the field equation that can model the gravitational field outside a massive object, like the Sun.

6.1 Eddington-Finkelstein coordinates

In the last chapter, we discovered that the Schwarzschild metric is singular for $r = 2m$ (i.e. the event horizon). But why exactly is this? We can better understand this in a set of new coordinates.

Derivation 6.1 (Eddington-Finkelstein coordinates) We first attack the rather nasty $1 - \frac{2M}{r}$ term by performing a transformation of the radial coordinate and turning it into the so-called tortoise coordinate:

$$dr^* = \left(1 - \frac{2M}{r}\right)^{-1} dr \rightarrow r^* = r + 2M \ln\left|\frac{r}{2M} - 1\right| \tag{6.1}$$

Now the metric is of the form

$$ds^2 = \left(1 - \frac{2M}{r}\right)(-dt^2 + dr^{*2}) + r^2 d\Omega^2 \tag{6.2}$$

We then define the *advanced null coordinate* or rather *advanced time* and set

$$v = t + r^* \tag{6.3}$$

The metric hence becomes

Definition 6.1 (Schwarzschild metric in ingoing Eddington-Finkelstein coordinates)

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dv^2 + 2dvdr + r^2 d\Omega^2 \tag{6.4}$$

This subtype is so-called because it is useful for studying paths falling into the black hole. Paths moving away from the black hole are alternatively described by the *retarded null coordinate* or rather *retarded time*

$$u = t - r^* \tag{6.5}$$

which yields the metric

Definition 6.2 (Schwarzschild metric in outgoing Eddington-Finkelstein coordinates)

$$ds^2 = -\left(1 - \frac{2M}{r}\right) du^2 - 2dudr + r^2 d\Omega^2 \tag{6.6}$$

Remark 6.1 The metric in both Eddington-Finkelstein coordinates types is very peculiar as it is non-diagonal. From this can note that the singularity $r = 2M$ no longer exists:

- We can then conclude that the singularity $r = 2M$ that originally existed in Schwarzschild coordinates is only really a restriction imposed by the geometry of the metric and holds no real physical significance.
- Eddington-Finkelstein coordinates therefore allow for the study of the causal structure of black holes at the event horizon and an object's experience as it crosses the event horizon.

Remark 6.2 One interesting note is that the concept of *white holes* can be developed from the outgoing coordinates. By investigating the ingoing and outgoing metrics one can conclude that the outgoing metric is simply the *time reverse* of the ingoing metric. If we ignore that time is reversed for the outgoing metric, the scenario it physically describes would be a large mass that is spewing out stuff - a white hole. The horizon $r = 2M$ is then one for which events that occur outside it could never be seen from inside.

6.2 Kruskal-Szekeres coordinates

In 1960, Joseph Kruskal and George Szekeres[1] independently developed a coordinate system covering the entire manifold which would later be named after them. In Eddington-Finkelstein Coordinates, the light cones near and inside the horizon are highly deformed. In Kruskal-Szekeres coordinates this problem is eliminated.

Derivation 6.2 (Kruskal-Szekeres coordinates) To introduce coordinates that are adapted to the light-cone structure, we start with the Schwarzschild metric in *both* advanced and retarded time:

$$ds^2 = -\left(1 - \frac{2M}{r}\right) du dv + r^2 d\Omega^2 \tag{6.7}$$

Unfortunately here we see that the nasty singularity has returned. To eliminate it, we impose two more coordinate transformations:

$$U = -4Me^{-u/4M} \quad V = 4Me^{v/4M} \tag{6.8}$$

and the metric becomes

$$ds^2 = -\frac{2M}{r} e^{-r/2M} dU dV + r^2 d\Omega^2 \tag{6.9}$$

Still this looks a bit awkward. We can, however, reintroduce coordinates that correspond to t and r:

$$T = \frac{1}{2}(V + U) \quad R = \frac{1}{2}(V - U) \quad \text{for} \quad r > 2M \tag{6.10}$$

$$T = \frac{1}{2}(V - U) \quad R = \frac{1}{2}(V + U) \quad \text{for} \quad r > 2M \tag{6.11}$$

thus giving

Definition 6.3 (Kruskal-Szekeres coordinates)

$$ds^2 = \frac{2M}{r} e^{-r/2M} (-dT^2 + dR^2) + r^2 d\Omega^2 \tag{6.12}$$

where r is now implicitly defined as

$$16M^2 \left(\frac{r}{2M} - 1\right) e^{r/2M} = R^2 - T^2 \tag{6.13}$$

6.3 Inside the mass: Schwarzschild interior solution

Ironically, the Schwarzschild solution is a vacuum solution due to the fact that the geometry excludes the interior of the mass. The interior of the mass can alternatively be described by the *Schwarzschild*

[1] Not to be confused with his son Peter, who was responsible for one of the references of this book.

interior solution, which assumes the mass to be a spherical and uniform perfect fluid.

Derivation 6.3 (Schwarzschild interior solution) We recall that the stress-energy tensor of a perfect fluid is defined by Equation 3.18, where u is the 4-velocity observing

$$u = (-e^{A(R)/2}, 0, 0, 0) \tag{6.14}$$

We can them derive the stress-energy tensor

$$T_{\mu\nu} = \begin{pmatrix} \rho e^{A(r)} & 0 & 0 & 0 \\ 0 & pe^{B(r)} & 0 & 0 \\ 0 & 0 & pr^2 & 0 \\ 0 & 0 & 0 & \sin^2\theta \end{pmatrix} \tag{6.15}$$

Knowing that the mass of a sphere is $\frac{4\pi}{3}\rho_0 r^3$, we can find

$$e^{-B} = 1\frac{2m(r)}{r} \tag{6.16}$$

As energy and momentum is conserved:

$$\nabla_\mu T^\mu_\nu = 0 \tag{6.17}$$

we can likewise determine A

$$A(r) = \log\left(\frac{C}{\rho_0 + p(r)}\right)^2 \tag{6.18}$$

and thus, the Schwarzschild interior metric

Definition 6.4 (Schwarzschild interior metric)

$$ds^2 = -\left(\frac{\rho_0 + p_c}{\rho_0 + p(r)}\right)^2 dt^2 + \frac{dr^2}{1 - (8\pi/3)\rho_0 \pi^2} + r^2 d\Omega^2 \tag{6.19}$$

6.4 Charged nonrotating mass: Reissner-Nordström solution

Definition 6.5 (Reissner-Nordström metric)

$$ds^2 = dt^2 = \left(1 - \frac{2M}{r} + \frac{r_q^2}{r^2}\right) dt^2 - \left(1 - \frac{2M}{r} + \frac{r_q^2}{r^2}\right)^{-1} dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta d\varphi^2 \tag{6.20}$$

where we have $r_q^2 = \frac{q^2}{4\pi\varepsilon_0}$. This is a term representing charge.

6.5 Uncharged rotating mass: Kerr solution

Definition 6.6 (Kerr metric)

$$ds^2 = -\frac{\Delta_r}{\varrho^2}\left(dt - a\sin^2\theta d\phi\right)^2 + \frac{\varrho^2}{\Delta_r} dr^2 + \varrho^2 d\theta^2 + \frac{\sin^2\theta}{\varrho^2}\left(adt - \left(r^2 + a^2\right) d\phi\right)^2 \tag{6.21}$$

where

$$\varrho^2 = r^2 + a^2\cos^2\theta \quad \Delta_r = r^2 - 2Mr + a^2 \tag{6.22}$$

M is the mass, and $a = J/M$ is a term representing rotation associated the angular momentum J.

Remark 6.3 This coordinate system is known as the *Boyer-Lindquist coordinates*. You may have noticed at this point that it is the 4D analogue to spherical coordinates.

Exercise 6.1 Show that in the limit $a \to 0$, the Kerr metric becomes the Schwarzschild metric.

Remark 6.4 The singularities, assuming $M \neq 0$, are $r = 0$ and $\theta = \pi/2$.
We consider the case where $\Delta_r = 0$. Assuming $M^2 - a^2 > 0$, r has two roots:

$$r_\pm = M \pm \sqrt{M^2 - r^2} \tag{6.23}$$

These are essentially horizons of a Kerr black hole, In the limit $a \to 0$, r_+ and r_- correspond to the two Schwarzschild event horizons $r = 2M$ and $r = 0$.
For $m = 0$ one yields the reduced metric

$$ds^2 = -dt^2 + \left(r^2 + a^2 \cos^2\theta\right)\left(\frac{dr^2}{r^2 + a^2} + d\theta^2\right) + \left(r^2 + a^2\right)\sin^2\theta d\phi^2 \tag{6.24}$$

We now consider the time component g_{tt}. It vanishes at the spacelike surface

$$r_{e\pm} = M \pm \sqrt{M^2 - a^2\cos^2\theta} \tag{6.25}$$

r_{e+} is known as the *stationary limit* while r_{e-} is not known to have a name[2]. The region $r_+ < r < R_{e+}$ is known as the *ergosphere*. Within it, any observer must co-rotate with the black hole. It is not possible to 'stand still', but it is possible to escape to infinity.

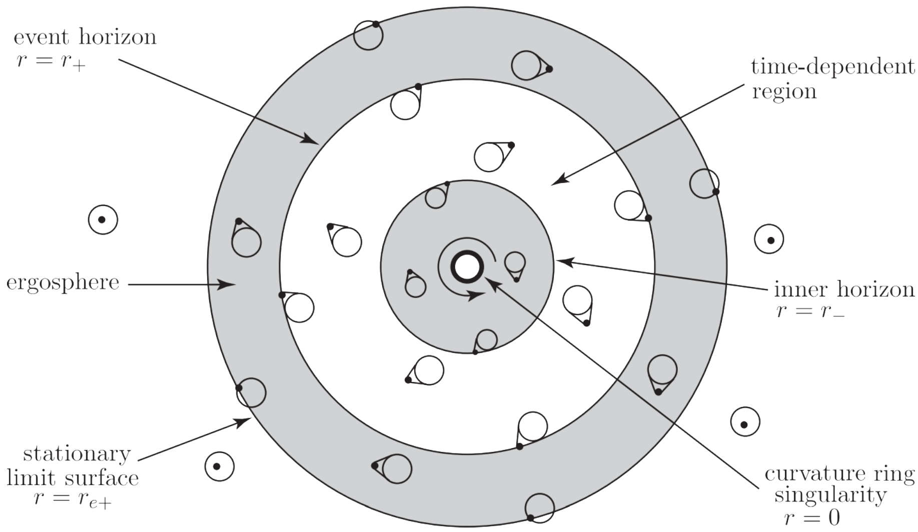


Figure 6.1: The ergosphere.

As the metric of solutions becomes increasingly complex, so do their geodesics. So far we have been using the proper time τ as the parameter. This parameter can be changed as a mathematical convenience, and the new parameter is called *Mino time*. We take the Kerr metric for example:

Definition 6.7 (Mino time in the Kerr metric) In the Kerr metric, the Mino time λ observes

$$\frac{d\tau}{d\lambda} = \varrho^2 \tag{6.26}$$

Remark 6.5 Note that Mino time is a general term for such parameter changes. It is whatever 'rescaled' parameter that happens to simplify calculations in the metric you are working with.

6.6 Charged rotating mass: Kerr-Newman solution

Finally, the Kerr-Newman solution, discovered by Roy Kerr and Alfred Schild in 1963, describes black hole solutions with both charge and rotation. It is basically a combination of the Reissner-Nordström and Kerr solutions.

[2] Aww!

Definition 6.8 (Kerr-Newman metric) At first glance, the Kerr-Newman metric is identical to the Kerr metric

$$ds^2 = -\frac{\Delta_r}{\varrho^2}\left(dt - a\sin^2\theta d\phi\right)^2 + \frac{\varrho^2}{\Delta_r}dr^2 + \varrho^2 d\theta^2 + \frac{\sin^2\theta}{\varrho^2}\left(adt - \left(r^2 + a^2\right)d\phi\right)^2 \tag{6.27}$$

where we still have $\varrho^2 = r^2 + a^2\cos^2\theta$, but Δ_r is modified to include charge:

$$\Delta_r = r^2 - 2Mr + a^2 + r_q^2 \tag{6.28}$$

Once again, when both the charge- and angular momentum-related terms are zero, the metric reduces to the Schwarzschild metric.

Figure 6.2: Schwarzschild takes a dirt nap (J. Giesen, 2004)

Chapter 7

Classical tests of GR

Quote 7.1 Everyone was silent for a minute. Then Filby said he was damned.

H. G. Wells, The Time Machine

In the last chapter of Part I, we will familiarise ourselves with the three classical tests of GR: advance of Mercury perihelion, deflection of starlight and gravitational redshift.

7.1 Perihelion precession of Mercury

Quote 7.2 But then I said, ‘In that frame of reference, the perihelion of Mercury would have precessed in the opposite direction.’

Stephen Hawking, in ‘Descent, Part 1’, 1993

Figure 7.1: ‘That is a great story!’

Definition 7.1 (Perihelion and aphelion) The perihelion q and aphelion Q are the nearest and farthest points respectively of a body's direct orbit around the Sun.

In the old days, it was very easy to calculate orbits, as they were treated as idealised systems. The orbits of objects moving around the Sun were treated as ellipses. However, if we account for various other masses in the Solar System (as we reasonably would), *smol* variations in a body's orbits are inevitably introduced[1], and its perihelion (and with that, its aphelion) is said to *precess*:

Definition 7.2 (Precession) We define the *precession* $\delta\phi$ as follows:

$$\underbrace{\phi(u_{\max})}_{\text{orbit } n} - \underbrace{\phi(u_{\max})}_{\text{orbit } n-1} = 2\pi + \delta\phi \quad \text{where} \quad (n-1)\pi \leq \phi_n \leq n\pi \tag{7.1}$$

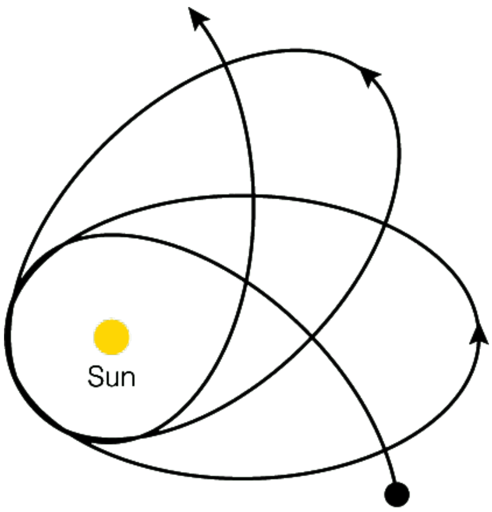


Figure 7.2: Unrealistically exaggerated diagram of Mercury precession.

In all other planets, Newtonian mechanics give good predictions of how their perihelions precess. However, this is not the case for Mercury, which is significantly closer to the Sun. Perhaps more outrageously, it was then discovered that the GR correction term in the effective potential was causing Mercury's orbit to vary from classically predicted values very slightly[2]. This has since become one of the most infamous effects caused by GR.

Derivation 7.1 (Perihelion procession) We seek a (non-periodical!) function $u(\phi)$ as the solution. As we have a planet, the case is that of a massive particle ($L = -1$). Already we can see that the orbit equation reduces to

$$\left(\frac{du}{d\phi}\right)^2 = \frac{E^2-1}{\ell^2} + \frac{2M}{\ell^2}u - u^2 + 2Mu^3 \tag{7.2}$$

For simplicity, we assume u to be small[a] and study first the Newtonian term by leaving out the last term. We can use a substitution $v = u - \frac{M}{\ell^2}$.

- **Newtonian orbit equation:** After substitution, we have

$$\left(\frac{dv}{d\phi}\right)^2 = \frac{E^2-1}{\ell^2} + \frac{2M^2}{\ell^4} - v^2 \tag{7.3}$$

  The general solution is hence

$$v = C\cos(\phi+\phi_0) \to \frac{1}{r} = \frac{M}{\ell^2} + C\cos(\phi+\phi_0) \tag{7.4}$$

[1] We're doomed!
[2] We recall that the GR correction dominates in smaller distances.

which is expectedly periodical.

- **Relativistic orbit equation:** We can attempt the same substitution again:

$$\left(\frac{dv}{d\phi}\right)^2 = \frac{E^2-1}{\ell^2} + \frac{M^2}{\ell^4} + \frac{2M^4}{\ell^6} + \frac{6M^3}{\ell^4}v - \left(1-\frac{6M^2}{\ell^2}\right)v^2 \tag{7.5}$$

Not much better! But it does give us a clue about another substitution:

$$w = v - \frac{3M^3}{\ell^4}\left(1-\frac{6M^2}{\ell^2}\right)^{-1} \tag{7.6}$$

The equation becomes

$$\left(\frac{dw}{d\phi}\right)^2 = a^2 - b^2w^2 \tag{7.7}$$

where a is a constant that collects all the uninteresting terms and $b = \sqrt{1-\frac{6M^2}{\ell^2}}$. The general solution then is

$$w = \frac{a}{b}\cos(b\phi) \tag{7.8}$$

Representing this in terms of r gives

$$\frac{1}{r} = c + \frac{a}{b}\cos(b\phi) \tag{7.9}$$

where yet again the constant c collects all the new uninteresting terms.

Remark 7.1 a is a scaling constant while c is essentially a phase factor.

A simple comparison between the Newtonian and GR solutions reveals that the only real interesting term is b. Why is this? In an idealised world tailored to physicists, we would have $b = 1$, and the orbit would be periodical. But this is obviously not the case, and we hence find the physical significance of b.

> **Quote 7.3** There is a slight failure of the object to return to its starting point after one turn.
>
> *Christian G. Böhmer, in his book*

We can then write down precession by its definition:

$$\delta\phi = \frac{2\pi}{b} - 2\pi = \frac{2\pi}{\sqrt{1-\frac{6M^2}{\ell^2}}} - 2\pi \tag{7.10}$$

Assuming $\frac{M^2}{\ell^2} \ll 1$, we can expand $\delta\phi$ and eliminate any higher-order terms:

$$\delta\phi = \frac{6\pi M^2}{\ell^2} \tag{7.11}$$

Now we try to somehow relate this equation to the radius, which is surprisingly possible. Recall that the perihelion is a stationary point which leads to $\frac{dV_{\text{eff}}}{dr} = 0$ and use $L = -1$:

$$r^2 - \frac{\ell^2}{M}r + 3\ell^2 = 0 \to \ell^2 = \frac{rM}{1-\frac{3M}{r}} \tag{7.12}$$

Once again we will have to expand this to get ourselves anywhere. In the first order we have $\ell^2 = rM$. Plugging this in and we finally have an approximation for the perihelion precession:

$$\delta\phi = \frac{6\pi M}{r} \tag{7.13}$$

Using the mean radius, which is 58×10^6 km, we find that the Mercury makes a procession 1 arcsecond every 880 days[b].

[a] This arises from the usually large r.
[b] A year on Mercury is 88 Earth days.

Remark 7.2 Historically, derivations of perihelion precession usually start with the Newtonian orbit equations and introduce a very small perturbation term. The proof described in this book only emerged at a later point.

Quote 7.4 Older books were written for an audience that was extremely familiar with Newtonian gravity. All ODEs appearing in Newtonian gravity and their solutions in various forms were considered basic. This is no longer the case and thus some of these texts can seem hard and difficult to follow.

Christian G. Böhmer, 14 May 2024

7.2 Light deflection by the Sun

Around an object with a large mass, light is deflected or rather 'bent'. This is best seen during a solar eclipse.
During this process, the light comes in along one asymptote in a negative ϕ_-, reaches the point of closest approach or rather the so-called *impact parameter* r_0, and then moves away along another asymptote in ϕ_+, thus forming an angle $2\phi_+$. In the absence of gravity, the angle subtended would have been π. The difference is hence the deflection angle $\Delta\phi$:

Definition 7.3 (Deflection angle) The *deflection angle* is defined as

$$\Delta\phi = 2\phi - \pi \tag{7.14}$$

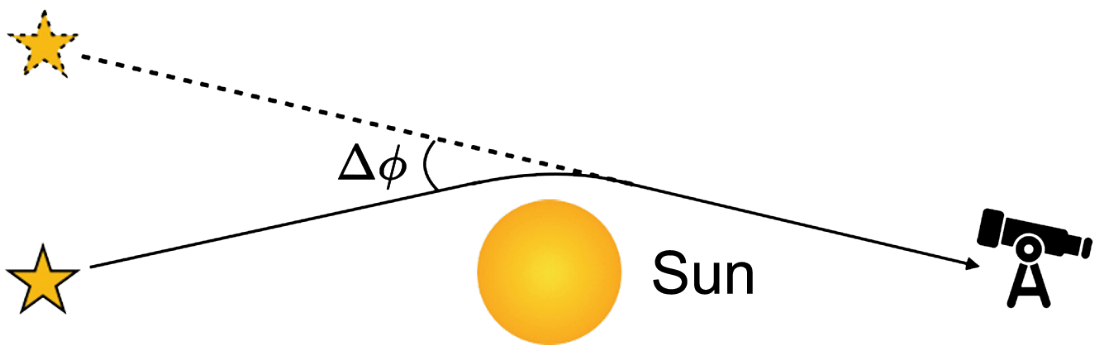


Figure 7.3: Likewise unrealistically exaggerated diagram of light deflection.

Derivation 7.2 (Light deflection) Here the case is that of a photon ($L = 0$). We again consult the relativistic orbit equation, which becomes

$$\left(\frac{dr}{d\phi}\right)^2 = \frac{E^2 - \left(1 - \frac{2M}{r}\right)\frac{\ell^2}{r^2}}{\ell^2/r^4} \tag{7.15}$$

We can consider the impact parameter, at which point the light ray is neither moving closer to nor further from the Sun in the radial direction. i.e. the radial motion is zero, manifesting in $\frac{dr}{d\phi} = 0$. We thus have

$$E^2 = \left(1 - \frac{2M}{r_0}\right)\frac{\ell^2}{r_0^2} \tag{7.16}$$

Inserting this back into the orbit equation, it follows that

$$\frac{d\phi}{dr} = \frac{1}{r^2}\left[\left(1 - \frac{2M}{r_0}\right)\frac{1}{r_0^2} - \left(1 - \frac{2M}{r}\right)\frac{1}{r^2}\right]^{-\frac{1}{2}} \tag{7.17}$$

Note that we have turned the result upside down, which allows us to integrate the conveniently separated variables. Here we essentially integrate over the trajectory of the photon from afar at infinite distance $\lim_{\bar{r}\to\infty} r$ and incoming angle $\phi_- = \phi_+$ to the impact parameter r_0 and angle 0:

$$\phi_+ = \lim_{\bar{r}\to\infty} \int_{r_0}^{\bar{r}} \frac{dr}{\left[\left(1-\frac{2M}{r_0}\right)\frac{1}{r_0^2} - \left(1-\frac{2M}{r}\right)\frac{1}{r^2}\right]^{1/2} r^2} \tag{7.18}$$

This looks very menacing. In the spirit of the emerging trend we assume a *smol* mass, which allows us to ignore any terms with M^2 and higher orders. We then want to find

$$\phi_+ = \phi_+(M=0) + \left.\frac{\partial \phi_+}{\partial M}\right|_{M=0} \tag{7.19}$$

This expression gives

$$\phi_+ = \frac{\pi}{2} + \frac{2M}{r_0} \tag{7.20}$$

and therefore from $\delta\phi = 2\phi_+ - \pi$ we find

Theorem 7.1 (Light deflection)

$$\delta\phi = \frac{4GM}{r_0} \tag{7.21}$$

This is about 1.74 arcseconds for the Sun, where $r_0 = R_\odot$.

7.3 Gravitational redshift

Gravitational redshift arises due to the difference in gravitational potential between the point of emission and the point of observation, not due to the relative motion of the emitter and observer. The implication is that we do not consider the Doppler effect, and the emitter is considered stationary.

Definition 7.4 (Redshift) The *redshift* z is defined as

$$z = \frac{\lambda_{\text{obs}} - \lambda_{\text{e}}}{\lambda_{\text{e}}} \tag{7.22}$$

where λ_{obs} and λ_{e} are the wavelengths of the observed and emitted signals respectively.

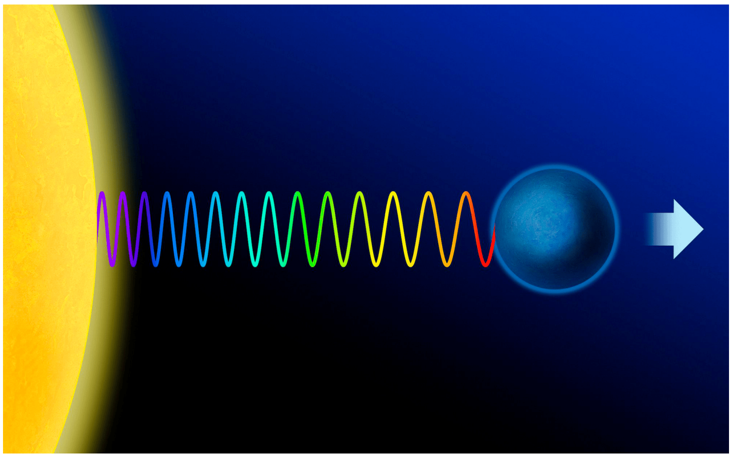


Figure 7.4: A third unrealistically exaggerated diagram, this time of gravitational redshift.

Derivation 7.3 (Gravitational redshift) We write s as τ in this section due to its physical meaning as proper time and recall the norm:

$$d\tau_{\text{e}} = \sqrt{-g_{\mu\nu} dX^\mu dX^\nu} \tag{7.23}$$

Given the stationary emitter, all the spatial components of the metric are considered zero:

$$d\tau_{\mathrm{e}}^2 = -\left(1 - \frac{2M}{r}\right) dt_{\mathrm{e}}^2 \tag{7.24}$$

We then find the emitter's proper time:

$$d\tau_{\mathrm{e}} = \sqrt{1 - \frac{2M}{r_{\mathrm{e}}}} dt_{\mathrm{e}} \tag{7.25}$$

Likewise, the observer's proper time is

$$d\tau_{\mathrm{e}} = \sqrt{1 - \frac{2M}{r_{\mathrm{obs}}}} dt_{\mathrm{obs}} \tag{7.26}$$

Remark 7.3 This inadvertently leads to the all-too-familiar time dilation equation.

$$dt = \frac{1}{\sqrt{1 - \frac{2M}{r}}} d\tau \tag{7.27}$$

Recalling that $\lambda = \frac{v}{f} = t$, we combine redshift and proper time, yielding

Theorem 7.2 (Gravitational redshift)

$$z = \frac{dt_{\mathrm{obs}}}{dt_{\mathrm{e}}} - 1 = \sqrt{\frac{1 - \frac{2M}{r_{\mathrm{obs}}}}{1 - \frac{2M}{r_{\mathrm{e}}}}} - 1 \tag{7.28}$$

Using the Sun as an emitter, we find that $z \approx 2 \times 10^{-6}$.

Figure 7.5: 'You shouldn't be able to pump solids, but you can pump peanut butter.' (Ulysses Tyson, 23 November 2023)

Part II

Advanced topics

Chapter 8

Tetrads

Quote 8.1 Time is nature's way to keep everything from happening all at once.

Ray Cummings, The Girl in the Golden Atom

As a general rule, curved spaces are more mathematically complex than flat spaces. Quite often, it is useful to reduce the local reference frame to a 4D flat Minkowski space. This is accomplished by transforming the metric of interest into the Minkowski metric using tetrad fields. As a formalism, the *tetrad formalism* does not alter predictions; it is rather a calculational technique.

8.1 Tetrad formalism

First we ask ourselves: what is a tetrad? The name comes from 'four' in Greek. It is also called a *Vierbein*[1].
We follow the convention used in most literature: the coordinate indices are represented by Greek letters, and the tetrad indices by Latin letters:

Definition 8.1 (Tetrad basis vector) A *tetrad basis vector* or *tetrad basis*, much like a coordinate basis, is simply the set of axes of an *orthonormal* coordinate system:

$$\gamma_a = (\gamma_0, \gamma_1, \gamma_2, \gamma_3) \tag{8.1}$$

The corresponding *tetrad metric* is

$$\gamma_{ab} = \gamma_a \gamma_b \tag{8.2}$$

Remark 8.1 In this book, the tetrad *basis* γ_a should not be confused with the tetrad *metric* γ_{ab}.

Note 8.1 In theory, γ_{ab} can be any metric we want to transform the target into. However, in GR, γ_{ab} is almost *always* the Minkowski metric η_{ab}.

Tetrad bases are quite distinct from coordinate bases (i.e. coordinate systems):

- A coordinate system provides a global description of spacetime positions, while a tetrad provides a local orthonormal basis at each point.
- Coordinate basis vectors are generally not orthonormal, whereas tetrad basis vectors are orthonormal.

Remark 8.2 In Riemannian geometry the vector spaces include the tangent space, the cotangent space, and the higher tensor spaces constructed from these. In gauge theories, we are concerned with 'internal' vector spaces. The distinction is that while the tangent space and its relatives are associated with the manifold itself, an internal vector space can be of any dimension we like, and has to be defined as an independent addition to the manifold.

[1]German for 'four legs' - this is a generalisation of the *Vielbein* or 'many legs'.

Definition 8.2 (Fibre bundle) The union of the base manifold with the internal vector spaces (defined at each point) is a *fibre bundle*, and each copy of the vector space is called a *fibre* (in accordance with our definition of the tangent bundle).

Quote 8.2 Why introduce tetrads?

1. The physics is more transparent when expressed in a locally inertial frame (or some other frame adapted to the physics), as opposed to the coordinate frame, where Salvador Dali rules.
2. If you want to consider spin-$\frac{1}{2}$ particles and quantum physics, you better work with tetrads.
3. For good reason, much of the general relativistic literature works with tetrads, so it's useful to understand them.

Andrew J. S. Hamilton

To transform between the tetrad basis γ_b and the coordinate basis X_a, we use an object called the *tetrad field*:

Definition 8.3 (Tetrad field) The tetrad field or *frame field* e_a^μ converts the coordinate basis X_μ to the tetrad basis γ_a:

$$\gamma_a = e_a^\mu X_\mu \tag{8.3}$$

Theorem 8.1 (Tetrad field transformations) The tetrad field can transform under a *local Lorentz transformation* Λ_b^a which acts on the tetrad index or a *general coordinate transformation* $\frac{\partial X^\nu}{\partial X'^\mu}$ which acts on the coordinate index, or both (as shown below):

$$e'^a_\mu = \Lambda^a_b e^b_\nu \frac{\partial X^\nu}{\partial X'^\mu} \tag{8.4}$$

Remark 8.3 While the tetrad behaves like a vector under general coordinate transformations with respect to the spacetime index μ, it does not transform purely under general coordinate transformations. For the internal Lorentz index a, the tetrad transforms as a vector *under local Lorentz transformations* instead. From the tetrad field's failure to transform purely under general coordinate transformations, it then follows that a tetrad field does not transform tensorially. It is thus *not* a tensor. Similarly, as the tetrad field does not transform purely under Lorentz transformations, it is *not* a spinor either. In tetrad formalism, the Lorentz transformation is a 'local' analogue of the more 'global' general coordinate transformation.

Theorem 8.2 (Tetrad field properties) A tetrad field observes the following properties:

- Contractions:
$$e_i^\mu e_\mu^j = \delta_j^i \quad e_i^\mu e_\nu^i = \delta_\nu^\mu \tag{8.5}$$
where δ_j^i and δ_ν^μ are the all-too-familiar Kronecker delta.
- Transformation between a coordinate metric and a tetrad metric using two tetrad fields:
$$\gamma_{ab} = e_a^c e_d^b g_{cd} \tag{8.6}$$

Exercise 8.1 Suppose we encode a tetrad field as

$$e^0_\mu dX^\mu = \left(1 - \frac{2M}{r}\right)^{1/2} dt \quad e^1_\mu dX^\mu = \left(1 - \frac{2M}{r}\right)^{-1/2} dr \quad e^2_\mu dX^\mu = r d\theta \quad e^3_\mu dX^\mu = r\sin\theta d\phi \tag{8.7}$$

Show that it transforms the coordinate Schwarzschild metric into the tetrad metric/Minkowski metric η_{ab}.

Remark 8.4 This is only but one example. Quite often, we want our set of coordinates in GR to look like those in SR. i.e. we want a local spacetime that looks like flat Minkowski space. This can simplify calculations, particularly when dealing with spinor fields or when working in contexts where a local inertial frame is advantageous.

8.2 Tetrad derivatives

Definition 8.4 (Tetrad derivative) We define the tetrad derivative with respect to a coordinate derivative:

$$\partial_m = \gamma_m \cdot \partial = \gamma_m \cdot e^\mu \frac{\partial}{\partial X^\mu} = e^\mu_m \frac{\partial}{\partial X^\mu} \tag{8.8}$$

Unlike coordinate derivatives, they do not commute:

Theorem 8.3 (Tetrad derivative commutation relation)

$$[\partial_m, \partial_n] = e^\nu_m \frac{\partial e^\nu_n}{\partial X^\nu} \frac{\partial}{\partial X^\mu} - e^\nu_n \frac{\partial e^\nu_m}{\partial X^\nu} \frac{\partial}{\partial X^\mu} \tag{8.9}$$

which is *not* a tetrad tensor.

We can shorten this as

$$[\partial_m, \partial_n] = (d^k_{mn} - d^k_{nm})\partial_k \tag{8.10}$$

where we have the *tetrad field derivative*:

Definition 8.5 (Tetrad field derivative)

$$d^k_{mn} = e^\mu_m e^\nu_n \frac{\partial e^k_\mu}{\partial X^\nu} \tag{8.11}$$

This is again *not* a tetrad tensor.

With these in mind, we can consider recreating the covariant derivative in the tetrad formalism.

Derivation 8.1 (Tetrad covariant derivative) Consider a tetrad 4-vector A^m and its abstract 4-vector $A = \gamma_m A^m$. The directional derivative is

$$\partial_n A = \gamma_m \partial_n A^m + (\partial_n \gamma_m) A^m \tag{8.12}$$

Note that we are still working with the tetrad basis. We convert the directional derivative to the coordinate basis:

$$\partial_n \gamma_m = e^\mu_n \frac{\partial \gamma_m}{\partial X^\mu} \tag{8.13}$$

We define the directional derivative of the tetrad basis as a *connection* and contract. This yields

$$\partial_n \gamma_m = \Gamma^k_{mn} \gamma_k \tag{8.14}$$

The original direction derivative of the abstract tensor then becomes

$$\partial_n A = \gamma_k (\nabla_n A^k) \tag{8.15}$$

We thus find that the covariant derivative works almost identically in a tetrad frame. The only difference is that we are using tetrad Christoffels instead of coordinate Christoffels.
But wait! How do we derive tetrad Christoffels from coordinate Christoffels? Consider the derivative

$$\frac{\partial e_\mu}{X^\nu} = \Gamma^\kappa_{\mu\nu} e_\kappa = \frac{\partial e^m_\mu \gamma_m}{\partial X_\nu} \tag{8.16}$$

This becomes

$$\Gamma^\kappa_{\mu\nu} e_\kappa = e^m_\mu e^n_\nu (d^k_{mn} + \Gamma^k_{mn}) \gamma_k \tag{8.17}$$

which yields the relation

Theorem 8.4 (Tetrad Christoffel-coordinate Christoffel relation)

$$d_{lmn} + \Gamma_{lmn} = e_l^\lambda e_m^\mu e_n^\nu \Gamma_{\lambda_\mu \nu} \tag{8.18}$$

where γ_{lmn} is simply a tetrad Christoffel of the first kind. The commonly-used tetrad Christoffel of the second kind can expectedly be recovered via

$$\Gamma_{lmn} = \gamma_{lk} \Gamma^k_{mn} \tag{8.19}$$

8.3 Newman-Penrose formalism: the null tetrad

As interesting as it is, the tetrad formalism is a very generalised technique, and any tetrad basis can be chosen. In 1962, Ezra T. Newman and Roger Penrose devised a notation system based on the so-called *null tetrad*. This is the *Newman-Penrose formalism* or *NP formalism*, which is a subset of the tetrad formalism.

We first recall what a null vector in general is:

Definition 8.6 (Null vector) If the sum of the squares of the components of a vector is zero, it is called a *isotropic vector* or a *null vector*.

Remark 8.5 Compare with the null geodesics, for which the line element (which stems from the metric) observes $ds^2 = 0$.

We can construct a set of tetrad coordinates in which the tetrad basis is null (light-like). This is accomplished by converting the 4 coordinates (i.e. the existing tetrad (t, x, y, z)) to the null tetrad:

Definition 8.7 (Null tetrad) The null tetrad or the *Newman-Penrose null tetrad* $(k, l, m, \bar{m})$ is a tetrad basis that consists of four vectors. The first two are null vectors

$$k = \frac{1}{\sqrt{2}}(t + z) \quad l = \frac{1}{\sqrt{2}}(t - z) \tag{8.20}$$

and the second two are a complex vector and its conjugate

$$m = \frac{1}{\sqrt{2}}(x - iy) \quad \bar{m} = \frac{1}{\sqrt{2}}(x + iy) \tag{8.21}$$

Recalling the tetrad basis-tetrad field relation, we know they are related to the tetrad field via

$${e_0}^\mu = l^\mu \quad {e_1}^\mu = n^\mu \quad {e_2}^\mu = m^\mu \quad {e_3}^\mu = \bar{m}^\mu \tag{8.22}$$

The metric is then

$$g_{ab} = -l_a n_b - n_a l_b + m_a \bar{m}_b + \bar{m}_a m_b \tag{8.23}$$

Remark 8.6 Every vector in the null tetrad can be represented in spinors (or 'spin notation' in the original paper), and this is why the NP formalism is a spinor-based formalism. In practice, however, this is very rarely used, and we represent each vector as a vector instead of a dual spinor-spinor pair.

Theorem 8.5 (Null tetrad properties) The null tetrad satisfies the orthogonality and normalisation conditions:

$$l^a k_a = -1, \quad m^a \bar{m}_a = 1, \quad l^a m_a = l^a \bar{m}_a = k^a m_a = k^a \bar{m}_a = 0 \tag{8.24}$$

Theorem 8.6 (Null tetrad transformations) A null tetrad can undergo the following transformations. Together, these represent the six-parameter group of Lorentz transformations.

- **Boost transformation:**

$$k' = Bk, \quad l' = B^{-1} l, \quad m' = m, \quad \bar{m}' = \bar{m} \tag{8.25}$$

where B is a real, positive *boost parameter*.

- **Spin transformation:**

$$k' = k, \quad l' = l, \quad m' = e^{i\Phi}m, \quad \bar{m}' = e^{-i\Phi}\bar{m} \tag{8.26}$$

where Φ is a real phase parameter.

- **Null rotation about k (parameter L):**

$$k' = k, \quad l' = l + L\bar{m} + \bar{L}m + L\bar{L}k, \quad m' = m + Lk, \quad \bar{m}' = \bar{m} + \bar{L}k \tag{8.27}$$

where L is a complex parameter representing a null rotation about the vector k.

- **Null rotation about l (parameter K):**

$$k' = k + K\bar{m} + \bar{K}m + K\bar{K}l, \quad l' = l, \quad m' = m + Kl, \quad \bar{m}' = \bar{m} + \bar{K}l \tag{8.28}$$

where K is a complex parameter representing a null rotation about the vector l.

8.4 Weyl scalars and their implications

From the tetrad $(k, l, m, \bar{m})$, we can encode the 10 independent components of the Ricci tensor through the following scalar quantities called the *Newman-Penrose scalars*:

Definition 8.8 (Newman-Penrose scalars) Real NP scalar components:

$$\Phi_{00} = \frac{1}{2}R_{ab}k^a k^b \quad \Phi_{11} = \frac{1}{2}R_{ab}(k^a l^b + m^a \bar{m}^b) \quad \Phi_{22} = \frac{1}{2}R_{ab}l^a l^b \quad \Lambda = \frac{R}{24} \tag{8.29}$$

Complex NP scalar components:

$$\Phi_{01} = \bar{\Phi}_{10} = \frac{1}{2}R_{ab}k^a m^b \quad \Phi_{02} = \bar{\Phi}_{20} = \frac{1}{2}R_{ab}m^a m^b \quad \Phi_{12} = \bar{\Phi}_{21} = \frac{1}{2}R_{ab}l^a m^b \tag{8.30}$$

The Weyl tensor C_{abcd} can also be decomposed into scalar components with respect to the null tetrad. These components are known as the *Weyl scalars* and break down the gravitational wave/radiation into more manageable chunks:

Definition 8.9 (Weyl scalars)

$$\Psi_0 = C_{abcd}l^a m^b l^c m^d \tag{8.31}$$

This is a transverse component of the gravitational field/wave propagating in the l^a-direction.

$$\Psi_1 = C_{abcd}l^a k^b l^c m^d \tag{8.32}$$

This is a longitudinal component of the gravitational field/wave in the l^a-direction.

$$\Psi_2 = C_{abcd}l^a m^b \bar{m}^c k^d \tag{8.33}$$

This is a Coulomb-like component that represents a sort of a 'gravitational charge' that decreases with distance. It represents the dominant, non-radiative part of the gravitational field, akin to the electrostatic field of a point charge and the Newtonian gravitational potential.

$$\Psi_3 = C_{abcd}l^a k^b \bar{m}^c k^d \tag{8.34}$$

This is a longitudinal component of the gravitational field/wave in the k^a-direction.

$$\Psi_4 = C_{abcd}k^a \bar{m}^b k^c \bar{m}^d \tag{8.35}$$

This is a transverse component of the gravitational field/wave propagating in the k^a-direction.

Remark 8.7 Remember that a gravitational wave is a wave. The transverse and longitudinal components are the same as in any other wave.

Remark 8.8 Iin standard GR and typical vacuum gravitational waves, the significance longitudinal components Ψ_1 and Ψ_3 are constrained to the context of tidal forces due to being much smaller than the transverse components Ψ_0 and Ψ_4. However, the gravitational field can gain a significant longitudinal component in the case of near sources[2] of strong gravitational waves, complex spacetimes and certain modified gravity theories[3].

Remark 8.9 The Weyl tensor is not always zero in the original metric, while it is always so in the Minkowski metric (i.e. the NP tetrad metric). This is not inconsistent as the tetrad metric is only a local descriptor while the original metric is a global one. This links to the purpose of the NP formalism, which is to describe a space as flat locally. In the NP formalism, instead of the Weyl tensor, it is the Weyl scalars that encode conformal curvature - otherwise this information would have been lost.

From the Weyl scalars, we can then find the *Petrov type* of a specific spacetime.

Theorem 8.7 (Petrov classification) In the *Petrov classification*, also known as the *classification of gravitational fields*, we have the following possible types:

- **Type I:**

$$\Psi_0 = 0 \tag{8.36}$$

- **Type II:**

$$\Psi_0 = \Psi_1 = 0 \tag{8.37}$$

  Type II regions combine the effects for types D, III, and N in a rather complicated nonlinear way.

- **Type III:**

$$\Psi_0 = \Psi_1 = \Psi_2 = 0 \tag{8.38}$$

  Type III regions are associated with a kind of longitudinal gravitational radiation. In such regions, the tidal forces have a shearing effect. This possibility is often neglected, in part because the gravitational radiation which arises in weak-field theory is type N, and in part because type III radiation decays faster than type N radiation. Certain Robinson/Trautman vacuums are everywhere type III.

- **Type N:**

$$\Psi_0 = \Psi_1 = \Psi_2 = \Psi_3 = 0 \tag{8.39}$$

  Type N regions are associated with transverse gravitational radiation, which is the type astronomers have detected with LIGO. The quadruple principal null direction corresponds to the wave vector describing the direction of propagation of this radiation. The long-range radiation field is type N. The pp-wave spacetimes are everywhere type N.

- **Type D:**

$$\Psi_0 = \Psi_1 = \Psi_3 = \Psi_4 = 0 \tag{8.40}$$

  Type D regions are associated with the gravitational fields of isolated massive objects, such as stars. More precisely, type D fields occur as the exterior field of a gravitating object which is completely characterised by its mass and angular momentum. The Kerr vacuum is everywhere type D.

- **Type O:**

$$\Psi_0 = \Psi_1 = \Psi_2 = \Psi_3 = \Psi_4 = 0 \tag{8.41}$$

  In type O, the Weyl tensor vanishes. The FLRW models are everywhere type O.

[2] In which case the waves could induce stresses in matter.

[3] In which they could manifest as additional polarisation modes of gravitational waves.

Chapter 9

Taub-NUT spacetime

Quote 9.1 Avoid fruit and nuts. You are what you eat.

Jim Davis, in Garfield

The Taub-NUT spacetime is a specific solution of Einstein's field equation that investigates gravitomagnetic properties of spacetime. The underlying Taub space was found by Abraham Haskel Taub, and extended to a larger manifold by Ezra T. Newman, Louis A. Tamburino, and Theodore W. J. Unti - hence the name 'Taub-NUT'. In this chapter, we will introduce the Taub-NUT solution and investigate some of its problems.

9.1 Taub-NUT solution

Definition 9.1 (Taub-NUT metric)

$$ds^2 = -f(r)(dt - 2l\cos\theta d\phi)^2 + \frac{dr^2}{f(r)} + (r^2 l^2)(d\theta + \sin^2\theta d\phi^2) \tag{9.1}$$

where one has

$$f(r) = \frac{r^2 - 2Mr - l^2}{r^2 + l^2} \tag{9.2}$$

M is the mass and l is a sort of a twist parameter called the NUT charge or the NUT parameter.

Remark 9.1 When $l = 0$ and $m \neq 0$, the metric reduces to the Schwarzschild solution.
We note that the metric is not well-behaved at both $\theta = 0$ and $\theta = \pi$. However, we can eliminate this by performing $\bar{t} = t + 2l\phi$. The metric then reads

$$ds^2 = -f(r)\left(dt - 4l\sin\frac{\theta}{2}d\phi\right)^2 + \frac{dr^2}{f(r)} + (r^2 l^2)(d\theta + \sin^2\theta d\phi^2) \tag{9.3}$$

which is the more commonly used form.
For $f(r) = 0$, r has two solutions:

$$r_\pm = m \pm \sqrt{m^2 + l^2} \tag{9.4}$$

which corresponds to the two Killing horizons - null surfaces on which the Killing vectors vanish.
We have $f(r) > 0$ for $r < r_-$ and $r > r_+$, in which case we have a spacelike r coordinate and a stationary metric. these two regions are called the NUT_- and NUT_+, or the *NUT regions* as a whole. Conversely, the region $r_- < r < r_+$ is known as the *Taub region*.

Definition 9.2 (Taub-NUT metric in LWP form) In stationary axisymmetric systems, we can represent the Taub-NUT metric in the LWP form:

$$ds^2 = -e^{2U}(d\bar{t} + Ad\phi)^2 + e^{-2U}(e^{2\gamma}(d\rho^2 + dz^2) + \rho^2 d\theta^2) \tag{9.5}$$

where the NUT solution satisfies

$$e^{2U} = \frac{(R_+ + R_-)^2 - 4(m^2 + l^2)}{(R_+ + R_- + 2m)^2 + 4l^2} \quad A = -\frac{l}{\sqrt{m^2 + l^2}}(R_+ - R_-) \tag{9.6}$$

in which

$$e^{2\gamma} = \frac{(R_+ R_-)^2 - 4(m^2 + l^2)}{4R_+ R_-} \quad R_\pm^2 = \rho^2 + (z \pm \sqrt{m^2 + l^2})^2 \tag{9.7}$$

Exercise 9.1 Prove that one can return the Taub-NUT metric from the LWP form to its original form using the coordinate transformations

$$\rho = \sqrt{r^2 - 2mr - l^2}\sin\theta \quad z = (e - m)\cos\theta \tag{9.8}$$

Hint: One then yields $R_\pm = r - m \pm \sqrt{m^2 + l^2}\cos\theta$.

9.2 Rise of gravitoelectromagnetism

The gravitational field produced by a rotating object (or indeed any rotating mass-energy) can be described by equations that resemble Maxwell's equations. This is called *gravitoelectricmagnetism* or GEM. It is especially useful in weak-field, slow-motion scenarios, where it simplifies the analysis of gravitational effects without solving the full Einstein field equations.

Remark 9.2 Despite the name, it has *nothing* to do with actual electromagnetism. Rather it is a formalism that makes gravitation *look* like electromagnetism.

Derivation 9.1 (GEM equations) The interesting parameter in the Taub-NUT solution is the so-called NUT charge l:

- In general l acts as a source of the gravitomagnetic field. It gives rise to a *gravitomagnetic monopole* which leads to a rotational effect similar to the one produced by a magnetic monopole in electromagnetism.
- If we expand the first term of the Taub-NUT metric, we will find off-diagonal terms containing l. This introduces a gravitomagnetic vector potential to the metric. This is a gravitational analogue to the EM vector potential:

$$A^i = (0, 0, 0, 2l\cos\theta) \to A^\phi = 2l\cos\theta \to A = 2l\cos\theta\hat{\phi} \tag{9.9}$$

Remark 9.3 Just like the magnetic monopole, no gravitomagnetic monopoles have so far been discovered.

We recall that the gravitational (scalar) potential is naturally $\Phi = -\frac{Gm}{r}$. We can then construct a 'gravitoelectric field' E_g:

$$E_g = -\nabla\Phi - \frac{\partial A}{\partial t} = \frac{Gm}{r^2}\hat{r} \tag{9.10}$$

The 'gravitomagnetic field' is thus

$$B_g = \nabla \times A = \frac{4l}{r^2}\hat{r} \tag{9.11}$$

Theorem 9.1 (GEM equations) Structure equations:

$$\nabla \cdot B_g = 0 \quad \nabla \times E_g = -\frac{\partial B_g}{\partial t} \tag{9.12}$$

Source equations:

$$\nabla \cdot E_g = -4\pi G\rho_g \quad \nabla \times B_g = -4\pi G J_g + \frac{\partial E_g}{\partial t} \tag{9.13}$$

where ρ_g is the mass density.

9.3 Misner strings

We consider the first term in the Taub-NUT metric

$$-f(r)(dt - 2l\cos\theta d\phi)^2 \tag{9.14}$$

At $\theta = 0$ and $\theta = \pi$, the term $2l\cos\theta$ becomes $\pm 2n$ respectively. Recalling the periodic nature of $2l\cos\theta$, we see that singularities arise at $\theta = 0$ and $\theta = \pi$.

Definition 9.3 (Misner string) The *Misner strings* are lines of singularity that extend from $r = 0$ to $r = \infty$ at $\theta = 0$ and $\theta = \pi$.

They are regions where the space-time has non-trivial topological structures, or physically, the 'twist' in space-time introduced by the NUT parameter. They are required to account for the gravitomagnetic monopole's NUT charge.
Remark 9.4 This is analogous to the Dirac string, which accounts for the magnetic monopole's magnetic charge.
In some extensions of general relativity, efforts are made to remove or regularise Misner strings.

9.4 Closed timelike curves

We can avoid Misner strings by considering curves on which t, r and θ are constant. We have the metric

$$da^2 = -(4l^2 f(r)(1 - cos\theta)^2) - (r^2 + l^2)\sin^2\theta) d\theta^2 \tag{9.15}$$

when $f(r) \leq 0$, such intevrals are spacelike. However, when $f(r) > 0$, they become timelike, and

$$\cos\theta < -\frac{r^2 + l^2 - 4l^2 f}{r^2 + l^2 + 4l^2 f} \tag{9.16}$$

As ϕ is a periodical coordinate, these timelike curves loop back to themselves. They are thus called *closed timelike curves* (CTCs).
Remark 9.5 This effectively gives rise to the possibility of *time travel*.

Figure 9.1: Einstein after travelling through a CTC.

Quote 9.2 Most relativists still regard the existence of regions containing closed timelike curves with suspicion. Yet it would appear that they occur naturally and regularly within many families of exact solutions of Einstein's equations.

Jerry B. Griffiths and Jiří Podolský

Chapter 10

Stationary, axially symmetric solutions

Quote 10.1 The usefulness of solution generating techniques in general relativity is renowned.

Marco Astorino, Riccardo Martelli, and Adriano Viganò, in Black holes in a swirling universe

Another significant family of solutions are stationary, axially symmetric solutions. These solutions are interesting in that one can, from the basic Weyl metric, generate more specific solutions using so-called solution generating techniques.

10.1 Weyl metric

Stationary, axially symmetric solutions are spacetimes which exhibit both time-translation symmetry and rotational symmetry. They are described by the so-called *Weyl metric*, which can be generated by the commuting Killing vectors ∂_t and ∂_ϕ.

Remark 10.1 If one recalls the shorthand $\partial_a = \frac{\partial}{\partial X^a}$, they will find that the Killing vectors are simply tangent vectors to the unit directions.

Definition 10.1 (Weyl metric)

$$ds^2 = -e^{2U}dt^2 + e^{-2U}(e^{2\gamma}(d\rho^2 + dz^2) + \rho^2 d\phi^2) \tag{10.1}$$

where U and γ are functions of *Weyl's canonical coordinates* ρ and z. They are often called the *metric functions*. γ is the *conformal factor* which ensures that the metric satisfies the Einstein equations in the (ρ, z) plane.

Definition 10.2 (Weyl's canonical coordinates) In Weyl's canonical coordinates:

- t is unsurprisingly time.
- ρ is the cylindrical radial distance from the axis of symmetry.
- z is the position along the axis of symmetry.
- ϕ is the angle around the axis of symmetry z.

They are useful for stationary (i.e. static) and axially symmetric spacetimes.

Remark 10.2 Note that this is essentially a cylindrical coordinate system with an extra time dimension. Because of this it is also called the *cylindrical Weyl coordinates*.

For the more general case in which the sources are rotating but their fields remain stationary, the line element may be extended to the *Lewis-Weyl-Papapetrou form* or *LWP form*:

Definition 10.3 (Lewis-Weyl-Papapetrou form)

$$ds^2 = -f(dt + Ad\phi)^2 + f^{-1}(e^{2\gamma}(d\rho^2 + dz^2) + \rho^2 d\phi^2) \tag{10.2}$$

where f and A are yet more metric functions of ρ and z:

- f is the *lapse function* that is the tt component of the metric.
- A is the *shift vector component* which describes how much space-time is dragged around by the rotation of the mass. Note that it is simply a component of the *vector potential.*

10.2 Vacuum, axially symmetric solutions and the Ernst equation

As it turns out, when solving for axially symmetric solutions of the Einstein field equations, the field equations can be replaced with the so-called *Ernst equation.*
Here we come across the concept of *solution generating techniques*, which simply means deriving more interesting solutions from a generic 'base metric' which we call a *seed.*
Remark 10.3 If you find this poorly defined, you'd be right.

Derivation 10.1 (Ernst equation) We use the LWP form as a seed and plug it into the vacuum field equations. This yields

$$\left(\rho^{-1}f^2\, A_{,\rho}\right)_{,\rho} + \left(\rho^{-1}f^2\, A_{,z}\right)_{,z} = 0 \quad \gamma_{,z} = \frac{1}{2}\rho f^{-2} f_{,\rho} f_{,z} - \frac{1}{2}\rho^{-1} f^2\, A_{,\rho} A_{,z} \tag{10.3}$$

We can define a 'potential' function such that

$$\varphi_{,\rho} = -\rho^{-1}f^2\, A_{,z} \quad \varphi_{,z} = \rho^{-1}f^2\, A_{,\rho} \tag{10.4}$$

which allows us to rewrite the equations as

$$f\nabla^2 f - f_{,\rho}^2 - f_{,z}^2 + \varphi_{,\rho}^2 + \varphi_{,z}^2 \tag{10.5}$$

$$f\nabla^2\varphi - 2f_{,\rho}\varphi_{,\rho} - 2f_{,z}\varphi_{,z} = 0 \tag{10.6}$$

where we have the Laplacian $\nabla^2 = \rho^{-1}\partial_\rho(\rho\partial_\rho) + \partial_z^2$. We introduce the *Ernst potential*

$$\mathcal{E} = f + i\varphi \tag{10.7}$$

which is basically a complex-valued gravitational potential[a]. φ is the *twist potential* and is related to the rotational aspect of the gravitational field.
The equations are then reduced to

Theorem 10.1 (Ernst equation)

$$\mathrm{Re}(\mathcal{E})\nabla^2\mathcal{E} = 2(\nabla\mathcal{E})^2 \tag{10.8}$$

where $\bar{\mathcal{E}}$ is the complex conjugate of $\mathcal{E}$.

Remark 10.4 In spinor form the Ernst equations are integrable. This is satisfied in both a vacuum and an EM field. The gradient and the Laplacian can then be derived:

$$(\nabla\mathcal{E})^2 = g^{ab}\mathcal{E}_{,a}\mathcal{E}_{,b} \quad \nabla^2\mathcal{E} = (g^{ab}\mathcal{E}_{,b})_{;a} = \frac{1}{\sqrt{-g}}(\sqrt{-g}g^{ab}\mathcal{E}_{,b})_{,a} \tag{10.9}$$

where we have the shorthand $T_{;i} = \nabla_i T$.

[a] $\mathcal{E}$ being complex is merely a mathematical trick/convenience.

Remark 10.5 Since the Ernst equation is integrable, a huge number of solutions are now known and others can straightforwardly be constructed.
But how exactly do we recover the metric from this?

Derivation 10.2 (Ernst generating technique) To recover the metric, we simply solve for all the metric functions present in the LWP form. This process is known as the *Ernst generating technique*:

- f is easily solved from the definition of the Ernst potential:

$$f = \mathrm{Re}(\mathcal{E}) \tag{10.10}$$

- We can solve for A from φ bu recalling their relation, which gives

$$\nabla A = \frac{\rho}{f^2}(\partial_z \varphi \hat{\rho} - \partial_\rho \varphi \hat{z}) \tag{10.11}$$

- γ can be found by integrating the equations:

$$\partial_\rho \gamma = \frac{\rho}{4f^2}\left[(\partial_\rho f)^2 - (\partial_z f)^2 + (\partial_\rho A)^2 - (\partial_z A)^2\right] \tag{10.12}$$

We can then plug these back into the LWP form.

A (sometimes more convenient) form of the equation involves relabelling $\mathcal{E}$ with ξ[1]:

$$\mathcal{E} = \frac{\xi - 1}{\xi + 1} \tag{10.13}$$

The Ernst equation then becomes

$$(\xi\bar{\xi} - 1)\nabla^2 \xi = 2\bar{\xi}(\nabla\xi)^2 \tag{10.14}$$

One may also be tempted to use a different set of coordinates (x, y):

$$\rho = \sqrt{x^2+1}\sqrt{1-y^2} \quad z = xy \tag{10.15}$$

This becomes

$$x = \frac{R_+ + R_-}{2} \quad y = \frac{R_+ - R_-}{2} \tag{10.16}$$

where $R_\pm^2 = \rho^2 + (z \pm 1)^2$.

10.3 Axially symmetric electrovacuum space-times

Derivation 10.3 (Ernst equation in an electrovacuum field) We can generalise the Ernst equation to include a non-zero electrovacuum field. We consider the Einstein-Maxwell equations, in which the EM field dominates the matter component.

$$G_{ab} = R_{ab} - \frac{1}{2}g_{ab}R = \kappa\left(F_{ac}F_b^c - \frac{1}{4}\eta_{ab}F_{mn}F^{mn}\right) \tag{10.17}$$

For the unitary U(1) gauge, we can consider the Faraday tensor to be

$$F_{ab} = \partial_a A_b - \partial_b A_a \tag{10.18}$$

where we have two commuting Killing vectors ∂_a and ∂_b and the EM 4-potential A_a. Its only non-zero components are A_t (the electric potential) and A_ϕ. Solving for the magnetic potential A'_ϕ is slightly trickier:

$$A'_{\phi,\rho} = \rho^{-1} f(A_{\phi,z} - AA_{t,z}) \quad A'_{\phi,z} = -\phi^{-1} f(A_{\phi,\rho} - AA_{t,\phi}) \tag{10.19}$$

where we can recall that $fA = -g_{t\phi}$. The complex potential is therefore

$$\Phi = A_t + iA'_\phi \tag{10.20}$$

The Ernst potential, now incorporating this complex EM potential, is

$$\mathcal{E} = f - \Phi\bar{\Phi} + i\varphi \tag{10.21}$$

[1] Not the Killing vector!

The equations involving the twist potential become

$$(\mathcal{E}+\bar{\mathcal{E}}+2\Phi\bar{\Phi})\nabla^2\mathcal{E} = 2\left(\nabla\mathcal{E}+2\bar{\Phi}\nabla\Phi\right)\cdot\nabla\mathcal{E} \tag{10.22}$$

$$(\mathcal{E}+\bar{\mathcal{E}}+2\Phi\bar{\Phi})\nabla^2\Phi = 2\left(\nabla\mathcal{E}+2\bar{\Phi}\nabla\Phi\right)\cdot\nabla\Phi \tag{10.23}$$

The Einstein field equations (now the *Einstein-Maxwell field equations* due to the inclusion of the EM field) is

Theorem 10.2 (Ernst equation in an electrovacuum field)

$$(\mathcal{E}+\bar{\mathcal{E}}+2\Phi\bar{\Phi})\nabla^2\mathcal{E} = 2(\nabla\mathcal{E}+2\bar{\Phi}\nabla\Phi)\cdot\nabla\mathcal{E} \tag{10.24}$$

$$(\mathcal{E}+\bar{\mathcal{E}}+2\Phi\bar{\Phi})\nabla^2\Phi = 2(\nabla\mathcal{E}+2\bar{\Phi}\nabla\Phi)\cdot\nabla\Phi \tag{10.25}$$

When the EM field vanishes, this simply reduces to the normal Ernst equations.

To recover the metric, we still need γ, which can be calculated via

$$\gamma_{\rho,\rho} = \frac{1}{4}\rho f^{-2}\left[\left(\mathcal{E}_{,\rho}+2\Phi\bar{\Phi}_{,\rho}\right)\left(\bar{\mathcal{E}}_{,\rho}+2\bar{\Phi}\Phi_{,\rho}\right)-\left(\mathcal{E}_{,z}+2\Phi\bar{\Phi}_{,z}\right)\left(\bar{\mathcal{E}}_{,z}+2\bar{\Phi}\Phi_{,z}\right)\right]-\rho f^{-1}\left(\Phi_{,\rho}\bar{\Phi}_{,\rho}-\Phi_{,z}\bar{\Phi}_{,z}\right) \tag{10.26}$$

$$\gamma_{\rho,z} = \frac{1}{4}\rho f^{-2}\left[\left(\mathcal{E}_{,\rho}+2\Phi\bar{\Phi}_{,\rho}\right)\left(\bar{\mathcal{E}}_{,z}+2\bar{\Phi}\Phi_{,z}\right)+\left(\mathcal{E}_{,z}+2\Phi\bar{\Phi}_{,z}\right)\left(\bar{\mathcal{E}}_{,\rho}+2\bar{\Phi}\Phi_{,\rho}\right)\right]-\rho f^{-1}\left(\Phi_{,\rho}\bar{\Phi}_{,z}+\Phi_{,z}\bar{\Phi}_{,\rho}\right) \tag{10.27}$$

Also, we note that the complex fields have some very interesting symmetry properties. In all, there are five coordinate transforms under which the action and the Ernst equation remain invariant.

Theorem 10.3 (Action- & Ernst equation-invariant transforms)

$$\begin{aligned}
\mathrm{G}_1[b] &\quad \mathcal{E}\to\mathcal{E}'=\mathcal{E}+ib \quad \Phi\to\Phi'=\Phi\\
\mathrm{G}_2[\beta] &\quad \mathcal{E}\to\mathcal{E}'=\mathcal{E}-2\beta^*\Phi=\beta\beta^* \quad \Phi\to\Phi'=\Phi+\beta\\
\mathrm{D}[\lambda] &\quad \mathcal{E}\to\mathcal{E}'=\lambda\lambda^*\mathcal{E} \quad \Phi\to\Phi'=\lambda\Phi\\
\mathrm{E}[c] &\quad \mathcal{E}\to\mathcal{E}'=\frac{\mathcal{E}}{1+ic\mathcal{E}} \quad \Phi\to\Phi'=\frac{\Phi}{1+ic\Phi}\\
\mathrm{H}[\alpha] &\quad \mathcal{E}\to\mathcal{E}'=\frac{\mathcal{E}}{1-2\alpha^*\Phi-\alpha\alpha*\mathcal{E}} \quad \Phi\to\Phi'=\frac{\Phi+\alpha\mathcal{E}}{1-2\alpha^*\Phi-\alpha\alpha*\mathcal{E}}
\end{aligned} \tag{10.28}$$

G_1 and G_2 are the so-called *gauge transformations*, which transform the potentials but leave the metric and the gauge field invariant, D is a *duality-rescaling transformation*, E is called the *Ehlers transformation*, while H is called the *Harrison transformation*.

We can combine these transformations and the invariants will stay invariant. For example, applying G_1, D and E yields the *inversion transformation*

$$\mathcal{E}\to\mathcal{E}'=\frac{1}{\mathcal{E}} \quad \Phi\to\Phi'=\frac{\Phi}{\mathcal{E}} \tag{10.29}$$

Remark 10.6 A particular specialisation of this inversion transformation for the null electromagnetic field is known as the *Buchdahl transformation*.

Chapter 11

Swirling universe

Quote 11.1 Definitely, less has been said about the swirling spacetime.

José Barrientos, Adolfo Cisterna, Ivan Kolář, Keanu Müller, Marcelo Oyarzo and Konstantinos Pallikaris

While somewhat well-established, the swirling universe only received significant attention in recent years. Significant work on geodesics in a swirling universe was done by Prof Betti Hartmann, who utilised elliptic integrals to tackle geodesics in the swirling universe.

11.1 Bonnor-Melvin universe

If one performs magnetic Harrison and magnetic Ehlers transformations on the seed potentials of Minkowski spacetime, one obtains two interesting asymptotically nontrivial backgrounds: the Bonnor-Melvin and swirling spacetimes.

We begin by considering the Bonnor-Melvin solution, in which the universe is under a uniform magnetic field (i.e. the EM tensor $F_{ab} \neq 0$). This solution was discovered by William B. Bonnor and rediscovered by Mael Avrami Melvin. In Weyl cylindrical coordinates:

Definition 11.1 (Bonnor-Melvin metric)

$$ds^2 = V^2(-dt^2 + dz^2 + d\rho^2) + \frac{\rho^2}{V^2}d\phi^2 \tag{11.1}$$

where $V(\rho) = 1 + \frac{1}{4}B^2\rho^2$.

We can generalise this for an EM field. The resultant metric is called the *electromagnetic universe*:

Definition 11.2 (EM universe metric)

$$ds^2 = V^2(-dt^2 + dz^2 + d\rho^2) + \frac{\rho^2}{V^2}d\phi^2 \tag{11.2}$$

where $V(\rho) = 1 + X\bar{X}\rho^2$. We have $X = \frac{E+iB}{2}$, and that $\bar{X}$ is the complex conjugate of X.

This metric belongs to the Petrov type D class.

11.2 Swirling universe

In a *swirling universe*, the 'background' (i.e. the universe itself) is rotating.

Definition 11.3 (Swirling metric)

$$ds^2 = \frac{\rho^2}{1 + j^2\rho^4}(d\phi + 4jzdt)^2 + (1 + j^2\rho^4)(-dt^2 + d\rho^2 + dz^2) \tag{11.3}$$

where j is the *swirling parameter* (as named by Betti Hartmann and Jutta Kunz) and determines the ergoregion of the swirling universe.

This 'toy model' is specific in that in the $+z$ direction, the universe swirls in one direction and the $-z$ direction in the other.

Quote 11.2 This is almost like a 'twirling universe'.

Betti Hartmann, July 2024

This metric also belongs to the Petrov type D class.

Remark 11.1 The physical significance of the swirling parameter is not known: it has been called an *anti-NUT parameter* by Plebański because of its resemblance with the NUT parameter in the Plebański-Demiański spacetime.

Remark 11.2 When t is reversed in the swirling metric, the metric is not invariant. But if the same is donein the Bonnor-Melvin and EM universes, the metric stays the same.

One can even combine the swirling and EM universes. This yields the *electromagnetic swirling universe*. This is accomplished by imposing both the Ehlers and Harrison transforms.

Definition 11.4 (Electromagnetic swirling metric)

$$ds^2 = \frac{\rho^2}{V^2 + j^2\rho^4}(d\phi + 4jzdt)^2 + (V^2 + j^2\rho^4)(-dt^2 + d\rho^2 + dz^2) \tag{11.4}$$

11.3 Elliptic functions

We can represent swirling geodesics in elliptic functions. We begin by defining an *elliptic curve*:

Definition 11.5 (Elliptic curve) An elliptic curve over a field K satisfies

$$y^2 = x^3 + ax + b \tag{11.5}$$

and is constrained by the determinant

$$\Delta = -16(4a^3 + 27b^2) \neq 0 \tag{11.6}$$

This ensures that the curve is non-singular (i.e. smooth).

Remark 11.3 An elliptic curve has genus[1] $g = 1$. This can be understood as the curve resembling a torus (a doughnut shape), which has one 'hole'.

Like all rather complicated curves, we want to use something to parametrise it. This gives rise to *elliptic functions*. These are very peculiar functions which are doubly periodic. i.e. they have two distinct periods.

Definition 11.6 (Elliptic function) A function $f(z)$ is elliptic if there exist two non-zero complex numbers ω_1 and ω_2 (periods) such that:

$$f(z + \omega_1) = f(z) \tag{11.7}$$

$$f(z + \omega_2) = f(z) \tag{11.8}$$

for all complex numbers z and where ω_1 and ω_2 must be linearly independent over the real numbers $\mathbb{R}$. They are meromorphic.

Remark 11.4 This periodicity implies that the function repeats its values in a lattice pattern in the complex plane. A notable example of an elliptic function is the so-called *Weierstrass $\wp$-function*.

[1] The number of 'holes' in the geometry.

Definition 11.7 (Weierstrass $\wp$-function) A Weierstrass[a] $\wp$-function or a *Weierstrass elliptic function* $\wp(z)$ is defined by

$$\wp(z, \omega_1, \omega_2) = \wp(z) = \frac{1}{z^2} + \sum_{\lambda \in \Lambda \setminus \{0\}} \left(\frac{1}{(z-\lambda)^2} - \frac{1}{\lambda^2} \right) \tag{11.9}$$

where the *period lattice* λ is defined by

$$\Lambda = \{m\omega_1 + n\omega_2\} \quad \text{where} \quad m, n \in \mathbb{Z} \tag{11.10}$$

[a]Named after Karl Weierstrass, or *Weierstraß* in German spelling.

Theorem 11.1 (Differential equation) The Weierstrass $\wp$-function satisfies the following nonlinear differential equation

$$\wp'(z)^2 = 4\wp(z)^3 - g_2\wp(z) - g_3 \tag{11.11}$$

This equation is fundamental in the study of elliptic functions and elliptic curves.

We can use Weierstrass elliptic functions to parametrise elliptic curves. This comes in the form

$$y^2 = 4x^3 - g_2x - g_3 \tag{11.12}$$

Another set of elliptic functions are the Jacobi elliptic functions, denoted as sn, cn, and dn, which also exhibit doubly periodic behavior. We can generalise the concept of elliptic curves to higher dimensions as *hyperelliptic curves*.

Definition 11.8 (Hyperelliptic curve) A hyperelliptic curve of genus g satisfies

$$y^2 = f(x) \tag{11.13}$$

where $f(x)$ is a polynomial of degree $2g + 1$ or $2g + 2$ with distinct roots.

Remark 11.5 For genus 1, hyperelliptic curves reduce to elliptic curves.
Likewise, *hyperelliptic functions* generalise elliptic functions to higher-genus curves. They parametrise hyperelliptic curves.

11.4 Elliptic integrals

Historically, the study of elliptic integrals led to the development of elliptic functions. Elliptical integrals are a class of integrals that arise in the calculation of the arc length of an ellipse. They cannot generally be expressed in terms of elementary functions.

Definition 11.9 (Elliptic integral of the first kind) Elliptic integrals can be generalised as

$$f(x) = \int_c^x R(t, \sqrt{P(t)})dt \tag{11.14}$$

where R is a rational function of its two arguments, P is a polynomial of degree 3 or 4 with no repeated roots, and c is a constant.

There are three main types of elliptical integrals, all of which depend on two variables: an *amplitude* (or *angle*) ϕ and the *elliptic modulus* or *modulus* k[2].

Definition 11.10 (Incomplete elliptic integral of the first kind)

$$F(\phi, k) = \int_0^\phi \frac{d\theta}{\sqrt{1 - k^2 \sin^2\theta}} \tag{11.15}$$

This integral represents the arc length of an ellipse from the major axis to an angle ϕ.

[2]Sometimes the parameter $m = k^2$ is used instead. This footnote should not be read as an exponential.

Definition 11.11 (Incomplete elliptic integral of the second kind)

$$E(\phi, k) = \int_0^{\phi} \sqrt{1 - k^2 \sin^2 \theta}\, d\theta \tag{11.16}$$

This integral is related to the calculation of the arc length of an ellipse along a different path.

Definition 11.12 (Incomplete elliptic integral of the third kind)

$$\Pi(n; \phi, k) = \int_0^{\phi} \frac{d\theta}{(1 - n \sin^2 \theta)\sqrt{1 - k^2 \sin^2 \theta}} \tag{11.17}$$

This integral generalises the first kind, incorporating an additional parameter n called the *characteristic.*

Remark 11.6 These *incomplete elliptic integrals* are so-called because ϕ is variable and is not guaranteed to extend over the full period of the elliptic integrand. They are used in problems where the range of the variable of integration is not fixed, such as in various physical and geometric applications. We hence introduce the concept of *complete elliptic integrals*, which depend only on k.

Definition 11.13 (Complete elliptic integral of the first kind)

$$K(k) = F\left(\frac{\pi}{2}, k\right) = \int_0^{\frac{\pi}{2}} \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}} \tag{11.18}$$

Definition 11.14 (Complete elliptic integral of the second kind)

$$E(k) = E\left(\frac{\pi}{2}, k\right) = \int_0^{\frac{\pi}{2}} \sqrt{1 - k^2 \sin^2 \theta} d\theta \tag{11.19}$$

Definition 11.15 (Complete elliptic integral of the third kind)

$$\Pi(n, k) = \Pi\left(n; \frac{\pi}{2}, k\right) = \int_0^{\frac{\pi}{2}} \frac{d\theta}{(1 - n \sin^2 \theta)\sqrt{1 - k^2 \sin^2 \theta}} \tag{11.20}$$

Remark 11.7 These integrals are so-called because the integration spans the full range needed to describe the elliptic curve (i.e. one full cycle of it). They often appear in the solutions of integrals over complete cycles or symmetric limits, such as in the calculation of the period of pendulums and other periodic systems.

11.5 Jacobi elliptic functions

Elliptic functions can be seen as the inverse functions of elliptic integrals. This can be seen in the *Jacobi elliptic functions.*
We rewrite the incomplete elliptic integral of the first kind as u:

$$u = \int_0^{\phi} \frac{d\theta}{\sqrt{1 - k^2 \sin^2 \theta}} \tag{11.21}$$

and the amplitude ϕ as the *Jacobi amplitude* am: ϕ as the *Jacobi amplitude* am:

$$\mathrm{am}(u, m) = \varphi \tag{11.22}$$

Definition 11.16 (Jacobi elliptic functions) In this framework, the Jacobi elliptic functions are given by

- **Elliptic sine**:

$$\mathrm{sn}(u, m) = \sin \mathrm{am}(u, m) \tag{11.23}$$

- **Elliptic cosine**:
$$\operatorname{cn}(u, m) = \cos \operatorname{am}(u, m) \tag{11.24}$$
- **Delta amplitude**:
$$\operatorname{dn}(u, m) = \frac{d}{du} \operatorname{am}(u, m) \tag{11.25}$$
where, if $u \in \mathbb{R}$ and $0 \leq m \leq 1$ (as is typically the case), we can also write $dn(u, m)$ as $\sqrt{1 - m \sin^2 \operatorname{am}(u, m)}$.

Theorem 11.2 (Jacobi elliptic function properties) The Jacobi elliptic functions have various important relationships and properties:

- The functions $\operatorname{sn}(u|m)$, $\operatorname{cn}(u|m)$, and $\operatorname{dn}(u|m)$ are periodic in u.
- Identity I:
$$\operatorname{sn}^2(u, m) + \operatorname{cn}^2(u, m) = 1 \tag{11.26}$$
- Identity II:
$$\operatorname{dn}^2(u, m) + m \cdot \operatorname{sn}^2(u, m) = 1 \tag{11.27}$$

We know from the last section that the elliptic integral $F(\phi, k)$ accumulates the 'arc length from 0 to ϕ on the ellipse, and ϕ depends on the amplitude. The Jacobi elliptic functions $\operatorname{sn}(u, k)$, $\operatorname{cn}(u, k)$, and $\operatorname{dn}(u, k)$ can be seen as giving the amplitude ϕ (or its sine, cosine, and delta values) for a given 'arc length' u. Therefore, elliptic functions are essentially the inverse functions of elliptic integrals. i.e. if we have:

$$u = \int_0^{\phi} R(\theta, k) d\theta \tag{11.28}$$

then the inverse function $\phi = \phi(u)$ can be expressed using elliptic functions.
Remark 11.8 In practice, this allows us to solve problems with elliptic functions instead of elliptic integrals. This is beneficial as elliptic functions are well-studied and come with established tables and computer algorithms.

Chapter 12

ADM formalism

Quote 12.1 Remember: you will eventually come back to me - everyone does, in the end - and your fate will not be enviable.

Philip K. Dick

Perhaps the most exciting topic in this book is the so-called ADM formalism. In our pursuit of a theory of quantum gravity, it is imperative that we construct GR analogues of the Schrödinger equation. However, difficulties exist in that the differentiation of the 4-position by proper time is poorly defined. By eliminating this problem via the decomposition of spacetime into spatial and temporal dimensions, we can eventually arrive at the Wheeler-DeWitt equation, the equivalent of the Schrödinger in GR.

12.1 3+1 decomposition of spacetime

From this point on, our goal is to construct a Hamiltonian formulation of GR. We recall *Hamilton's equations* in classical mechanics:

Theorem 12.1 (Hamilton's equations)

$$\frac{\partial H}{\partial X_i} = -\dot{p}_i \quad \frac{\partial H}{\partial p_i} = \dot{X}_i \tag{12.1}$$

However, one problem soon becomes evident. The generalised velocity $\dot{X}^i$ is not well-defined in GR. We solve this by 'granting previlige' to one of the four coordinates (e.g. time) so that we can define the generalised velocity using it. To this end, we 'split' the spacetime coordinates X^μ into a time coordinate t and three spatial coordinates x^α which we look at as a single object

$$X^\mu = (t, x^\alpha) \quad \text{where} \quad \alpha = 1, 2, 3 \tag{12.2}$$

The end result we have in mind are a series of *constant time hypersurfaces* which we will further investigate in Part III. In an n-dimensional manifold, it is possible to embed an $n-1$-dimesional *hypersurface* Σ_t. With respect to itself, it can be thought of as a 'slice' of a higher dimension. This is because we have fixed one of the n coordinates to get the $n-1$-dimensional hypersurface.
The implication is that we can now think of time evolution not as the change in the time coordinate t but rather the evolution through an infinite series of hypersurfaces. These spacelike constant time hypersurfaces (effectively *slices* or *foliations*) are defined by the tetrad frame

$$\gamma_m = (\xi_0, \xi_a) \quad \text{where} \quad a = 1, 2, 3 \tag{12.3}$$

where ξ_0 is a unique future-pointing unit normal (defined to have unit length) which is orthogonal to the spatial tangent axes ξ_a[1] (also called a *triad*), which are the coordinates on the hypersurfaces:

$$\xi_0 \cdot \xi_0 = -1 \quad \xi_0 \cdot \xi_a = 0 \tag{12.4}$$

[1] Here, ξ has nothing to do with Killing vectors.

The tetrad metric γ_{mn} becomes

$$\gamma_{mn} = \begin{pmatrix} -1 & 0 \\ 0 & \gamma_{ab} \end{pmatrix} \tag{12.5}$$

where γ_{ab} is the so-called *spatial matric*. Now consider the spatial part of some generic coordinate metric of interest $g_{\alpha\beta}$. One can convert it to the spartial (tetrad) metric via

$$g_{\alpha\beta} = \gamma_{ab} e^a_\alpha e^b_\beta \tag{12.6}$$

The *spatial tetrad field* and the *inverse spatial tetrad* field are

Definition 12.1 (Spatial tetrad field and inverse spatial tetrad field)

$$e^a_\alpha = \begin{pmatrix} \alpha & 0 \\ -e^a_\alpha \beta^\alpha & e^a_\alpha \end{pmatrix} \quad e^\alpha_a = \begin{pmatrix} 1/\alpha & \beta^\alpha/\alpha \\ 0 & e^\alpha_a \end{pmatrix} \tag{12.7}$$

where we introduce the two following terms:

Definition 12.2 (Lapse function) The *lapse function* α is the rate at which the proper time τ of the tetrad rest frame advances per unit coordinate time t.

$$\alpha = \frac{d\tau}{dt} \tag{12.8}$$

Physically, this is the rate of advance of proper time between the constant time hypersurfaces.

Definition 12.3 (Shift vector) The *shift vector* β^α is the velocity at which the tetrad rest frame moves through the spatial coordinate X^α per unit coordinate time t.

$$\beta^\alpha = \frac{dX^\alpha}{dt} \tag{12.9}$$

Physically, this is the rate at which coordinates shift between the constant time hypersurfaces.

Remark 12.1 The *spatial tetrad field* e^a_α and its inverse, the *inverse tetrad field* e^α_a observe

$$e^0_\alpha = e^t_a = 0 \tag{12.10}$$

The original 4D metric and inverse metric are then

$$g_{\mu\nu} = \begin{pmatrix} -\alpha^2 + \beta_\mu\beta^\mu & \beta_\nu \\ \beta_\mu & \gamma_{\mu\nu} \end{pmatrix} \quad g^{\mu\nu} = \begin{pmatrix} -\frac{1}{\alpha^2} & \frac{\beta^\nu}{\alpha^2} \\ \frac{\beta^\mu}{\alpha^2} & \gamma_{\mu\nu} - \frac{\beta^\mu\beta^\nu}{\alpha^2} \end{pmatrix} \tag{12.11}$$

and the determinants observe

$$\sqrt{-g} = \alpha\sqrt{y} \tag{12.12}$$

Rewriting in terms of the line element gives

Definition 12.4 (Original 4D line element)

$$ds^2 = (-\alpha^2 + \beta_\mu\beta^\mu)dt^2 + 2\beta_\mu dt dx^\mu + \gamma_{\mu\nu}dx^\mu dx^\nu \tag{12.13}$$

where $\gamma_{\mu\nu}$ is the *spatial metric* on the spatial hypersurface Σ_t. It provides all the information about the intrinsic geometry of the hypersurfaces.

Now we solve for $\gamma_{\mu\nu}$. To do so we first consider the *normal vector*. i.e. the vector normal to the hypersurface:

Definition 12.5 (Normal vector)

$$n^\mu = (1/\alpha, -\beta^i/\alpha) \tag{12.14}$$

From this, we can express the (3D) spatial metric.

Definition 12.6 (Spatial metric)

$$\gamma_{\mu\nu} = g_{\mu\nu} + n_\mu n_\nu \tag{12.15}$$

In summation convention:

$$n^\mu = -\alpha \nabla^\mu t \tag{12.16}$$

where the minus sign is there to guarantee that n^μ is future pointing.

Remark 12.2 As this spatial metric is 'induced' from the 4D spacetime metric, we often call it the *induced metric*. However the same term is often used to refer to 2D metrics of embedded 2-surfaces.

Note 12.1 We will often label the 4D quantities with a superscript $^{(4)}$ before them to distinguish them from their 3D counterparts, which will have $^{(3)}$ before them. The exceptions are the shift vector, whose 3D equivalent is the *shift function* N^i and the 3D lapse function, which is labelled N.

12.2 Extrinsic curvature

The *extrinsic curvature* is also called the *second fundamental form*. Since there is a second fundamental form, we must presume that there is a *first fundamental form*. As it turns out that is indeed the case, and the first fundamental form is simply the induced metric we know and love. Consider a 3D hypersurface Σ embedded in a 4D spacetime and recall the induced metric:

$$\gamma_{ab} = g_{\mu\nu} e^\mu_a e^\nu_b \quad ds^2 = g_{\mu\nu} dx^\mu dx^\nu = \gamma_{ab} dy^a dy^b \tag{12.17}$$

We then consider the decomposed 3+1D spacetime, which has the parametric equation

$$X^\alpha = X^\alpha(Y_a) \tag{12.18}$$

where Y^a is the 2+1D coordinates on the (constant time) hypersurface.
Noting that tetrad fields can be used to induce a 3D hypersurface metric from a 4D metric, we can write

$$ds^2 = g_{\mu\nu} dX^\mu dX^\nu = h_{ab} dY^a dY^b \tag{12.19}$$

where $h_{ab} = g_{\mu\nu} e^\mu_a e^\nu_b$ is the induced hypersurface metric, and there exist tetrad fields $e^\mu_a = \frac{\partial X^\mu}{\partial Y^a}$.
More generally, this induced metric is known as the first fundamental form in differential geometry.

Definition 12.7 (First fundamental form) The first fundamental form is the inner product on the tangent space of a surface. Suppose we have two tangent vectors $aX_u + bX_v$ and $cX_u + dX_v$). We perform an inner product between them:

$$\mathrm{I} = ac\langle X_u, X_u\rangle + (ad+bc)\langle X_u, X_v\rangle + bd\langle X_v, X_v\rangle = Eac + F(ad+bc) + Gbd \tag{12.20}$$

One can write this in matrix notation and the form of the metric:

$$\mathrm{I} = x^T \begin{pmatrix} E & F \\ F & G \end{pmatrix} y \quad g_{\mu\nu} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} = \begin{pmatrix} E & F \\ F & G \end{pmatrix} \tag{12.21}$$

Remark 12.3 Interestingly, the first fundamental form can be written as $\mathrm{I} = g_{ij} dX^i dX^j$. This is equivalent to the Langragian in particle geodesics.
But how do we solve for quantities on hypersurfaces when all we have are quantities on the manifold? This is where the so-called *projection operator* comes into play.

Definition 12.8 (Projection operator) The projection operator projects objects on the manifold onto the hypersurface:

$$P^\alpha_\beta = \delta^\alpha_\beta + n^\alpha n_\beta \tag{12.22}$$

At this point, you might see a pattern emerging in our train of thought. We want to operate on 3D constant time hypersurfaces within the 4D spacetime instead of the 4D spacetime *itself*. In our absolute loyalty to this doctrine we also define the *extended covariant derivative*. It is the functional equivalent to the covariant derivative. However, unlike the covariant derivative, it restricts itself purely to the 3D constant time hypersurface.

Definition 12.9 (Extended covariant derivative) The *extended covariant derivative* is

$$D_\mu = P^\alpha_\mu \nabla_\alpha \tag{12.23}$$

Remark 12.4 At first the name may seem confusing. Nonetheless one can make sense of it by understanding it as the 'extension' of the 4D covariant derivative to the 3D hypersurface.
Now we proceed[2] to the second fundamental form II. Consider the 3D constant time hypersurface that is defined by $f(X^\mu(Y^a))$ and its tangent plane at the origin $f(X^\mu(Y^a)) = 0$. This then implies the vanishing of f and its x- and y-partial derivatives at $(0,0)$.
As long as the hypersurface is not timelike/null, one can define a *unit normal vector*:

Theorem 12.2 (Unit normal on a hypersurface)

$$n_\mu = \epsilon \left| g^{\alpha\beta} \frac{\partial f}{\partial X^\mu} \frac{\partial f}{\partial X^\mu} \right|^{-\frac{1}{2}} \frac{\partial f}{\partial X^\mu} \tag{12.24}$$

where n_μ observes $n_\mu e^\mu_a$ and for convenience, we define an arbitrary scalar $\epsilon = n^\mu n_\nu$. ϵ satisfies

$$n^\mu n_\nu = \epsilon = \begin{cases} -1 & \text{if } \Sigma \text{ is spacelike} \\ 1 & \text{if } \Sigma \text{ is timelike} \end{cases} \tag{12.25}$$

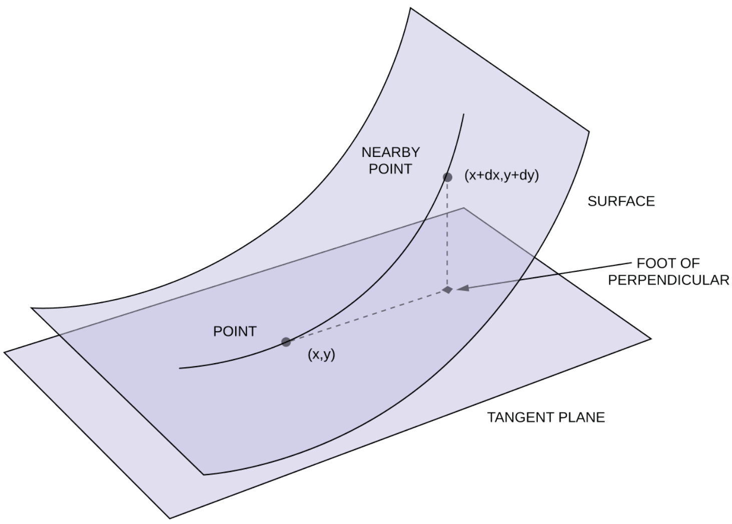


Figure 12.1: A surface and its tangent plane.

As the hypersurfaces Σ_t are embedded in the manifold in the 3+1 formalism, we introduce the concept of *extrinsic curvature* or the *second normal form*[3] in contrast to the *intrinsic curvature* of the manifold we usually use.
While the intrinsic curvature is the normal curvature that describes the geometry within a manifold itself as encoded by the Riemann, Ricci and Weyl tensors, the extrinsic curvature describes the curvature of a submanifold *within* a higher-dimensional embedding space. Mathematically, the extrinsic curvature corresponds to the variation of the normal vector of the hypersurface n along a tangent vector.
Remark 12.5 Note that the extrinsic curvature is *not* the curvature of the hypersurface *itself*. The curvature of the hypersurface is the *intrinsic curvature* of the hypersurface.

Definition 12.10 (Extrinsic curvature) The extrinsic curvature $K_{\mu\nu}$ describes how the normal vector changes as one moves along the submanifold. Given a submanifold with a normal vector n^μ, it is

[2] Or rather go back?
[3] There is also the *third fundamental form*, although this is outside the scope of the book.

defined as:

$$K_{ij} = -n_\mu \nabla_j e^\mu_i = -e^a_i e^b_j n_\mu \left(\partial_b e^\mu_a + \omega_{b\,a}{}^c e^\mu_c\right) \tag{12.26}$$

where ∇_j is the covariant derivative in the higher-dimensional (embedding) space and ω^c_{ba} are the spin connection coefficients associated with the tetrad field.

Remark 12.6 For example, if you take a 2D surface (e.g. a sphere or a plane) embedded in 3D space, the extrinsic curvature tells you how this surface bends in the third dimension.
Expectedly, there is a relation between the extrinsic and intrinsic curvatures. This is shown in the *Gauss-Codazzi equations*:

Theorem 12.3 (Gauss-Codazzi equations)

- **Gauss equation**:
$$P^\rho_\alpha P^\beta_\sigma P^\gamma_\mu P^\delta_\nu \, {}^{(4)}R^\alpha_{\beta\gamma\delta} = \mathcal{R}^\rho_{\sigma\mu\nu} + K^\rho_\mu K_{\sigma\nu} - K^\rho_\nu K_{\sigma\mu} \tag{12.27}$$
- **Codazzi equation**:
$$P^\alpha_\rho P^\mu_\beta P^\nu_\gamma \, {}^{(4)}\mathcal{R}^\rho_{\sigma\mu\nu} n^\sigma = D_\gamma K^\alpha_\beta - D_\beta K^\alpha_\gamma \tag{12.28}$$

We can then derive the Ricci scalar from the extrinsic curvature:

Theorem 12.4 (Ricci scalar)

$${}^{(4)}R = R + K^2 + K^{ij}K_{ij} - 2\nabla_n K = \frac{2}{N} D^i D_i N \tag{12.29}$$

Now we consider how the stress-energy tensor can be projected into Σ_t. Conceptually speaking, the stress-energy tensor encompasses both stress and energy. This manifests in the following relation for the trace of $T_{\alpha\beta}$:

$$T = S - E \tag{12.30}$$

Here $E = T_{\mu\nu} n^\mu n^\nu$ is simply the energy density. S is the traced stress which implies a stress tensor $S_{\alpha\beta}$. This is

Definition 12.11 (Stress tensor)

$$S_{\alpha\beta} = T_{\mu\nu} P^\nu_\alpha P^\nu_\beta \tag{12.31}$$

With these projected quantities we have the total projection of the Einstein field equations.

12.3 3+1 field equations

Theorem 12.5 (3+1 field equations)

- **Total projection onto Σ_t**:
$$\mathcal{L}_m K_{ij} = -D_i D_j N + N[R_{ij} - 2K_{il}K^l_j + KK_{ij} + 4\pi(\gamma_{ij}(S_E) - 2S_{ij})] \tag{12.32}$$
Note that $\mathcal{L}_m$ is the Lie derivative we have encountered earlier. The Lagrangian is L as usual.
- **Total projection along n^μ**:
$$R - K_{ij}K^{ij} + K^2 = 16\pi E \tag{12.33}$$
This is known as the *Hamiltonian constraint*.
- **Mixed projection onto Σ_t and along n^μ**:
$$D_j K^i_j - D_i K = 8\pi p_i \tag{12.34}$$
where $p_i = -T_{\mu\nu} n^\mu \gamma^\nu_i$ is the momentum density. This is known as the *momentum constraint*.

12.4 ADM, action and canonical momentum

GR operates under a Lagrangian framework. In the grand quest of quantum gravity, we need to quantise things, which is easier done in a Hamiltonian framework. The ADM formalism, as devised by Richard Arnowitt, Stanley Deser, and Charles W. Misner, represents GR in a Hamiltonian framework.
The ADM formalism is formulated by applying the 3+1 framework to action and Hamiltonian mechanics in general. In classical mechanics, the so-called *canonical*[4] *momentum density* is defined as

$$\pi = \frac{\partial L}{\partial \dot{q}} \tag{12.35}$$

In GR, the tensorial equivalent is defined as

Definition 12.12 (ADM canonical momentum)

$$\pi^{ij} = \frac{\partial L}{\partial \dot{\gamma}_{ij}} = -\sqrt{\gamma}(K^{ij} - \gamma^{ij} K) \tag{12.36}$$

where the dot simply refers to the time derivative.

As we have already derived the tensorial 3D metric and canonical momentum, it is easy to find the ADM Hamilton's equations:

Theorem 12.6 (ADM Hamilton's equations)

$$\frac{\partial H}{\partial \pi^{ij}} = \dot{\gamma}_{ij} \quad \frac{\partial H}{\partial \gamma^{ij}} = -\dot{\pi}_{ij} \tag{12.37}$$

Theorem 12.7 (Evolution equations)

- **Spatial metric**:

$$\dot{\gamma}_{ij} = D_i\beta_j + D_j\beta_i - 2\alpha K_{ij} \tag{12.38}$$

- **Extrinsic curvature**:

$$\begin{aligned}\dot{K}_{ij} =& - D_i D_j \alpha + \alpha(R_{ij} - 2K_{il}K^l_j + KK_{ij}) + \beta^k \partial_k K_{ij} \\ &+ K_{ik}\partial_j \beta^k + K_{jk}\partial_i\beta^k - 8\pi G\alpha \left(S_{ij} - \frac{1}{2}\gamma_{ij}(S-\rho)\right)\end{aligned} \tag{12.39}$$

- **Canonical momentum**:

$$\begin{aligned}\dot{\pi}^{ij} =& - N\sqrt{\gamma}\left(R^{ij} - \frac{1}{2}\gamma^{ij}R\right) + \frac{N}{2\sqrt{\gamma}}\left(\pi_{cd}\pi^{cd} - \frac{\pi^2}{2}\right)\gamma^{ij} \\ &- \frac{2N}{\sqrt{\gamma}}\left(\pi^{ic}\pi_c{}^j - \frac{1}{2}\pi\pi^{ij}\right) + \sqrt{\gamma}\left(D^iD^jN - \gamma^{ij}D_cD^cN\right) \\ &+ D_c\left(\pi^{ij}N^c\right) - \pi^{ic}D_cN^j - \pi^{jc}D_cN^i\end{aligned} \tag{12.40}$$

12.5 Wheeler-DeWitt equation

The final step in developing a quantum formulation would be converting previously known quantities into a tensorial form. Wavefunctions $|\Psi\rangle$ become *wave functionals* $\Psi[\gamma_{ab}]$:

$$\hat{\gamma}_{ij}(t, x^k)|\Psi\rangle \to \gamma_{ij}(t, x^k)\Psi[\gamma_{ab}] \tag{12.41}$$

$$\hat{\pi}^{ij}(t, x^k)|\Psi\rangle \to -i\frac{\delta}{\delta\gamma_{ij}(t, x^k)}\Psi[\gamma_{ab}] \tag{12.42}$$

[4] Here 'canonical' simply means conforming to Hamiltonian mechanics. This is significant as due to the inclusion of various potentials we do not necessarily have $p = mv$.

The first and third evolution equations then become

$$\hat{R}^0|\Psi\rangle \doteq -\left[\sqrt{\gamma}R + \frac{1}{\sqrt{\gamma}}\left(\frac{\pi^2}{2} - \pi^{ij}\pi_{ij}\right)\right]\Psi\left[\gamma_{kl}\right] = 0 \tag{12.43}$$

$$\hat{R}^i|\Psi\rangle \doteq -2D_j\pi^{ij}\Psi\left[\gamma_{kl}\right] = 0 \tag{12.44}$$

where first one can be rewritten as

Theorem 12.8 (Wheeler-Dewitt equation)

$$\left[\sqrt{\gamma}R - \frac{\hbar^2}{\sqrt{\gamma}}\left(\frac{1}{2}\gamma_{ab}\gamma_{cd} - \gamma_{ac}\gamma_{bd}\right)\frac{\delta}{\delta\gamma_{ab}}\frac{\delta}{\delta\gamma_{cd}}\right]\Psi\left[\gamma_{kl}\right] = 0 \tag{12.45}$$

The Wheeler-DeWitt equation is the analogue to the Schrödinger equation in quantised GR. Now, the significance of the ADM formalism in quantum gravity can be seen: we have combined GR and quantum mechanics in an emergence of quantum gravity.

Part III

Cosmology

Chapter 13

Cosmology before GR

Quote 13.1 The evolution of the world can be compared to a display of fireworks that has just ended: some few red wisps, ashes and smoke. Standing on a well-chilled cinder, we see the slow fading of the suns, and we try to recall the vanished brilliance of the origin of the worlds.

Georges Lemaître, 1931

As a field of study, cosmology predates GR, and will probably also outlive GR should a new theory of gravitation emerge. In this chapter, we will cover the portion of cosmology developed before GR as well as some key concepts which will be used throughout this part of the book.

13.1 Comoving frames

We are interested in a special frame of reference called the *cosmological reference frame* or the *cosmic frame* in which cosmological quantities are homogeneous and isotropic. Its coordinates are called the *comoving coordinates*:

Definition 13.1 (Comoving coordinates) The *comoving coordinates cosmological reference frame* is a set of coordinates in which physical quantities are homogeneous and isotropic.

Remark 13.1 They are called 'comoving' coordinates because they move along with the Hubble flow.

Note 13.1 We thus have a series of values in comoving coordinates like the *comoving observer* R (represented by their position) and the comoving distance d. They are all derived by scaling (dividing) their non-comoving counterpart by various powers of the scale factor $a(t)$. Some of them will be discussed below.

Definition 13.2 (Scale factor) The time-dependent *scale factor* $a(t)$ (in effect an expansion parameter) essentially rescales any parameter into their comoving counterpart. We see the following example with the comoving distance:

$$\vec{r}(t) = \frac{a(t)}{a_0}\vec{r}_0 \quad \text{where} \quad a_0 = a(t_0) \tag{13.1}$$

Remark 13.2 The scale factor is dimensionless and characterises the expansion of the universe.

Definition 13.3 (Comoving observer) The *comoving observer* is an observer at rest in comoving coordinates.

Definition 13.4 (Cosmic time) *Cosmic time* is the proper time t measured by a comoving observer, starting with $t = 0$ at the Big Bang.

The cosmic microwave background provides the best method of determining the comoving coordinates: In comoving coordinates, the CMB is perfectly isotropic save for *smol* fluctuations associated with primeval galaxies.

Remark 13.3 We observe a large dipole in the CMB that results from the proper motion of the Milky Way with respect to comoving coordinates.

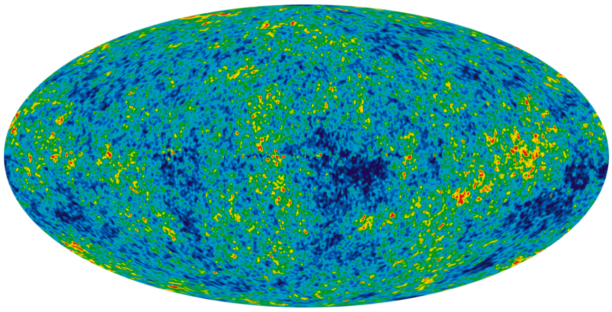


Figure 13.1: 'These are the axes of evil... you must conquer each!'

13.2 Historical overview

Now with all the basic concepts by our side, we can proceed with a recap of the history of cosmology.

> **Quote 13.2** The key question of classical cosmology is: Which solutions of the Einstein field equations describe the (idealised) universe that we observe?
>
> *Christian G. Böhmer, 2009*

Definition 13.5 (Homogeneity) A system is *homogeneous* if it is invariant under translations

$$X_a \to X_a + X_c \tag{13.2}$$

Definition 13.6 (Isotropy) A system is *isotropic* if it is invariant under rotations. i.e. it looks the same in all directions.

Theorem 13.1 (Cosmological principle) There exist coordinate systems such that the universe appears spatially isotropic and homogeneous on sufficiently large scales for every observer.

Remark 13.4 Note that we are explicitly stating here that *space* is isotropic and homogeneous and not *space-time*. We assume time to be absolute and deal with the spatial components.

Theorem 13.2 (Axioms) In cosmology, we assume the following *axioms*:

1. The laws of physics as we know them today are valid in the past and will be valid in the future. This includes the non-variability of physical constants like c.
2. The universe is connected.
3. When describing the universe, we are not interested in substructures like galaxies, planetary systems, stars etc.[a], but in the universe on its largest scales. Then we can conclude
   - Electromagnetic forces cancel **on average** as we expect to have the same number of positive and negative charges in the universe.
   - The strong and weak nuclear forces are acting only in the nucleus of the atom ($d \approx 10^{-15}$m), i.e. are not relevant for us.
   - The only fundamental interaction relevant is **gravity**.

[a] This is astrophysics!

> **Quote 13.3** The universe is a very specific system because you only have one. (...) There is only one universe. That is a very restricting and important

observation.

Betti Hartmann, 9 January 2024

Currently, we use the aforementioned FLRW model. Nonetheless, we note that an assumption may very well be wrong. We take, for example, Einstein's initial assumption that the universe is static, infinite and homogeneous.

Definition 13.7 (Staticity) Static solutions are models in which the universe remains constant in size.

Derivation 13.1 (Olbers's paradox, or why Einstein was wrong) Olbers's paradox was formulated in 1826. Assume that the number density of stars is given by

$$n(r) = n_* = \text{constant} \tag{13.3}$$

Then the number of stars in a spherical shell located at distance r with thickness dr is

$$N_* = 4\pi n_* r^2 dr \tag{13.4}$$

Assume a star located at r has luminosity L. Then, the apparent luminosity (flux density) is

$$F' = \frac{L}{(4\pi r^2)} \tag{13.5}$$

We can then calculate the total luminosity (flux density) that reaches us from all stars in an infinite, static universe:

$$\int_0^\infty F'(r) 4\pi r^2 dr = \int_0^\infty \frac{L}{4\pi r^2} 4\pi n_* r^2 dr = n_* \int_0^\infty dr = \infty \tag{13.6}$$

This means that the night sky would be infinitely bright - which it is not.

Remark 13.5 To resolve this paradox, there are two possibilities:

- The universe is not static.
- The universe is finite.

We thus arrive at what Hubble did in 1929. Light from distant galaxies is 'redshifted', and this redshift z is proportional to the distance d of the galaxies:

$$d \propto z := \frac{\Delta\lambda}{\lambda} \tag{13.7}$$

We thus have *Hubble's law*:

Theorem 13.3 (Hubble's law)

$$\vec{v} = H(t)\vec{r} \tag{13.8}$$

where significantly, H is independent of $\vec{r}$ and is only a function of t.

Note 13.2 The *Hubble parameter* H is not really a constant when considering the evolution of the universe. In fact, $H = H(t)$ and the present value is often denoted by $H(t_0) = H_0$. Likewise, we usually denote the present value of some parameter by a subscript $_0$. e.g. *present time* is t_0.

Quote 13.4 He did them for galaxies so close that all the assumptions are fine.

Betti Hartmann, on Hubble's law, 9 January 2024

From the data on universe expansion, there are two scenarios:

- The galaxies are moving away from us (e.g. the Andromeda galaxy is moving towards us).
- Space is expanding.

Remark 13.6 A new development is the *Hubble tension*.

13.3 Non-relativistic cosmology

As we stand now, the split of space and time suggests that we could use non-relativistic equations. This non-relativistic approach to cosmology (or rather *Newtonian cosmology*) has since been eclipsed by GR-based (or *Friedmann*) cosmology, but we will explore it briefly for its historical significance and relative[1] simplicity. Since we are only interested in the universe 'on average', we can assume it to be a perfect fluid.

Remark 13.7 This fluid is made of 'fluid elements', which, for the non-relativistic model, we assume to be the galaxies, i.e. the matter we can observe in the universe.

Theorem 13.4 (Non-relativistic fluid equations) The equations of non-relativistic fluid dynamics are:

- **Continuity equation**: This describes the conservation of mass in the fluid.

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0 \tag{13.9}$$

  where ρ is the fluid density (i.e. matter density), $\vec{v}$ is the fluid velocity, and ∇ denotes the gradient.

- **Euler's equation**: This is a statement of the conservation of momentum.

$$\frac{D\vec{v}}{Dt} = \frac{\partial \vec{v}}{\partial t} + \vec{v} \cdot \underbrace{\nabla \vec{v}}_{\text{covariant derivative}} = -\frac{\nabla p}{\rho} + \vec{f} \tag{13.10}$$

  where D is the *material derivative* we saw in Equation 2.46, p is the pressure, $\vec{f}$ is the force per unit mass, Φ is the gravitational potential and $\vec{r}$ is the position vector[a].

- **Gravitational potential equation**: Also known as *Poisson's equation*, this equation emerges when the only force is gravity. i.e. when $\vec{f} = \vec{g}$:

$$\nabla^2 \Phi = 4\pi G \rho \tag{13.11}$$

  where G is the gravitational constant[b]. This equation states that the curvature of the gravitational potential is directly proportional to the local mass density.

[a] Note that $\nabla \vec{v}$ is the all-too-familiar covariant derivative - you can only have so many symbols.
[b] Omitted from the next chapter on.

Remark 13.8 Note that these equations are not intrinsically cosmology - they are merely fluid dynamics equations which happen to find their use in Newtonian cosmology, as we will see later on.

Note 13.3 Remember that $\vec{r} = t\vec{v}$ and all that follows.

Remark 13.9 From the fluid equations, we can infer that homogeneity implies that $\vec{f}$ and $\vec{v}$ are constant vectors. But this would give a preferred direction, meaning that a system cannot be both homogeneous and isotropic! The comoving frame solves this problem.

Definition 13.8 (Hubble parameter) We can represent the *Hubble parameter* H by the scale factor:

$$H(t) = \frac{\dot{a}(t)}{a(t)} \tag{13.12}$$

Noting Equation 13.8 and that:

- The divergence of ϕ observes, by the definition of $\vec{g}$:

$$\nabla \Phi = \vec{g} \tag{13.13}$$

- The universe, being isotropic and homogeneous, has no pressure gradients[2]:

$$\nabla p = 0 \tag{13.14}$$

[1] See what I did?
[2] Another way to consider this is that the pressure p is negligible compared to the energy density ρ.

We can rewrite the fluid equations for Newtonian cosmology:

Theorem 13.5 (Newtonian cosmology fluid equations) Continuity equation:

$$\frac{\partial \rho}{\partial t} + 3\rho H(t) = 0 \tag{13.15}$$

Euler's equation:

$$\frac{\partial H}{\partial t} + H(t)^2 = \vec{f}(t) \tag{13.16}$$

Gravity equation:

$$g(t) = -\frac{4\pi G\rho}{3} \tag{13.17}$$

Finally, for completeness, we simplify the Friendmann equations, which we will later derive from the Einstein field equations. If this is your first reading, you should skip to the next chapter and only return after having seen the Friedmann equations.
We consider $p = k = 0$ for Equation 15.7, Equation 15.8 and Equation 15.9. They reduce to

Theorem 13.6 (Friedmann equations for non-relativistic matter)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{\Lambda}{3} \tag{13.18}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\rho \tag{13.19}$$

$$\frac{d}{dt}(\rho a^3) = 0 \tag{13.20}$$

Einstein introduced the cosmological constant to maintain the static universe[3]:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho - \frac{\Lambda c^2}{4\pi G}\right) \tag{13.21}$$

Remark 13.10 Up to this point, we have not acknowledged the cosmological constant, which induces a negative pressure. Due to the negative sign in the term, the effect of the cosmological constant is sometimes called *antigravity*.

[3] We will see this later as *Einstein's static solution* or the *static universe*.

Chapter 14

FLRW model

> **Quote 14.1** Wilhelm Killing, der die Behandlung der Grundlagen der Geometrie zu seiner Lebensaufgabe machte, veröffentlichte eine Reihe von Lehrbüchern über Geometrie und Elementarmathematik.
>
> *Universitäts- und Landesbibliothek Münster website*

In this chapter, we will derive the currently best-accepted Friedmann-Lemaître-Robertson-Walker (FLRW) metric, which is an exact solution to the field equations. We will go through some assumptions of the FLRW model and derive them using constant time hypersurfaces.

14.1 Local isometry

In the introductory chapter, we have investigated the universe as a whole. But what about smaller scales?

Definition 14.1 (Isometry) A coordinate transformation $X^a \to X'^b$ is called a *local isometry* if the metric fulfills

$$g'_{\mu\nu} = g_{\mu\nu} \tag{14.1}$$

Derivation 14.1 (Example isometries) The vector V generates an isometry if

$$\mathcal{L}_V g_{\mu\nu} = 0 \tag{14.2}$$

If the Schwarzschild metric is static, the following generates an isometry

$$V_t^\mu = (1, 0, 0, 0) = \delta_0^\mu \tag{14.3}$$

Similarly, the following generates an isometry (i.e. rotational invariance)

$$V_\phi^\mu = \delta_\phi^\mu \tag{14.4}$$

14.2 Killing equation

We invoke the covariant derivative and *insert*[1] its definition into the Lie derivative. The Lie derivative of the metric then reads

Theorem 14.1 (Lie derivative of the metric)

$$\mathcal{L}_\xi g_{\mu\nu} = \nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu \tag{14.5}$$

For local isometries this then leads to the following:

[1] Oh, no, you don't!

Theorem 14.2 (Killing equation)

$$\mathcal{L}_\xi g_{\mu\nu} = 0 \quad \text{or} \quad \nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu = 0 \tag{14.6}$$

It has 10 independent components (4 translations, 3 rotations, 3 boosts) in 4 space-time dimensions and allows us to find the vectors ξ^μ that generate the isometries. It is so useful that one can almost say that it is *to die for*.

Definition 14.2 (Killing vector) Such ξ^μ vectors are called the *Killing vectors*, which are of *capital* importance in the exact solutions of the Einstein field equations.

Remark 14.1 Linear combinations of Killing vectors are also Killing vectors.

Quote 14.2 Killing is not killing the metric. It is named after Wilhelm Killing.

Betti Hartmann, 19 January 2024

Figure 14.1: Wilhelm Killing, c. 1889-1891.

Note 14.1 Remember that the Killing vector is tensorial. Switching its indices would require applying the metric.

Quote 14.3 Since the second covariant derivatives of vectors involve the Riemann curvature tensor, the Killing vectors have to satisfy certain integrability conditions (We will not discuss these issues). Killing vectors are important in general relativity and cosmology because they can be used to define conserved quantities.

Christian G. Böhmer, 2009

Definition 14.3 (Maximally symmetric space) A *maximally symmetric space* observes

$$R_{\mu\nu} = \frac{R}{d} g_{\mu\nu} \tag{14.7}$$

where the Ricci scalar R is constant for $d \geq 2$. i.e.

$$R_{\mu\nu} \propto g_{\mu\nu} \tag{14.8}$$

Theorem 14.3 (Properties to die for) The following statements hold:

- A homogeneous d-dimensional space admits the maximal number d of translational Killing vectors.
- An isotropic d-dimensional space admits the maximal number $d(d-1)/2$ of rotational Killing vectors.

Remark 14.2 An example would be Minkowski space, which, in 4D, has $\frac{4 \cdot 5}{2} = 10$ and is hence maximally symmetric and homogeneous.

Theorem 14.4 (Other equivalences) We also have

- Every maximally symmetric space is homogeneous and isotropic.
- An isotropic space is also homogeneous.
- Every maximally symmetric metric is conformally flat (*but not vice versa*). The reason for this is that the conformal transformations locally preserve angles and orientations.

Exercise 14.1 Give an example of metric such that $R_{\mu\nu} = cg_{\mu\nu}$ for some constant c, but which is not maximally symmetric.

14.3 Constant time hypersurfaces

The spacetimes discussed in modern cosmology are in 4D, with 3 spatial dimensions and 1 temporal dimension, which is cosmic time. If we take a snapshot of the spacetime at some particular t_0, what we get is a 3D manifold embedded in the 4D spacetime or a so-called hypersurface that is called a *constant time hypersurface*[2].
From homogeneity and isotropy, it follows that this hypersurface has constant curvature. We can classify them quite easily:

Theorem 14.5 (Spaces of constant curvature) For spaces of constant curvature:

- If $k = 1$, we have a sphere.
- If $k = 0$, we have an Euclidean space.
- If $k = -1$, we have a hyperbolic space.

where k is the *curvature parameter* which we will discuss almost immediately.

As just mentioned, we want to preserve isotropy (i.e. the rotational components), yet we want a rotational component that is purely dependent on r. Such a metric can be written as

Definition 14.4 (Conformally-adjusted hypersurface metric)

$$ds^2 = \frac{1}{\left(1 + \frac{kr^2}{4}\right)^2} \left(dr^2 + \tilde{r}^2 d\Omega_2^2\right) \tag{14.9}$$

where $1/\left(1 + \frac{kr^2}{4}\right)^2$ is the *conformal factor*.

This can obviously be rescaled[3]. We introduce a new radial coordinate

$$\rho = r\left(1 + \frac{kr^2}{4}\right)^{-1} \tag{14.10}$$

which gives

$$ds^2 = \frac{d\rho^2}{1 - k\rho^2} + \rho^2 d\Omega^2 \tag{14.11}$$

[2] If you have read Part II, you will be able to recall this.
[3] We do this in GR exercises all the time.

But what is this k? This can be investigated by solving for the spatial part of the Ricci tensor, which is

$$R_{\text{spatial}} = \frac{6k}{a^2(t)} \tag{14.12}$$

Already, the Ricci scalar is understood to be the scalar curvature. As such, we conclude the physical significance of k:

Definition 14.5 (Curvature parameter) k is the scale-adjusted version of the scalar curvature which we call the *curvature parameter*.

This will become significant in later chapters where we investigate k under different eras.

Derivation 14.2 (3-sphere) Considering the 3-sphere and for simplicity, setting $k = 1$, we impose another coordinate transformation

$$\rho = \sin\chi \tag{14.13}$$

Definition 14.6 (Metric for a constant time 3-sphere)

$$ds^2 = d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta d\phi^2) \tag{14.14}$$

Exercise 14.2 Use the following coordinate transformations

$$x = \sin\chi\sin\theta\cos\phi \quad y = \sin\chi\sin\theta\sin\phi \quad z = \sin\chi\cos\theta \quad w = \cos\chi \tag{14.15}$$

to find that this metric is indeed that of a 3-sphere.

Derivation 14.3 (Flat space) For flat spaces, we do not need any coordinate transformations. The metric is already that of an Euclidian space in spherical polar coordinates.

$$ds^2 = \frac{d\rho^2}{1 - k\rho^2} + \rho^2 d\Omega^2 \tag{14.16}$$

Derivation 14.4 (Hyperboloid) Here we have the so-called *hyperboloid* or *hyperbolic space* $\mathbb{H}^3$. We can impose a hyperbolic coordinate transformation

$$\rho = \sinh\chi \tag{14.17}$$

We can insert this into the metric for a 4D constant time hypersurface, which is identical to Equation 14.14 except for the addition of the time coordinate. From here, we can write down the hyperbolid metric:

$$ds^2 = -dt^2 + a(t)^2\left[d\chi^2 + S_k(\chi)^2(d\theta^2 + \sin^2\theta d\phi^2)\right] \tag{14.18}$$

where $S_k(\chi)$ represents the spatial curvature term and observes

$$S_k(\chi) = \begin{cases} \sin\chi & \text{for } k = +1 \\ \chi & \text{for } k = 0 \\ \sinh\chi & \text{for } k = -1 \end{cases} \tag{14.19}$$

14.4 FLRW metric

We recall the introductory content and ask ourselves: What are the implications of the cosmological principle? Or in other words, what are the implications of homogeneity and isotropy?

Theorem 14.6 (FLRW metric assumptions) Below are three main assumptions that lead to the FLRW metric.

- The universe is (statistically) homogeneous and isotropic on large scales[a].
- The constant time hypersurfaces are of constant curvature.

This stems from the fact that the curvature parameter k, as seen previously, can only admit values of 1, 0 and -1.

- The dynamical behaviour of the spacetime is described by the scale factor $a(t)$.

[a] This means that you have only one independent function which is $a(t)$.

All this yields the FLRW model. It is the currently accepted model for the universe that seems to agree with all known observations to date. By rewriting Equation 14.18 in terms of k, we have

Definition 14.7 (FLRW metric) In isotropic coordinates, the FLRW metric is

$$ds^2 = -dt^2 + a^2(t)\frac{dx^2 + dy^2 + dz^2}{\left(1 + \frac{k}{4}(x^2 + y^2 + z^2)\right)^2} \tag{14.20}$$

In polar coordinates, it is

$$ds^2 = -dt^2 + a(t)^2\left[\frac{dr^2}{1 - kr^2} + r^2 d\Omega^2\right] \tag{14.21}$$

where, notably, t should be understood as cosmic time.

We note its implications:

- The 3D constant time hypersurface metric (which represents only one point in time) is multiplied by the scale factor $a(t)$, which represents dynamic evolution of the universe *over time*.
- $\frac{dt}{d\tau} = 1$ describes an observer in free fall (moving along a geodesic). Such an observer is a co-moving observer and $(t, r, \theta\phi)$ are co-moving coordinates.
- Given the 1 term in the denominator, the metric becomes spacelike when x, y and z are small.
- $k/4$ is a mathematically convenient term that incorporates the curvature parameter k.

Chapter 15

Friedmann equations and exact solutions

The Friedmann equations are simply the field equations when the metric is fixed to the FLRW metric. From them, we can investigate several significant concepts in cosmology as well as their solutions under different choices of constants.

15.1 Friedmann equations

From the Einstein field equations, we can derive the so-called *Friedmann equation* or the *cosmological field equations* which describe the evolution of the universe.

Derivation 15.1 (Friedmann equations) By substitution, one can find the Einstein tensor components corresponding to the FLRW metric as:

$$G^t_t = -9\frac{\dot{a}^2}{a^2} - 9\frac{k}{a^2} \quad G^x_x = G^y_y = G^z_z = -2\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} - \frac{k}{a^2} \tag{15.1}$$

Inserting these results to the field equations with the cosmological term, as seen in Equation 5.28 and the stress-energy tensor for a perfect fluid in Equation 3.18 gives the first two Friedmann equations:

$$-3\frac{\dot{a}^2}{a^2} - 3\frac{k}{a^2} + \Lambda = -8\pi\rho \tag{15.2}$$

$$2\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} - \frac{k}{a^2} + \Lambda = 8\pi p \tag{15.3}$$

The third equation is slightly tricky. We first differentiate Equation 15.2:

$$2\left(\frac{\dot{a}}{a}\right)\left(\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2}\right) = \frac{8\pi}{3}\dot{\rho} + 2\frac{k}{a^2}\frac{\dot{a}}{a} \tag{15.4}$$

Now we want to remove the term $\frac{\ddot{a}}{a}$. To do so, we can add Equation 15.2 and Equation 15.3, which yields

$$-\frac{\ddot{a}}{a} = \frac{4\pi}{3}(\rho + 3p) - \frac{\Lambda}{3} \tag{15.5}$$

Substitute $\frac{\ddot{a}}{a}$ with this result, and we find the third equation:

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0 \tag{15.6}$$

To sum up:

Theorem 15.1 (Friedmann equations) Field equations:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi\rho}{3} + \frac{\Lambda}{3} - \frac{k}{a^2} \tag{15.7}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3}(\rho + 3p) + \frac{\Lambda}{3} \tag{15.8}$$

Conservation equation:

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0 \tag{15.9}$$

15.2 Equation of state and beginnings of domination

Theorem 15.2 (Equation of state) Perfect fluids observe the so-called *equation of state*:

$$p = w\rho \tag{15.10}$$

where w is a constant called the *equation of state parameter*.

Derivation 15.2 (Density evolution with various ws) We find that the time evolution of density directly relates to the equation of state parameter. The derivation is quite formulaic: We derive p in terms of ρ, insert p into Equation 15.9 and solve for $\rho(t)$.

- $w = 0$: 'normal' matter (dust) and dark matter. The pressure is vanishing because the matter particles do not interact. This leads to $p = 0$ and

$$\rho(t) \propto 1/a^3(t) \tag{15.11}$$

- $w = 1/3$: radiation. This leads to $p = \rho/3$ and

$$\rho(t) \propto 1/a^4(t) \tag{15.12}$$

- $w = -1/3$: curvature. This leads to $p = -\rho/3$ and

$$\rho(t) \propto 1/a^2(t) \tag{15.13}$$

- $w = 1$: stiff matter[a]. This leads to $p = \rho$ and

$$\rho(t) \propto 1/a^6(t) \tag{15.14}$$

  Stiff matter is a hypothetical matter whose speed of light is equal to its speed of sound. This is verified in its equation of state.

- $w = -1$: cosmological constant/dark energy. this leads to $p = -\rho$ and

$$\rho = \text{constant} \tag{15.15}$$

[a] So-called because its high rigidity makes it resistant to being compressed.

We will later see that different equation of state parameters lead to scenarios in which different parameters act as the main driving force of the universe's expansion. This marks the beginning of the concept of *domination* which will be revisited later on.

15.3 Energy-momentum conservation

The definition of perfect fluids can be adapted for cosmology.

Definition 15.1 (Perfect fluids in cosmology) A *perfect fluid* is a fluid that can be completely characterised by its rest frame mass density ρ_m and isotropic pressure p.

We again recall the stress-energy tensor in a perfect fluid in Equation 3.18. In comoving coordinates, one has $u = (1, 0, 0, 0)$. As such, the stress-energy tensor of a perfect fluid in comoving coordinstes is

$$T_{\mu\nu} = \rho u_\mu u_\nu = \mathrm{diag}(\rho, p, p, p) \quad \text{or} \quad T^\mu_\nu = \mathrm{diag}(-\rho, p, p, p) \tag{15.16}$$

Remark 15.1 But is there something that we have forgotten?
Surprise, surprise! Due to the existence of the energy-momentum tensor, energy and momentum conservation are actually not independent[1]. Recalling that the Einstein tensor satisfies the contracted Bianchi identity in Equation 5.25. If $\Lambda = 0$, the same applies to the stress-energy tensor:

$$\nabla_\mu T^{\mu\nu} = 0 \tag{15.17}$$

The implication? We then recall our previous derivation:

$$\dot{\rho} - 3H(\rho + p) = 0 \tag{15.18}$$

This clearly shows that the total energy of the universe is *not* conserved.
However, there indeed exist conserved quantities. Assuming a co-moving volume v, we have the *proper volume*

Definition 15.2 (Proper volume)

$$V(t) = a^3 v \tag{15.19}$$

which is the 'scaled' version of the volume.

Differentiation by t yields

$$\frac{dV}{dt} = 3a^2\dot{a} = 3HV \tag{15.20}$$

Note 15.1 If we insert this into the definition of entropy, we will find that instead of energy, the property conserved is *entropy*.

Remark 15.2 As entropy represents information, we can thus say that the universe does not create new information.
Again, we consider the following cases:

- $w = 0 \to p = 0$: energy is conserved! Matter particles do not interact.
- $w = 1/3 \to p = \rho/3$: radiation experiences redshift in their frequency $\propto a^{-1}$ in the expanding universe. Thus, energy density decays with a^{-4}.
- $w = -1 \to p = -\rho$: we now understand the notion of cosmological constant: it is constant in the sense that *the energy density stays constant*. The universe expands and the cosmological constant 'provides' the necessary contribution to keep the energy density constant, i.e. $E(t) \sim V(t)$.

Theorem 15.3 (Weak energy condition) The weak energy condition stipulates that for every timelike vector field, the matter density observed by the corresponding observers is always non-negative. i.e.

$$\rho T_{\mu\nu} V^\mu V^\nu \geq 0 \tag{15.21}$$

where V is a time-like vector.

Quote 15.1 There have been some speculations, but for the last 25 years, nobody has been able to solve that problem. (...) And of course, we can dream of everything, and if that works with data is another thing.

Betti Hartmann, on energy conditions, 30 January 2024

[1] This explains the existence and conservation of four-momentum in SR.

15.4 History and ultimate fate of the universe

Quote 15.2 Well, let's see. First, the Earth cooled, and then the dinosaurs came. But they got too big and fat, so they all died and they turned into oil. And then the Arabs came and they bought Mercedes-Benzes. And Prince Charles started wearing all of Lady Di's clothes. I couldn't believe it. He took her best summer dress out of the closet and put it on and went to town.

Johnny, in Airplane

Using the Hubble parameter in the first Friedmann equation, we have

$$H^2 = \frac{8\pi\rho}{3} + \frac{\Lambda}{3} - \frac{k}{a^2} \to 1 = \frac{8\pi\rho}{3H^2} + \frac{\Lambda}{3H^2} - \frac{k}{a^2H^2} \tag{15.22}$$

We can thus define a series of parameters.

Definition 15.3 (Density parameter) The *density parameters* are

$$\underbrace{\Omega_m = \frac{\rho_m}{\rho_c} = \frac{8\pi\rho}{3H^2}}_{\text{matter}} \quad \underbrace{\Omega_r = \frac{\rho_r}{\rho_c}}_{\text{radiation}} \quad \underbrace{\Omega_k = \frac{\rho_k}{\rho_c} = \frac{k}{a^2H^2}}_{\text{curvature}} \quad \underbrace{\Omega_\Lambda = \frac{\rho_\Lambda}{\rho_c} = \frac{\Lambda}{3H^2}}_{\text{dark energy}} \tag{15.23}$$

where we have the *critical density*

$$\rho_c = \frac{3H^2}{8\pi G} - \frac{3H^2}{8\pi} \tag{15.24}$$

which is the density that leads to a spherically flat universe. G is omitted as usual.

Remark 15.3 The radiation-dominated density parameter is missing in the Friedmann equations because they are formulated under the assumption that radiation's contribution to the universe's overall energy density is *smol*.
We can then consider the universe to be *dominated* by several quantities during various eras. By *domination* we mean the nature of the scale factor, which, as we recall, contributes to expansion.
Considering a radiation-inclusive version of the first Friedmann equation[2], we insert the density parameters:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}(\rho_m + \rho_r - \rho_k + \rho_\Lambda) \tag{15.25}$$

and find rather miraculously that

Theorem 15.4 (Density parameters)

$$\Omega_m + \Omega_r - \Omega_k + \Omega_\Lambda = 1 \tag{15.26}$$

Recalling the equation of state, we remember that each equation is proportional to some power of $a(t)$. We denote the proportionality constant as $\tilde{C}_w$ and scale it such that $C_w = \frac{8\pi\tilde{C}_w}{3}$.

Theorem 15.5 (Domination in various eras) Solving for the proportionality constant in each era, we have

$$\dot{a}^2 = \frac{C_0}{a} + \frac{C_{1/3}}{a^2} - k + \frac{\Lambda}{3}a^2 \tag{15.27}$$

where Λ is simply $C_{-1/3}$!

- At early universe, $C_{1/3}$ (i.e. radiation) dominates.
- At a later stage, C_0/a (i.e. matter) dominates.
- As k is believed to be 0, curvature never dominates.
- And from then up to now, C_{-1} (i.e. dark energy) dominates.

[2]Thus accounting for the entire history of the universe.

Going back to the density-scale parameter relations derived by the equation of state, one can consider the first Friedmann equation. For each domination, ignore all non-dominating terms and insert the corresponding density-scale parameter relationship. By doing so, one can ultimately derive the scale parameters in terms of time and some term a_0 made up of constants.
We have, for radiation, matter, curvature and dark energy-dominated universes:

Definition 15.4 (Various scale factors)

$$a(t) = a_0 \left(\frac{t}{t_0}\right)^{1/2} \quad a(t) = a_0 \left(\frac{t}{t_0}\right)^{2/3} \quad a(t) = a_0 \left(\frac{t}{t_0}\right) \quad a(t) = a_0 \exp\left[\pm\sqrt{\frac{\Lambda}{3}}(t - t_0)\right] \tag{15.28}$$

Note 15.2 In scenarios where it simplifies calculations and does not alter the interpretation of the physical situation being modelled (read: in the end-of-year exams), t_0 can be regarded as 1.

Remark 15.4 When dark energy dominates, we can infer that

$$\frac{\dot{a}}{a} = H = \pm\sqrt{\frac{\Lambda}{3}} \tag{15.29}$$

In the final sections of this chapter, we will investigate specific solutions that arise from these scale factors.
We then zoom in to the current universe and assume only matter and dark energy[3], we can define today's total density parameter $\Omega = \Omega_m + \Omega_\Lambda$[4] and therefore associate the curvature parameter k with Ω:

$$\frac{k}{a^2 H^2} = \Omega - 1 \tag{15.30}$$

In effect, it associates matter and energy content with geometry. We can thus infer the ultimate fate of the universe:

- $\Omega < 1 \to k = -1$: *Open universe* (i.e. the speed of expansion approaches a constant)
- $\Omega = 1 \to k = 0$: *Flat universe* (i.e. the speed of expansion approaches zero)
- $\Omega > 1 \to k = +1$: *Closed universe* (i.e. big crunch)

In fact, we can now even solve for the time evolution of the density parameter:

Theorem 15.6 (Time evolution of the density parameter)

$$\Omega(t) = 1 + \frac{k}{\dot{a}^2(t)} \tag{15.31}$$

Quote 15.3 Suppose the trilobites studied cosmology 500 million years ago. They would obtain this same equation, but with their t_0 dated 500 million years before ours.

Mitchell A. Berger, confirming that he is a trilobite, 2004

[3] The curvature term, although not negligible, is not included as it is a representation of the geometry of the universe instead of that of matter or energy.
[4] Simply add Ω_r when radiation is accounted for.

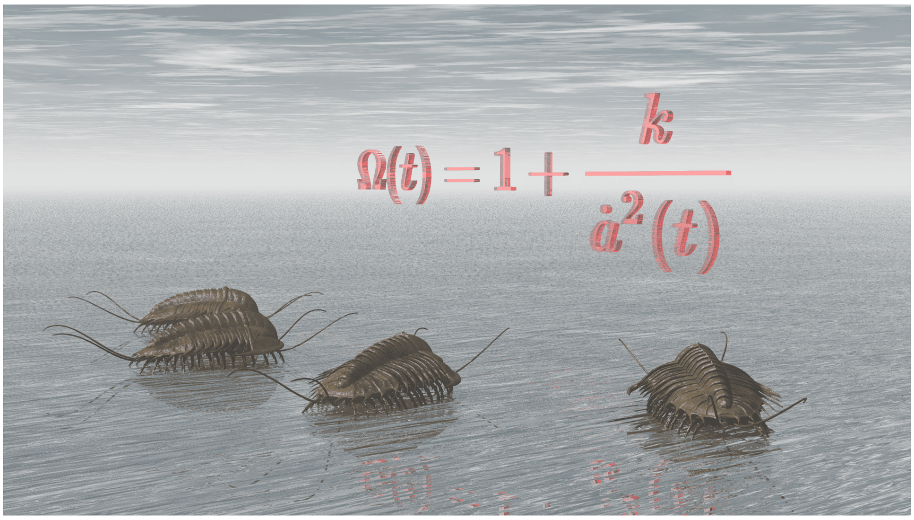


Figure 15.1: Trilobites discuss the density parameter (3D trilobite by Thorsten Brand).

We can also investigate if the expansion of the universe is accelerating or otherwise from the so-called *deceleration parameter*:

Definition 15.5 (Deceleration parameter)

$$q(t) = \frac{a(t)\ddot{a}(t)}{\ddot{a}^2(t)} \tag{15.32}$$

If we insert this into the last Friedmann equation, we get

$$q = \frac{4\pi}{3H^2}(\rho + 3p) \tag{15.33}$$

We conclude that the sign of q, i.e. whether the universe is decelerating or accelerating depends on the sign of $\rho + 3p$.

- For matter and radiation, $q > 0$ and cosmic expansion is decelerating.
- For the cosmological constant, $q < 0$ and cosmic expansion is accelerating.

Remark 15.5 The early universe was dominated by matter or radiation. In this case, $a \sim t^\gamma$ in which $0 < \gamma < 1$. We observe that $a > 0$, $\dot{a} > 0$ and $\ddot{a} < 0$.

15.5 Age of the universe

Naively, we might be tempted to insert $H_0 = 2.3 \times 10^{-18}$ into

$$t_0 = \frac{a_0 t}{a(t)} \tag{15.34}$$

which gives $t_0 = 4.4 \times 10^{18}$.

> **Quote 15.4** That's too much. The universe is not as old as that. The universe is approximately 13.8 billion years old.
>
> *Betti Hartmann, 6 February 2024*

Derivation 15.3 (Age of the universe) Recalling Equation 15.26, one can write down the expression

$$\frac{H^2}{H_0^2} = \frac{H^2}{H_0^2}(\Omega_m + \Omega_r - \Omega_k + \Omega_\Lambda) \tag{15.35}$$

Meanwhile, we note that

$$\frac{H^2}{H_0^2} = \left(\frac{a_0}{a}\right)^{3(1+w)} \tag{15.36}$$

Inserting this and taking the total derivative, we find

$$dt = \frac{da}{aH_0}\left[\left(\frac{a_0}{a}\right)^3 \Omega_m + \left(\frac{a_0}{a}\right)^4 \Omega_r - \left(\frac{a_0}{a}\right)^2 \Omega_k + \Omega_\Lambda\right]^{-1/2} \tag{15.37}$$

We consider both cases of the flat universe.

- **Matter only**

$$\Omega_r = \Omega_k = \Omega_\Lambda = 0 \quad \Omega_m = 1 \tag{15.38}$$

Integrating this equation, we can calculate that the age of the universe is approximately 9 billion years, but we have stars that are about 14 billion years old. As such the purely matter-dominated model of the universe is incorrect.

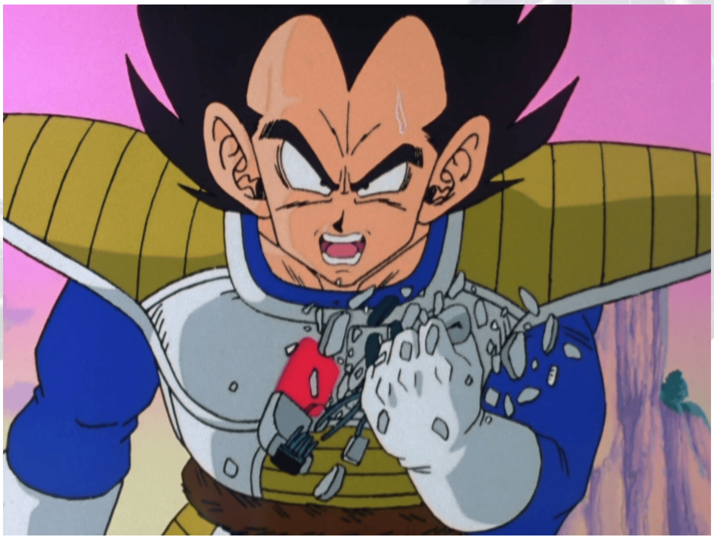


Figure 15.2: It's over 9,000,000,000!

- **Matter and cosmological constant**

$$\Omega_r = \Omega_k = 0 \quad \Omega_m = 1 - \Omega_\Lambda \quad \Omega_\Lambda \neq 0 \tag{15.39}$$

In this case, we find that the age of the universe is about 13 billion years old, a much better value.

Exercise 15.1 Calculate the age of the universe for the matter and radiation-dominated cases.

Quote 15.5 The radiation in the current era is negligible.

Betti Hartmann, 20 February 2024

15.6 Redshift

In Part I, we investigated redshift in the broader context of GR. In terms of the scale factor, it can be defined as

Definition 15.6 (Redshift in cosmology)

$$z = \frac{a(t_{\text{obs}})}{a(t)} - 1 \tag{15.40}$$

Taking the total derivative of this expression gives

$$dz = -\frac{a(t_{\text{obs}})}{a(t)^2}\dot{a}dt = -\frac{a(t_{\text{obs}})}{a(t)}Hdt = (1+z)Hdt \tag{15.41}$$

Integrating with respect to t:

Theorem 15.7 (Signals)

$$t_{\text{obs}} - t_{\text{e}} = \int_0^z \frac{1}{(1+z')H}dz' \tag{15.42}$$

where t_{e} is the time when the signal takes place.

The geodesics can also be investigated. As redshift involves EM signals, we consider radial null geodesics, and the metric reduces to

$$dr = \frac{da}{a^2H} \tag{15.43}$$

This can then be integrated to find the *distance-redshift relation.*

Theorem 15.8 (Distance-redshift relation)

$$r(z) = \frac{1}{a_0H_0}\int_0^z dz\left[(1+z)^3\Omega_m + (1+z)^4\Omega_r + \Omega_\Lambda\right]^{-1/2} \tag{15.44}$$

15.7 Particle horizons

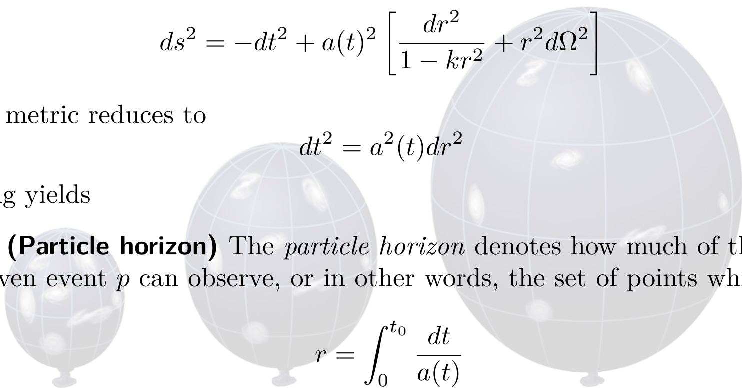
When analysing cosmological models the following question naturally arises: How much of our universe can be observed in principle at some event p? The particle horizon answers this problem[5].

Definition 15.7 (Conformal time) As the name suggests, the *conformal time* is time with the scale factor taken into consideration

$$d\eta = dt/a(t) \tag{15.45}$$

The FLRW model contains two horizons:

Derivation 15.4 (Particle horizon) We begin with the FLRW metric in isotropic coordinates:



$$ds^2 = -dt^2 + a(t)^2\left[\frac{dr^2}{1-kr^2} + r^2d\Omega^2\right] \tag{15.46}$$

For a photon, the metric reduces to

$$dt^2 = a^2(t)dr^2 \tag{15.47}$$

Further simplifying yields

Definition 15.8 (Particle horizon) The *particle horizon* denotes how much of the universe that an observer at a given event p can observe, or in other words, the set of points which are just coming into view.

$$r = \int_0^{t_0} \frac{dt}{a(t)} \tag{15.48}$$

where r is the particle horizon.

Remark 15.6 In terms of the comoving distance, the particle horizon is equal to the conformal time η that has passed since the Big Bang.

[5]Although you can make as convincing a case that it has created more problems than it has solved.

Definition 15.9 (Cosmological event horizon) If we set $t_0 \to \infty$ instead, we get the *cosmological event horizon*[a], which is the region of space from which the observer at $r = 0$ can never receive information:

$$r_\gamma(t) = \eta(\infty) = \eta(t) \tag{15.49}$$

[a]Not to be confused with its black hole cousin.

We can thus deduce various particle horizons:

Various particle horizons in the FLRW model

Ω_r	Ω_Λ	Ω_m	Horizon (r_γ)	Conclusion
0	0	1	$2/(H_0 a_0)$	Finite horizon, implying it exists
-1	0	0	a_0/H_0	Finite horizon, implying it exists
0	-1	0	$1/H_0(-1/a)\vert_0^{a_0}$	Horizon approaches infinity, implying *no* horizon

Exercise 15.2 Prove these horizon results.

Definition 15.10 (Causal patch) The *causal patch* D refers to a region of the universe that is causally connected. i.e. the collection of points which can communicate with each other. It is given by

$$D = \frac{H^{-1}}{a(t)} = -\frac{\ddot{a}(t)}{\dot{a}^2(t)} \tag{15.50}$$

Two regions greater than this length cannot communicate with each other.

Remark 15.7 The particle horizon is essentially the outer boundary of a causal patch.
The particle horizon is not to be confused with the *apparent horizon* or the *gravitational horizon*, which is equivalent to the *Hubble distance* $R_H(t)$:

Definition 15.11 (Apparent horizon) The *apparent* is the distance at which galaxies move away from us with the speed of light c.

$$R_H(t) = \frac{c}{H(t)} \tag{15.51}$$

Now we take a few pages off as an intermission, and discuss some exact solutions of Friedmann equations.

15.8 Solutions with $\Lambda = 0$

These solutions are also known as the *Friedmann solutions*. The derivations of these solutions from the $\Lambda = 0$ Friedmann equations are quite formulaic. We can simplify the field equations as

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}\rho - \frac{k}{a^2} \qquad \frac{\ddot{a}}{a} = -\frac{4\pi}{3}\rho \tag{15.52}$$

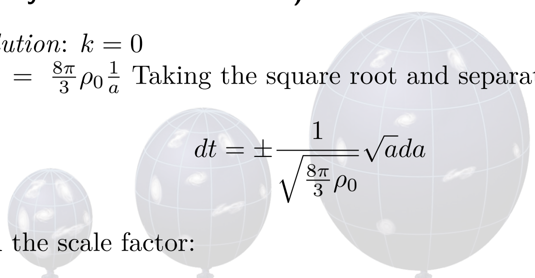
while the conservation equation stays the same. Integrating it yields

$$\rho = \frac{\rho_0}{a^3} \tag{15.53}$$

Derivation 15.5 (Matter-only Friedmann solutions) We first consider the matter-only case:

- *Einstein-de Sitter solution*: $k = 0$
  Rearranging gives $\dot{a}^2 = \frac{8\pi}{3}\rho_0\frac{1}{a}$ Taking the square root and separating variables yields

$$dt = \pm\frac{1}{\sqrt{\frac{8\pi}{3}\rho_0}}\sqrt{a}da \tag{15.54}$$



  Integrate and we find the scale factor:

$$a(t) = \pm(6\pi\rho_0)^{1/3}(t - t_0)^{3/2} \tag{15.55}$$

We can reasonably assume the universe started with zero volume[a]. i.e. $t_0 = a(t_0) = 0$

$$a(t) \propto t^{2/3} \tag{15.56}$$

The horizon is at

$$r = 3t/a \tag{15.57}$$

The other solutions follow this sequence: use the first Friedmann equation, take the square root, separate variables and integrate.

- *Friedmann closed universe* or *closed universe model*: $k = 1$
  For simplicity, we define a new set of coordinates:

$$\left(\frac{8\pi\rho_0}{3a} - 1\right) = \tan u \tag{15.58}$$

  Subbing this in, we can find

$$a(u) = \frac{4\pi}{3}\rho_0(1 - \cos u) \tag{15.59}$$

  where u is a new set of coordinates. This is the standard parametrisation of a cycloid.
  Again, $t_0 = u_0 = a(u_0) = 0$. This universe will keep expanding until it reaches $a_{\max} = 8\pi\rho_0/3$, when it doesn't anymore. The corresponding $t_{max} = 4\pi^2\rho_0/3$. At $t = 8\pi^2/3\rho_0$, the so-called *big crunch* takes place.

- *Friedmann open universe* or *open universe model*: $k = 1$
  We likewise employ an u, finding

$$a(u) = \frac{4\pi\rho_0}{3}(\cosh 2u - 1) \tag{15.60}$$

  This solution expands, as $t \to \infty$ it expands asymptotically like $a(t) \propto t$.

[a] In the beginning, Christian G. Böhmer created the observed universe and the metric. Now the metric was formless and rankless, flat space was over the horizons of the geodesics, (...) Christian G. Böhmer saw all that he had made, and it was very isotropic. And there was gravitational lensing, and there was cosmic time - the sixth day.

Derivation 15.6 (Radiation-only Friedmann solutions) Now we consider the radiation-only case:

- *Lemaître-Tolman metric*: $k = 0$

$$a(t) \propto t^{1/2} \quad a(t) = \sqrt{2}(8\pi\rho_0/3)^{1/4}t^{1/2} \tag{15.61}$$

  The density observes $\rho(t) \propto 1/t^2$, which is the same as the energy density of a matter-dominated universe.
  The horizon is at

$$r = 2t/a \tag{15.62}$$

Derivation 15.7 (Curvature-only Friedmann solutions) Now we consider the curvature-only case:

- *Milne model*: $k = -1$

$$a(t) \propto t \quad a(t) = H_0\sqrt{\Omega_k}t \tag{15.63}$$

  As mentioned previously, there is no curvature-dominated era, and the Milne model is inconsistent with cosmological observations and thus unphysical.

15.9 Solutions with $\Lambda \neq 0$

These solutions are also known as the *Friedmann-Lemaître solutions*. The derivations of these solutions are also quite formulaic.

Derivation 15.8 (Friedmann-Lemaître solutions) We likewise consider possible curvature parameters:

- *Einstein's static universe*: $k = 1$
  Remark 15.8 In his original discussion, Einstein assumed the universe to be filled solely with matter, i.e. he set $\rho_r = p_r = 0$.
  Static means that $a(t)$, $\rho(t)$ and $p(t)$ are all constant, and we have the so-called *static scale parameter*:
  $$a = \frac{1}{\sqrt{4\pi G\left(\rho_m + \rho_r + p_\lambda\right)}} \tag{15.64}$$
  As a is constant, its derivative of any order is zero. Using this to our advantage in the Friedmann equations, we can derive that
  $$\Lambda = 4\pi(\rho + 3p) \tag{15.65}$$
  Formally the age of a static universe is infinite. This is because $a = 0$ which causes the denominator of the time equation $H = \frac{\dot{a}}{a} = 0$.

- *De Sitter solution*: $k = 0$, $\rho = p = 0$ This is a solution without matter. we can formulaically derive
  $$a(t) = e^{\sqrt{\Lambda/3t}} \tag{15.66}$$
  In investigating the metric we realise that the coordinates (t, x, y, z) do not cover the complete manifold. For a more complete coordinate system, we set the manifold to be the 4D hyperbolic space $\mathbb{H}^4$ embedded in a 5D Minkowski space $\mathbb{R}^5$. The metric is thus
  $$ds^2 = -dv^2 + dw + 2 + dx^2 + dy^2 + dz^2 \tag{15.67}$$
  This can be accomplished by a series of coordinate transformations or *slices* which we call the *de Sitter foliations*:

  Definition 15.12 (De Sitter foliations)
  $$\begin{aligned}
  v &= \sqrt{\frac{3}{\Lambda}}\sinh\sqrt{\frac{\Lambda}{3}}t \\
  w &= \sqrt{\frac{3}{\Lambda}}\cosh\sqrt{\frac{\Lambda}{3}}t\cos\chi \\
  x &= \sqrt{\frac{3}{\Lambda}}\cosh\sqrt{\frac{\Lambda}{3}}t\sin\chi\cos\theta \\
  y &= \sqrt{\frac{3}{\Lambda}}\cosh\sqrt{\frac{\Lambda}{3}}t\sin\chi\sin\theta\cos\phi \\
  z &= \sqrt{\frac{3}{\Lambda}}\cosh\sqrt{\frac{\Lambda}{3}}t\sin\chi\sin\theta\sin\phi
  \end{aligned} \tag{15.68}$$

  Through the coordinate transformations
  $$T = \sqrt{\frac{3}{\Lambda}}\log\left[\frac{v+w}{\sqrt{3/\Lambda}}\right] \quad X = \sqrt{\frac{3}{\Lambda}}\frac{x}{v+w} \quad Y = \sqrt{\frac{3}{\Lambda}}\frac{y}{v+w} \quad Z = \sqrt{\frac{3}{\Lambda}}\frac{z}{v+w} \tag{15.69}$$
  the metric in so-called static coordinates can be recovered. For the sake of familiarity, we can rewrite capitals to small letters:
  $$ds^2 = -dt^2 + e^{2\sqrt{\Lambda/3t}}(dx^2 + dy^2 + dz^2) \tag{15.70}$$
  Remark 15.9 We can do the same thing in spaces with zero and negative curvature, both of which yield 5 slices each as well.

Derivation 15.9 (Staticity of the de Sitter solution) Consider a time translation to the de Sitter solution $t \to t' + T$. The metric then becomes

$$ds^2 = -dt'^2 + e^{2\sqrt{\Lambda/3}(t'+T)}(dz^2 + dy^2 + dz^2) \tag{15.71}$$

Only the spatial part is rescaled. We use a rather dirty trick and impose a second rescaling

$$X'^\mu = e^{2\sqrt{\Lambda/3}T} X^\mu \tag{15.72}$$

and everything returns to

$$ds^2 = -dt'^2 + e^{2\sqrt{\Lambda/3}t'}(dz'^2 + dy'^2 + dz'^2) \tag{15.73}$$

15.10 Solutions without matter & radiation

These solutions are also known as the *vacuum solutions*.

Derivation 15.10 (Vacuum solutions) Considering various cosmological constants and curvature parameters:

- *Minkowski space* (or at least part of it...): $\Lambda = 0$
  While we predictably have $k = 0$, $k = -1$ and $k = 1$, none of them yields a solution.
- *De Sitter space*: $\Lambda > 0$

  $$\left(\frac{\dot{a}}{a}\right)^2 = \frac{\Lambda}{3} - \frac{k}{a^2} \to \dot{a}^2 = \underbrace{\frac{\Lambda}{3}a^2 - k}_{\geq 0} \tag{15.74}$$

  for which we have

  - $k = 1$

    $$a^2 = \frac{\Lambda}{3} \pm 1 \to a(t) = \sqrt{\frac{3}{\Lambda}} \cosh\left(\pm\sqrt{\frac{\Lambda}{3}}t\right) \tag{15.75}$$

  - $k = -1$

    $$a(t) = \sqrt{\frac{3}{\Lambda}} \sinh\left(\pm\sqrt{\frac{\Lambda}{3}}t\right) \tag{15.76}$$

  - $k = 0$

    $$a(t) = \exp\left(\pm\sqrt{\frac{\Lambda}{3}}t\right) \cdot \tilde{c} \tag{15.77}$$

  Each curvature parameter represents a different choice of coordinates.
- *Anti-de Sitter space*: $\Lambda < 0$

  $$\dot{a}^2 = \underbrace{\frac{\Lambda}{3}a^2 - k}_{\geq 0} \tag{15.78}$$

  Remark 15.10 As the RHS has to be positive, $k = 1$ and $k = 0$ do not work.

  - $k = -1$

    $$a(t) = \sqrt{\frac{|\Lambda|}{3}} \sinh\left(\pm\sqrt{\frac{|\Lambda|}{3}}t + c\right) \tag{15.79}$$

  Remark 15.11 Anti-de Sitter space is negatively curved. Hence, it is unphysical not very relevant to cosmology. However, the study of its similarities with conformal field theory (CFT) is a significant field known as *AdS/CFT correspondence*.

Remark 15.12 True to their name, all three spaces are solutions to the Einstein equations $G_{\mu\nu} = \Lambda g_{\mu\nu} = 0$.

Figure 15.3: De Sitter standing

Chapter 16

Problems with the FLRW model

So far, our discussion has been restricted to classical cosmology. Unsurprisingly, however, the FLRW model runs into a few problems, which will be illustrated in this chapter.

16.1 Flatness problem

This is also known as the *oldness problem*. Solutions using the FLRW metric tell us that the universe is spatially flat

$$k = 0 \to \rho = \rho_c \to \Omega_0 = 1 \tag{16.1}$$

Observations give, with generous bounds:

$$0.9 \leq \Omega_0 \leq 1.1 \tag{16.2}$$

This itself is a problem. Why is Ω_0 so close to 1?
This is not even the only component of the flatness problem. We dial the clock back and investigate the early universe in two scenarios. How does the FLRW model behave in matter and radiation-dominated universes?

Derivation 16.1 (Matter-dominated universe) We consider the matter-dominated case, in which

$$a \propto t^{2/3} \quad \text{and} \quad H \propto 1/t \to aH \propto t^{-1/3} \tag{16.3}$$

and therefore

$$|\Omega - 1| \propto t^{2/3} \tag{16.4}$$

Noting that matter started to dominate at $a/a_0 \approx 1000$, we conclude that for a flat universe today, we need to make sure that at the time matter started to dominate, Ω was a factor of 10^{-3} of that today.

> **Quote 16.1** In physics, this is called fine-tuning, and not very natural.
>
> *Betti Hartmann, 23 February 2024*

Derivation 16.2 (Radiation-dominated universe) We consider the radiation-dominated case, in which

$$a \propto t^{1/2} \quad \text{and} \quad H \propto 1/t \to aH \propto t^{-1/2} \tag{16.5}$$

and therefore

$$|\Omega - 1| \propto t \tag{16.6}$$

We know that the age of the universe is $t_a \approx 10^{17}$s and that the Planck time is $t_{\mathrm{Pl}} \approx 10^{-43}s$. This results in

$$t/t_0 \approx 10^{-60} \tag{16.7}$$

This is extreme fine-tuning: we have to fine-tune Ω to be =1 in 60 digits to get it to $\Omega_0 = 1$ today!

Remark 16.1 The universe being spatially flat is not a 'natural' choice. It requires extreme fine-tuning in a very hot, very dense early universe!

16.2 Horizon problem

Formation of the CMB roughly 380,000 years after the Big Bang. matter and light decoupled and photons were free to travel without disturbance. We can then define the conformal-adjusted proper distance by recalling the particle horizon in Equation 15.7:

Definition 16.1 (Proper distance)

$$R_p(t_2) = a(t_2) \int_{t_1}^{t_2} \frac{dt}{a(t)} \tag{16.8}$$

We recall that for the scale factor $a(t)$, $a \propto t^{\beta}$:

- In a matter-dominated universe, we have $\beta = 2/3$.
- In a radiation-dominated universe, we have $\beta = 1/2$.

$$R_p(t_2) \sim \frac{1}{1-\beta} t_2^{\beta} (t_2^{1-\beta} - t_1^{1-\beta}) \tag{16.9}$$

The time since the Big Bang ($t_1 = 0$) until the decoupling of matter (formation of CMB) $t_2 = t_{\mathrm{ls}}$ (last scattering of photons). This was radiation-dominated ($\beta = 1/2$). i.e. regions separated by distances that are no larger than $2t_{\mathrm{ls}}$ were never in causal contact.

$$R_p(t_{\mathrm{ls}}) \sim 2t_{\mathrm{ls}} \tag{16.10}$$

Remark 16.2 That is to say, the causal patch has expanded due to the expansion of the universe. Its size is given by

$$\frac{a(t_0)}{a(t_{\mathrm{ls}})} R_p(t_{\mathrm{ls}}) = 1 + Z_{\mathrm{ls}} R_p(t_{\mathrm{ls}}) \tag{16.11}$$

The distance travelled by photons of the CMB since $t = t_{\mathrm{ls}}$ until today is the integral from before $t_1 = t_{\mathrm{ls}}$, $t_2 = t_0$:

$$R_p(t_0) = a(t_0) \int_{t_{\mathrm{ls}}}^{t_0} \frac{dt}{a(t)} \tag{16.12}$$

We assume that the universe is matter-dominated ($\beta = 2/3$):

$$R_p(t_0) \sim 3t_0^{2/3}(t_0^{1/3} - t_{\mathrm{ls}}^{1/3}) \sim 3t_0 \quad \text{for } t_0 \ll t_{\mathrm{ls}} \tag{16.13}$$

The light cone angle θ is then

$$\theta \approx \frac{(1+z_{\mathrm{ls}})R_p(t_{\mathrm{ls}})}{R_p(t_{\mathrm{ls}})} \approx \frac{(1+z_{\mathrm{ls}})2t_{\mathrm{ls}}}{3t_0} \approx 0.02 \approx 1.15^{\circ} \tag{16.14}$$

This means that, looking at patches of the CMB, under angles $> 1.15^{\circ}$, we see photons that have not been at causal contact at the moment of last scattering is formation of the CMB.

Note 16.1 But the CMB is isotropic to $1 : 10^5$ on the full sky! Why should that happen if photons have not been in causal contact?

16.3 Monopole problem

Some grand unified theories (GUTs) posit that fundamental forces are not fundamental forces but only arise due to spontaneous symmetry. In high temperatures, a number of heavy particles emerge, including the infamous *magnetic monopole* which is at about 10^{15} GeV. Given the universe was initially very hot, these monopole particles would have emerged. And more disturbingly...

Note 16.2 ...the universe would have recollapsed!

Remark 16.3 Meanwhile, we note that no monopoles have been detected up to now. With these problems in mind, we advance to modern cosmology where we introduce *inflation* and use it toi resolve the three problems.

Chapter 17

Inflation

Quote 17.1 You are probably too young for this.

Betti Hartmann, on an ongoing research topic, 5 March 2024

In this chapter, we will see that the previous problems can be eliminated by considering cosmological inflation driven by a scalar field. The entire process of inflation, from the slow roll phase to the end of inflation will be introduced.

17.1 Cosmological inflation

Inflation is essentially accelerated expansion of space. It can be formally defined as

Theorem 17.1 (Inflation) Inflation occurs when

$$\ddot{a} > 0 \tag{17.1}$$

or if the derivative of the so-called *comoving Hubble length* $H^{-1}/a(t)$ is negative:

$$\frac{d}{dt}\frac{H^{-1}}{a(t)} = \underbrace{\frac{d}{dt}\frac{1}{\dot{a}(t)}}_{①} = -\frac{\ddot{a}(t)}{(\dot{a})^2(t)} < 0 \tag{17.2}$$

Remark 17.1 The comoving Hubble length is the size of the casual patch and the two are conceptually identical.

Quote 17.2 Note that some authors use slightly different definitions and they might differ in some tiny detail. However, I have no objection to you point.

Christian G. Böhmer, on Remark 17.1, 15 April 2024

Remark 17.2 ① is important in that physically, it being smaller than 0 simply denotes that the scaled version of an arbitrary length (in this case, of 1) is becoming increasingly small due to inflation, even though the length itself has undergone no change.

Quote 17.3 Inflation means that the observable universe becomes smaller very rapidly. So, yes, you can call this scaling down but it is not the preferred notation.

Christian G. Böhmer, on Remark 17.2, 15 April 2024

In other words:

Quote 17.4 If we had a standard ruler with standard length L and we measured the length of an object to be rL, then L is the quantity that's changing, not r. i.e. the length with respect to the standard length remains the same.

Abhijeet Vats, on Remark 17.2, 15 April 2024

17.2 Inflation saves the day

Now we will see how inflation solves the FLRW problems.

- **Flatness problem**: as $\frac{d}{dt}\frac{H^{-1}}{a(t)} < 0$, $|\Omega - 1|$ is driven towards zero.
- **Horizon problem**: The horizon problem is solved because of the reduction of the causal patch during inflation[1].
- **Monopole problem**: Suppose monopoles were produced before inflation - the density of primordial monopoles would have decreased exponentially during the expansion.

17.3 Scalar fields

As before, the starting point is the FLRW metric as seen in Equation 14.21. We define a new real scalar field, or a so-called *inflation field*, which dominates over all other energy-momentum content during inflation.

Definition 17.1 (Lagrangian density of a real scalar field) The Lagrangian density for the real scalar field $\phi(t) \in \mathbb{R}$ is

$$\mathcal{L}_\phi = \underbrace{-\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi}_{\text{kinetic + gradient}} - \underbrace{V(\phi)}_{\text{potential}} \tag{17.3}$$

This is a standard result one will commonly find in classical and quantum field theories.

The stress-energy tensor under such a scalar field is given in Equation 3.14. Thus, for $\phi = \phi(t) \in \mathbb{R}$ abd the FLRW metric, we have

$$\mathcal{L}_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi) \quad T_{00} = \frac{1}{2}\dot{\phi}^2 + V(\phi) \tag{17.4}$$

The other diagonal components of the energy-momentum tensor are

$$T_{xx} = T_{yy} = T_{zz} = \left(\frac{1}{2}\dot{\phi}^2 - V(\phi)\right)a^2 \tag{17.5}$$

Noting that the components are simpny density and pressure, we can write

Definition 17.2 (Stress and press under a scalar field)

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi) \quad p_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi) \tag{17.6}$$

We then consider the dynamics of the scalar field. By inserting 17.4 into the Euler-Lagrange equations seen in Equation 2.33, we find that

$$\ddot{\phi} + \underbrace{\frac{3\dot{a}}{a}\dot{\phi}}_{\text{friction term}} + \frac{dV}{d\phi} = 0 \tag{17.7}$$

The dynamics of an FLRW universe filled solely with a real scalar field $\phi(t) \in \mathbb{R}$ is thus as follows:

Theorem 17.2 (Inflationary Friedmann equations)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}\mathcal{L}_\phi = \frac{8\pi}{3}\left(\frac{1}{2}\dot{\phi}^2 - V(\phi)\right) \tag{17.8}$$

[1] See previous bullet point.

$$-\frac{\ddot{a}}{a} = \frac{4\pi}{3}(\rho_\phi + 3p_\phi) = \frac{8\pi G}{3}((\dot{\phi})^2 - V(\phi)) \tag{17.9}$$

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} + \frac{dV}{d\phi} = 0 \tag{17.10}$$

Remark 17.3 We do not really use Equation 17.10. Nonetheless, it is there for completeness.

> **Quote 17.5** Because it's there.
>
> *George Mallory, on his motivation of climbing Mount Everest, 18 March 1923*

17.4 Slow roll

We consider the so-called *slow roll phase*, which describes a regime in which the inflationary phase of the universe's expansion happens at a relatively slow, controlled rate.

Definition 17.3 (Slow roll conditions) The *slow roll conditions* say that the inflaton potential must be flat (compared to the large vacuum energy) and that the inflaton particles must have a small mass

$$\ddot{\phi} \ll 3H\dot{\phi} \quad \dot{\phi}^2 \ll V(\phi) \tag{17.11}$$

Moreover, we use the reduced Planck mass, which substitutes several constants:

Definition 17.4 (Reduced Planck mass) On the scale of the reduced Planck mass, quantum effects of gravity become important.

$$M_{\mathrm{Pl}} = \sqrt{\frac{\hbar c}{8\pi}} = \frac{1}{\sqrt{8\pi}} \quad (\hbar = c = 1) \tag{17.12}$$

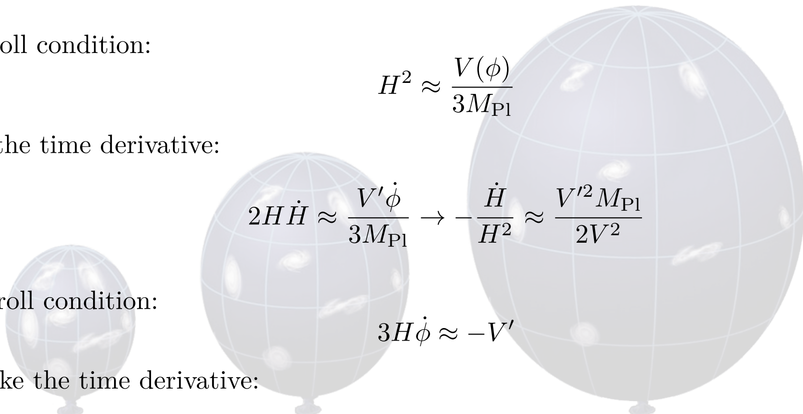
Inflation happens when the scalar field is in a 'slow roll'[2]. Via the slow roll conditions, we can rewrite the Friedmann equations as

Theorem 17.3 (Slow roll Friedmann equations)

$$3H\dot{\phi} \approx -\frac{dV}{d\phi} \quad H^2 \approx \frac{8\pi}{3}V(\phi) = \frac{1}{3M_{\mathrm{Pl}}^2}V(\phi) \tag{17.13}$$

Remark 17.4 The advantage of using a dynamical scalar field over a cosmological constant is that we can exit inflation when the slow roll conditions are no longer fulfilled.

Derivation 17.1 (Slow roll conditions) We want to write the slow roll conditions in terms of the potential only:



- 1^{st} slow roll condition:

$$H^2 \approx \frac{V(\phi)}{3M_{\mathrm{Pl}}} \tag{17.14}$$

  We take the time derivative:

$$2H\dot{H} \approx \frac{V'\dot{\phi}}{3M_{\mathrm{Pl}}} \to -\frac{\dot{H}}{H^2} \approx \frac{V'^2 M_{\mathrm{Pl}}}{2V^2} \tag{17.15}$$

- 2^{nd} slow roll condition:

$$3H\dot{\phi} \approx -V' \tag{17.16}$$

  Again, take the time derivative:

$$3\dot{H}\dot{\phi} + 3H\ddot{\phi} \approx -V'' \cdot \dot{\phi} \to \frac{\ddot{\phi}}{3H\dot{\phi}} \approx \frac{M_{\mathrm{Pl}}^2 V''}{3V} + \frac{1}{6}\frac{V'^2}{V^2} \tag{17.17}$$

[2] The reason behind whose naming will soon become apparent.

In summary:

Definition 17.5 (Slow roll conditions, roll 2)

$$-\frac{\dot{H}}{H^2} \approx \frac{V'^2 M_{\text{Pl}}}{2V^2} \quad \frac{\ddot{\phi}}{3H\dot{\phi}} \approx \frac{M_{\text{Pl}}^2 V''}{3V} + \frac{1}{6}\frac{V'^2}{V^2} \tag{17.18}$$

This allows us to consider inflation. We can write the second derivative of a in terms of the Hubble constant:

$$H = \frac{\dot{a}}{a} \to \ddot{a} = (\dot{H} + H^2) \tag{17.19}$$

During the very rapidly accelerated expansion era, we want that $\ddot{a} \ll 0$. Thus we have

$$\dot{H} + H^2 \ll 0 \to -\frac{\dot{H}}{H^2} \ll 1 \to \frac{V'^2}{2V} M_{\text{Pl}}^2 \ll 1 \tag{17.20}$$

Definition 17.6 (Slow roll parameters)

$$\epsilon = \frac{1}{2} M_{\text{Pl}}^2 \frac{(V')^2}{V^2} \quad \eta = M_{\text{Pl}}^2 \frac{V''}{V} \tag{17.21}$$

Remark 17.5 ϵ measures the steepness of the potential V. When $\epsilon \ll 1$, the potential is flat enough for the inflaton field to roll slowly, leading to prolonged inflation. η measures the curvature of the potential. When $\eta \ll 1$, it ensures that the potential remains flat and that the slow roll condition is maintained over time.

Remark 17.6 The slow roll parameters are *smol* in magnitude during the slow roll phase - a phase where the early universe expands slow enough for inflation to be sustained for a sufficient period, which indicates that the potential energy of the field dominates over its kinetic energy. This leads to a period of rapid but steady expansion of the universe, justifying the names.

17.5 End of inflation

Eventually, however, all good things come to an end. The end of inflation, which was made possible due to our choice of dynamic quantities, is defined to occur when either of the two is satisfied:

Theorem 17.4 (End of inflation conditions)

$$\epsilon \approx 1 \quad \eta \approx 1 \tag{17.22}$$

Theorem 17.5 (E-folding) The number of e-foldings $\mathcal{N}$ is given by

$$\mathcal{N}(t) = \log\left(\frac{a(t_{\text{end}})}{a(t)}\right) \tag{17.23}$$

where t_{end} is the time that inflation ends.

We thus have

$$a(t_{\text{end}}) = e^{\mathcal{N}} a(t) \tag{17.24}$$

Remark 17.7 The scale factor at t_{end} is $\mathcal{N}$ e-foldings larger than at t. Observations tell us that at $\mathcal{N} = 50 \sim 70$ solves all problems we have with the FLRW model[3].

Remark 17.8 A real scalar field in slow roll corresponds to exponential expansion of the universe.

In the end-of-inflation case, we can attack $\mathcal{N}$ such that it is written solely in terms of $V(\phi)$.

Derivation 17.2 (End-of-inflation case) In the end-of-inflation case:

$$H = \frac{\dot{a}}{a} = \frac{d}{dt}(\log a(t)) \tag{17.25}$$

[3]The slow roll parameters have risen!

$$\mathcal{N} = \log \frac{a(t_{\text{end}})}{a(t)} = \int_t^{t_{\text{end}}} H(t)dt \tag{17.26}$$

Now recall the first slow roll Friedmann equations seen in Equation 17.13. One can rewrite this as

$$3H\dot{\phi} \approx -V' \to 3H\frac{d\phi}{dt} \approx -\frac{dV}{d\phi} \to dt = \frac{3Hd\phi}{-V'} \tag{17.27}$$

By inserting this dt into 17.26

$$\mathcal{N}(t) = \frac{1}{M_{\text{Pl}}} \int_{\phi_{\text{end}}}^{\phi} \frac{V}{V'} d\phi \tag{17.28}$$

Derivation 17.3 (Scalar field potential) We introduce a massive scalar field. The potential, which then couples to mass, has the standard form

$$V = \frac{1}{2}m^2\phi^2 \tag{17.29}$$

> **Quote 17.6** Everyone's favourite potential is this one...
>
> *Betti Hartmann, on $V = \frac{1}{2}m^2\phi^2$, 1 March 2024*

In this case, we can derive

$$V' = m^2\phi \quad V'' = m^2 \tag{17.30}$$

As such, the slow roll conditions are

$$\epsilon = \eta = \frac{2M_{\text{Pl}}^2}{\phi^2} \tag{17.31}$$

Inflation happens when

$$\phi^2 \ll 2M_{\text{Pl}} \to \phi \ll \sqrt{2}M_{\text{Pl}} \tag{17.32}$$

Inflation ends when

$$\phi = \sqrt{2}M_{\text{Pl}} \tag{17.33}$$

This can be inserted into Equation 17.28. Setting the time interval as t_{init} and t_{end}, we find

$$\mathcal{N}(t_{\text{init}}) = \frac{1}{M_{\text{Pl}}^2} \int_{\phi_{\text{init}}}^{\phi_{\text{end}}} \frac{V}{V'} d\phi = \frac{1}{M_{\text{Pl}}^2} \int_{\phi_{\text{init}}}^{\phi_{\text{end}}} \frac{1}{2}\phi d\phi = \left(\frac{\phi_{\text{init}}}{2M_{\text{Pl}}}\right)^2 - \frac{1}{2} \tag{17.34}$$

When we have around 70 e-foldings:

$$\phi_{\text{init}} = 2M_{\text{Pl}}\sqrt{\frac{141}{2}} \approx 17M_{\text{Pl}} \tag{17.35}$$

Hence, in order to achieve 70 e-foldings, we need to *choose* the value of ϕ at the beginning of inflation to be approximately $17M_{\text{Pl}}$.

We summarise the main ideas:

- A scalar field that evolves dynamically such that we can enter and exit inflation.
- When the scalar field has reached the minimum, it oscillates.
- The potential energy of the scalar field acts like a cosmological constant (i.e. exponential expansion).
- For $V(\phi) = \frac{1}{2}m^2\phi^2$, we have a natural exit to inflation as scalar field rolls down to the minimum value $\phi = 0$.
- The oscillation and the minimum of the potential are a mechanism for reheating the universe. After inflation, the universe is *very* cold and has a redshift by a factor of $e^{\mathcal{N}}$[4].

[4] Here we see the nominal significance of the e-foldings.

Remark 17.9 (The appeal of inflation) Up to this point, it all looks rather like the fine-tuning we've seen before. So what is the appeal of inflation? We will see that the model of inflation is in good agreement with the fluctuations in the CMB.

Chapter 18

Emergence of cosmological perturbation theory

> **Quote 18.1** Cosmological perturbation theory can be its own module.
>
> *Betti Hartmann, March 2024*

Finally, we will introduce the concept of cosmological perturbations. Two types of perturbations exist. In cosmology, the more commonly studied form are metric perturbations, which bear some similarities to linearised gravity. In physical cosmology, where we are concerned with smaller-scale structures, the more commonly studied form are density perturbations, which essentially declare a region to have a larger or smaller density in comparison to its surroundings.

18.1 Metric perturbations

We first ask a few questions:

- Why do we see structures on a smaller scale (e.g. galaxies, clusters, etc.)?
- The CMB is isotropic to $1 : 10^{-5}$ (i.e. $\frac{\Delta T}{T} = 10^{-5}$), but where do these small deviations come from?

Remark 18.1 We conclude that there are quantum fluctuations in the scalar field. Every quantum field has fluctuations related to the uncertainty principle.
We consider perturbations to the FLRW metric and its 10 components. The perturbations will read:

Definition 18.1 (Metric perturbations)

$$g_{\mu\nu} = \underbrace{{}^{(0)}g_{\mu\nu}}_{\text{background FLRW metric}} + \underbrace{{}^{(1)}g_{\mu\nu}}_{\text{linear } (\delta g_{\mu\nu})} + \underbrace{{}^{(2)}g_{\mu\nu} + \ldots}_{\text{non-linear}} \tag{18.1}$$

To date, all perturbation theories have been linear, as it would be too difficult otherwise. We want the equations that tell us how $\delta g_{\mu\nu}$ will evolve with t. We then ask the following:

- How do we fix the freedom we have in coordinate choice?
- How do we implement the idea of homogeneity and isotropy while 'creating' the ground structure?

To answer these questions, we will encounter again our good friend, the Lie derivative.

18.2 Conformal coordinates

We recall *conformal time* defined in Equation 15.45, and realise that it would be better to use Cartesian coordinates. The FLRW metric $g_{\mu\nu}$ in (Cartesian and) conformal coordinates (η, x, y, z) is

Definition 18.2 (FLRW metric in conformal coordinates)

$$ds^2 = a^2(\eta)[-d\eta^2 + \gamma_{\mu\nu}dX^\mu dX^\nu] \quad \text{where} \quad \gamma_{\mu\nu}dX^\mu dX^\nu = \frac{dx^2 + dy^2 + dz^2}{\left(1 + \frac{k}{4}(z^2 + y^2 + z^2)\right)^2} \tag{18.2}$$

18.3 Scalar, vector and tensor perturbations

We decompose $\delta g_{\mu\nu}$ into scalar, vector and tensor perturbations:

$$\delta g_{\mu\nu} = \begin{pmatrix} \underbrace{\delta g_{00}}_{1\times 1} & \underbrace{\delta g_{0\nu}}_{1\times 3} \\ \underbrace{\delta g_{\mu 0}}_{3\times 1} & \underbrace{\delta g_{\mu\nu}}_{3\times 3} \end{pmatrix} \tag{18.3}$$

Rather unsurprisingly, this is called *scalar-vector-tensor decomposition*. This is the backbone of cosmological perturbation theory.
We then consider the degrees of freedom. A 3-vector is characterised by its absolute value (i.e. length/-modulus) as well as its direction.

- A scalar (i.e. the magnitude) degree of freedom of 1.
- Now that we have fixed the magnitude, considering that $v_x^2 + v_y^2 + v_z^2 = |v|^2$, we find that the components have 2 degrees of freedom.
- As such, a 3-vector has 3 degrees of freedom.

Theorem 18.1 (Helmholtz's theorem) Let v be a twice-differentiable vector field in $\mathbb{R}^3$ that vanishes as fast as $1/r$ for $r \to \infty$. Then v can be decomposed into a curl-free (i.e. rotation-free) and divergent-free component:

$$v = \underbrace{-\nabla\phi}_{\text{curl-free}} + \underbrace{\nabla \times A}_{\text{divergence-free}} \tag{18.4}$$

where ∇ is *not* the covariant derivative that produces the cute little Christoffel symbols.

We then arrive at the following parameterisation:

$$\delta g_{0\mu} = \nabla_\mu b + s_\mu \quad \mu = 1, 2, 3 \tag{18.5}$$

where b is a scalar field and s_μ is a divergence-free vector field.
Evgeny M. Lifshitz[1] showed in 1946 showed that one way to parameterise the 3×3 matrix is the following:

$$\delta g_{\mu\nu} = \nabla_\mu\nabla_\nu e + \frac{1}{2}(\nabla_\mu f_\nu + \nabla_\nu f_\mu) + h_{\mu\nu} \tag{18.6}$$

where e is a scalar, f_ν is a 3-vector and $h_{\mu\nu}$ is a 3×3 symmetric tensor.
We write the scalar perturbations...

Theorem 18.2 (Scalar perturbations)

$$\delta^{(s)} g_{\mu\nu} = a^2(\eta) \begin{pmatrix} -2\phi & \nabla_\nu B \\ \nabla_\mu B & -2(\psi\gamma_{\mu\nu} - \nabla_\mu\nabla_\nu E) \end{pmatrix} \tag{18.7}$$

where ϕ, B, ψ and E are scalar functions of space and time.

...the vector perturbations...

[1] Of the *Landau & Lifshitz* fame.

Theorem 18.3 (Vector perturbations)

$$\delta^{(v)} g_{\mu\nu} = a^2(\eta) \begin{pmatrix} 0 & -S_\nu \\ -S_\mu & \nabla_\nu F_\mu + \nabla_\mu F_\nu \end{pmatrix} \tag{18.8}$$

where S_μ and F_μ are 3-vector functions that are divergence-free.

Remark 18.2 Why are the functions divergence-free? In the universe, we do not have any physical process that creates vector perturbations. What is the physical implication of this? We remember that $\nabla \cdot B = 0$ represents a lack of magnetic monopoles. Likewise, we can represent this by having S_μ be divergence-free.
...and finally, the tensor perturbations.

Theorem 18.4 (Tensor perturbations)

$$\delta^{(t)} g_{\mu\nu} = a^2(\eta) \begin{pmatrix} 0 & 0 \\ 0 & h_{\mu\nu} \end{pmatrix} \tag{18.9}$$

where $h_{\mu\nu}$ is a symmetric 3-tensor function. It is divergence- and trace-free.

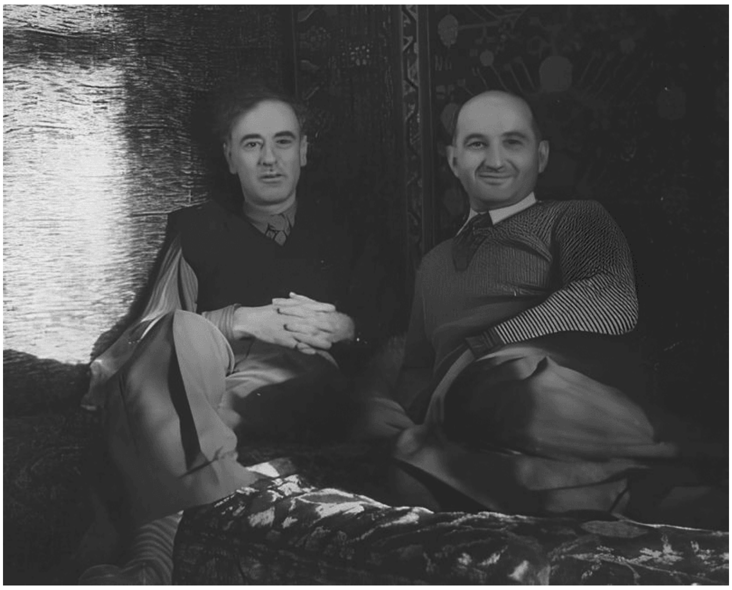


Figure 18.1: ‘What are ya looking at, punk?’

Quote 18.2 Not a word of Landau and not a thought of Lifshitz.

Common saying on Course of Theoretical Physics

We count all the degrees of freedom:

- Scalar: 4
- Vector: $6 - 2 = 4$
- Tensor: $6 - 1 - 3 = 2$

We end up with 10 degrees of freedom. This corresponds to the 10 independent components of the metric.
Remark 18.3 In the scalar-vector-tensor decomposition the dynamical evaluation of the perturbations are such that the equation for scalar, vector and tensor perturbations decouple!

18.4 Fixing the gauge for scalar perturbations

Nonetheless, there is still some freedom to choose coordinate systems. We consider an infinitesimal coordinate transformation (i.e. a *gauge*):

$$\eta \to \tilde{\eta} = \eta + \xi^0(\eta, x^\mu) \quad x^\mu \to \tilde{x}^\mu = x^\mu + \gamma^{\mu\nu}\nabla_\nu \xi(\eta, x^\mu) \quad \mu = 1,2,3 \tag{18.10}$$

We will now choose ξ^0 and ξ 'to our convenience' or, in field theory language, 'fix the gauge'.
Remark 18.4 Choosing the coordinate transformations to give in terms ξ^0 and ξ preserves the *scalar* nature of the perturbation.

Derivation 18.1 (Fixing the gauge) We recall, from *Spinor's Destiny*, the definition of Lie derivatives:

$$(\mathcal{L}g)_{\mu\nu} = g_{\mu\nu}(x) -^{(')} g_{\mu\nu}(x) \tag{18.11}$$

The decomposed metric is given by

$$g_{\mu\nu} = \underbrace{g_{\mu\nu}(x)}_{\text{FLRW metric}} \quad -^{(0)} g_{\mu\nu} + \delta g_{\mu\nu} \tag{18.12}$$

which we insert into the Lie derivative. This yields

$$\mathcal{L}_\xi g_{\mu\nu} = \delta g_{\mu\nu}(x) - \delta^{(')} g_{\mu\nu}(x) \tag{18.13}$$

Using Equation 18.7, it can be found that

$$\delta^{(')} g_{00} = -2a^2\phi + 2aa'\xi^0 + 2a^2(\xi^0) \tag{18.14}$$

From parametrisation, we have $\delta^{(')} g_{00} = 2a^2\tilde{\phi}$. Equating the two yields

$$\tilde{\phi} = \phi - \frac{a'}{a}\xi^0 - (\xi^0)' \tag{18.15}$$

In the same vein, we can derive the scalar functions seen in Equation 18.7:

$$\tilde{\psi} = \psi + \frac{a'}{a}\xi^0 \quad \tilde{E} = E - \xi \quad \tilde{B} = B + \xi^0 - \xi' \tag{18.16}$$

And now, having no more degrees of freedom left, we have fixed the gauge. The metric that includes the scalar perturbations thus reads

Definition 18.3 (Scalar perturbation metric)

$$ds^2 = a^2(\eta)\left[-(1+2\Phi)d\eta^2 + (1-2\Psi)\gamma_{\mu\nu}dx^\mu dx^\nu\right] \tag{18.17}$$

where

$$\Phi = \phi + \frac{1}{a}\left[(B - E')\,a\right]' \quad \Psi = \psi - \frac{a'}{a}(B - E') \tag{18.18}$$

where $\phi(\eta, x, y, z)$ and $\psi(\eta, x, y, z)$ are functions that depend on all coordinates, representing 2 degrees of freedom.

18.5 First-order scalar perturbation theory

We are now in a position to finally perform first-order scalar perturbation theory. Again, for simplicity, we consider scalar perturbations only. Assuming $\Phi \ll 1$ and $\Psi \ll 1$, we perturb the Einstein tensor:

$$G_{\mu\nu} =^{(0)} G_{\mu\nu} + \delta G_{\mu\nu} \quad g_{\mu\nu} =^{(0)} g_{\mu\nu} + \delta g_{\mu\nu} \quad T_{\mu\nu} =^{(0)} T_{\mu\nu} + \delta T_{\mu\nu} \tag{18.19}$$

This gives, for the background space-time equation:

$$^{(0)}G_{\mu\nu} = 8\pi G^{(0)} T_{\mu\nu} \tag{18.20}$$

and for the the first-order perturbation:

$$\delta G_{\mu\nu} = 8\pi G \delta T_{\mu\nu} \tag{18.21}$$

Using the metric, we can shift an index upwards:

$$\begin{aligned}
\delta G^0_0 &= \frac{2}{a^2}\left[3\frac{a'}{a}\left(\Psi' + \frac{a'}{a}\Phi\right) - \Delta\Psi\right] \\
\delta G^0_\mu &= -\frac{2}{a^2}\partial_\mu\left(\frac{a'}{a}\Phi + \Psi'\right) \\
\delta G^\nu_\mu &= \frac{2}{a^2}\left[\left(\Psi'' + \frac{a'}{a}(\Phi' + 2\Psi') + 2\frac{a''}{a}\Phi - \frac{a'^2}{a^2}\Phi\right)\delta^\nu_\mu + \frac{1}{2}\delta^\nu_\mu\Delta(\Phi - \Psi) - \frac{1}{2}\partial_\mu\partial^\nu(\Phi - \Psi)\right]
\end{aligned} \tag{18.22}$$

where we have the 3-dimensional Laplace operator

$$\Delta = \partial_{xx} + \partial_{yy} + \partial_{zz} \tag{18.23}$$

Remark 18.5 We note that although the background space-time and the corresponding Einstein tensor is diagonal, this is not true for the perturbed Einstein tensor. We then consider the energy-momentum tensor:

$$\rho =^{(0)} \rho + \delta\rho \quad p =^{(0)} p + \delta p \tag{18.24}$$

Again, only considering scalar perturbations, the perturbed energy-momentum tensor is

$$\delta T^\nu_\mu = \begin{pmatrix} -\delta\rho & (^{(0)} + \rho^{(0)p})a^{-1}\partial^\nu V \\ (^{(0)}\rho +^{(0)} p)a\partial_\mu V & \delta p \delta^\nu_\mu - \nabla_\mu\nabla^\nu\sigma \end{pmatrix} \quad \mu,\nu = 1,2,3 \tag{18.25}$$

where V and σ are scalar functions.
Perturbations in the inflation field lead to scalar perturbations:

$$\phi(\eta, x, y, z) = \phi_0(\eta) + \delta\phi(\eta, x, y, z) \tag{18.26}$$

Assuming $\sigma = 0$, we find that $\Phi = \Psi$.

Definition 18.4 (Primordial gravitational waves) These waves are created by tensor perturbations of the energy-momentum tensor of the scalar field ϕ.

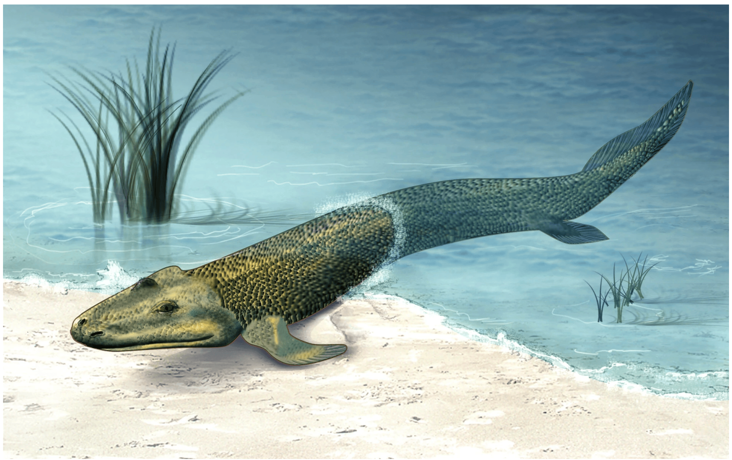


Figure 18.2: The first primordial gravitational wave emerges from water.

The perturbed energy-momentum tensor reads

$$\begin{aligned}
\delta T^0_0 &= \frac{1}{a^2}\left(\phi_0'^2\Phi - \phi_0'(\delta\phi)' - \frac{\partial V}{\partial\phi}a^2\delta\phi\right) \\
\delta T^\mu_0 &= -\frac{1}{2}\phi_0'\partial_\mu(\delta\phi) \\
\delta T^\mu_\nu &= \delta^\nu_\mu\left(-\phi_0'\Phi + \phi_0'(\delta\phi)' - \frac{\partial V}{\partial\phi}a^2\delta\phi\right)
\end{aligned} \tag{18.27}$$

where $\phi_0'^2\Phi$ is the perturbation to the metric and $\delta\phi$ is the scalar perturbation. Combine this to the perturbed Einstein field equation and use $\Phi = \Psi$, we have

$$(\delta\phi)'' + 2\frac{a'}{a}(\delta\phi)' - \Delta(\delta\phi) = 0 \tag{18.28}$$

This is very difficult to solve. Applying the Fourier transform, we can convert the x, y and z coordinates to Fourier coordinates K_x, K_y and K_z, where we have the wavevector $\vec{K} = (R_x, K_y, R_z)$[2].

$$\mathcal{F}[f(\eta, x, y, z)] = \int f(\eta, K_x, K_y, K_z)e^{-i\vec{K}\cdot\vec{r}}d^3x \tag{18.29}$$

This yields the damped harmonical oscillator solution:

$$(\delta\phi)'' + 2\frac{a'}{a}(\delta\phi)' + K^2\delta\phi = 0 \tag{18.30}$$

Remark 18.6 Note that this solution is in Fourier space, not real space!

18.6 Density perturbations

In physical cosmology, cosmological perturbations manifest in the form of very large *density perturbations* and collapsed structures such as stars and galaxies.

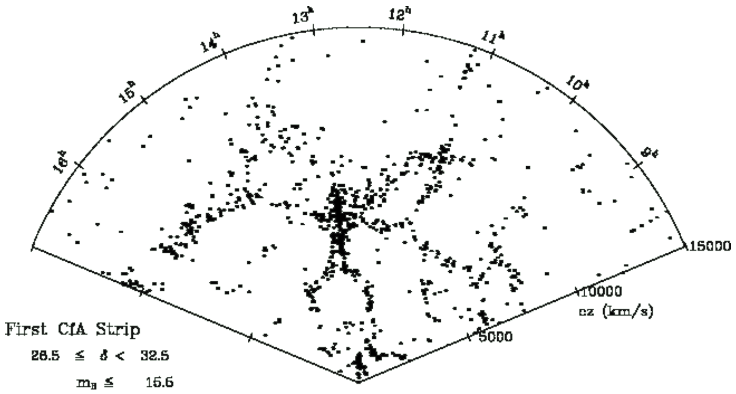


Figure 18.3: Infamous ‘Stick Man’ image from the CfA Redshift Survey.

Definition 18.5 (Overdensity and underdensity)

$$\delta(\vec{X}) = \frac{\rho(\vec{X}) - \bar{\rho}}{\bar{\rho}} \tag{18.31}$$

They are defined with respect to the mean density $\bar{\rho}$:

- If $\rho(\vec{X}) > \bar{\rho}$, then $\delta(\vec{X}) > 0$ and the density is called overdensity.
- If $\rho(\vec{X}) < \bar{\rho}$, then $\delta(\vec{X}) < 0$ and the density is called underdensity.

Remark 18.7 In cosmology, one is more often interested in the overdensities, because these are the ones that will evolve into the ‘most interesting’ structures, i.e. galaxies, groups, filaments, sheets, and clusters. The underdensities will evolve into voids, which are of course interesting too.
Assuming that such perturbations are *adiabatic*, we can equate pressure and density with[3] the *speed of sound*:

[2]This makes sense when you consider the definition of the wavevector.
[3]Not the equation of state!

Definition 18.6 (Speed of sound)

$$c_s = \sqrt{\frac{\partial p}{\partial \rho}} \tag{18.32}$$

With (over)density perturbations taken into consideration, we introduce a modified Friedmann equation or the so-called *growth equation*:

Theorem 18.5 (Growth equation)

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} = \delta\left(4\pi G\bar{\rho} - \frac{\vec{K}^2 c_s^2}{a^2}\right) \tag{18.33}$$

As seen previously, from a wavevector-based ($\vec{K}$) solution, we can use a Fourier transform and derive a position-based ($\vec{X}$) solution:

Definition 18.7 (Generalised density purturbation)

$$\delta(t, \vec{X}) \propto \int e^{-i\vec{K}\cdot\vec{X}}\delta(t, \vec{K})d^3\vec{K} \tag{18.34}$$

Exercise 18.1 Consider the various Friedmann equation solutions we discussed in this book.

- What does the growth equation look like in each of the solutions?
- How does the density perturbation evolve with time?

Derivation 18.2 (Perfect fluids: Mészáros effect) The $-\vec{K}^2 c_s^2/a^2$ term can be ignored in the case of dark matter being perfect fluids (often called *dust* or *cold matter*. We consider multi-component perturbations where we have

$$\delta_i = \frac{\rho_i - \bar{\rho}_i}{\bar{\rho}_i} \tag{18.35}$$

Considering the cold dark matter case:

$$\delta_m'' + 2\frac{\dot{a}}{a}\delta_m' - 4\pi G(\bar{\rho}_m\delta_m + \bar{\rho}_r\delta_r) = 0 \tag{18.36}$$

where ρ_r is the density of the relativistic fluid and δ_r is the associated overdensity. Considering the era of matter-radiation equality:

Theorem 18.6 (Mészáros effect)

$$\delta'' + \frac{2+3y}{2y(1+y)}\delta' - \frac{3}{2y(1+y)}\delta = 0 \tag{18.37}$$

where $y = \bar{\rho}_m/\bar{\rho}_r$.

Remark 18.8 During radiation domination, a dark matter perturbation does not grow. After matter-radiation equality, the perturbation will grow as the Einstein-de Sitter solution ($\delta \propto a$).

Derivation 18.3 (Relativistic case: big perturbations) We return to relativistic cosmology and consider the case where perturbations are larger than the particle horizon. Assume a flat universe with two causal patches with:

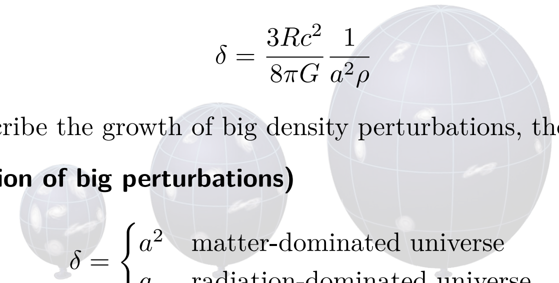
- The first patch is typical of the whole universe.
- The second patch has an overdensity parameter.

The Friedmann equations for the two patches become

$$H^2 = \frac{8\pi G}{3}\bar{\rho} \quad H^2 = \frac{8\pi G}{3}\rho - \frac{R}{a^2} \tag{18.38}$$

where R is the curvature. Combining the equations yield

$$\delta = \frac{3Rc^2}{8\pi G}\frac{1}{a^2\rho} \tag{18.39}$$



The solutions, which describe the growth of big density perturbations, then reads

Theorem 18.7 (Evolution of big perturbations)

$$\delta = \begin{cases} a^2 & \text{matter-dominated universe} \\ a & \text{radiation-dominated universe} \end{cases} \tag{18.40}$$

Likewise, for perturbations smaller than the particle horizon, we can derive

Theorem 18.8 (Evolution of *smol* perturbations)

$$\delta = \begin{cases} a^0 & \text{matter-dominated universe} \\ a & \text{radiation-dominated universe} \end{cases} \tag{18.41}$$

Exercise 18.2 Prove the *smol* perturbation evolutions above.

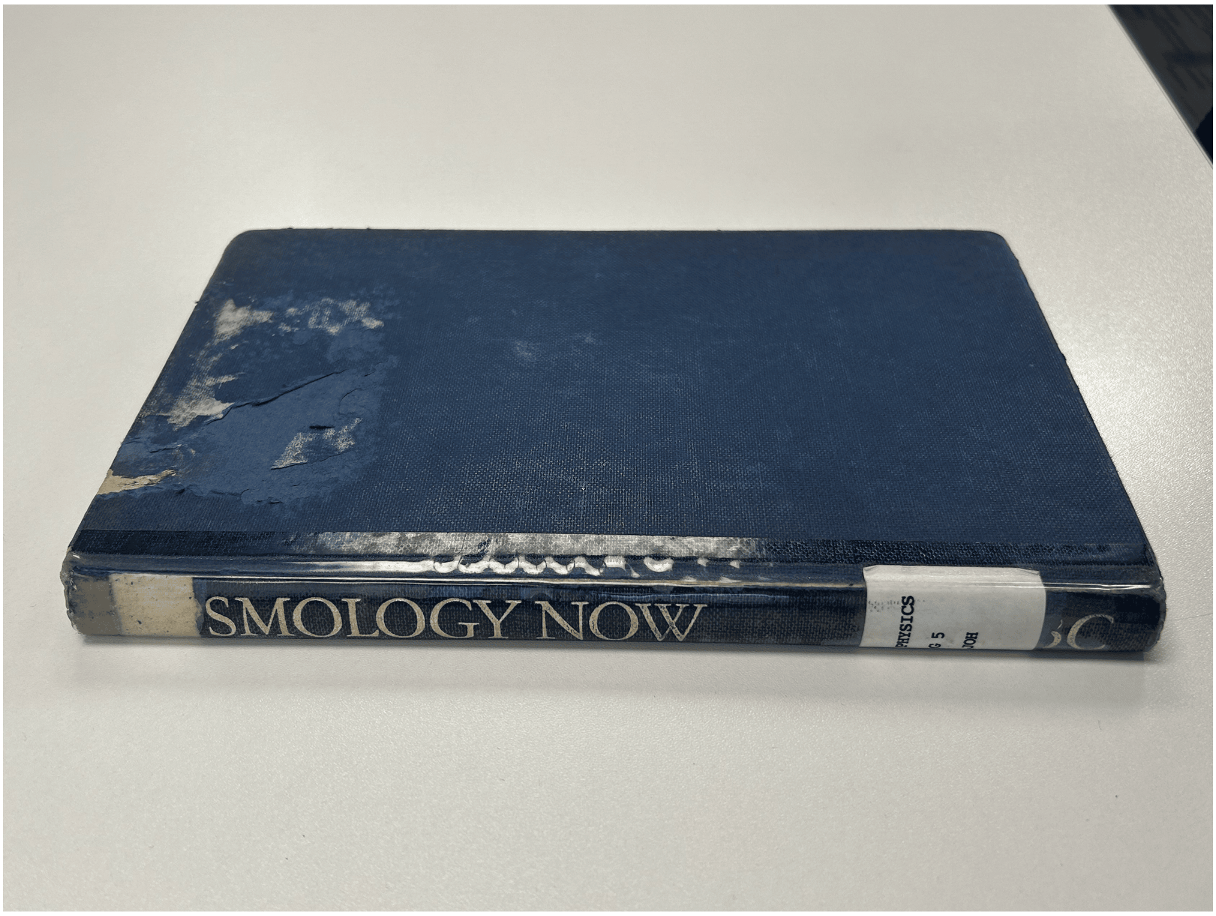


Figure 18.4: ‘Smology Now’