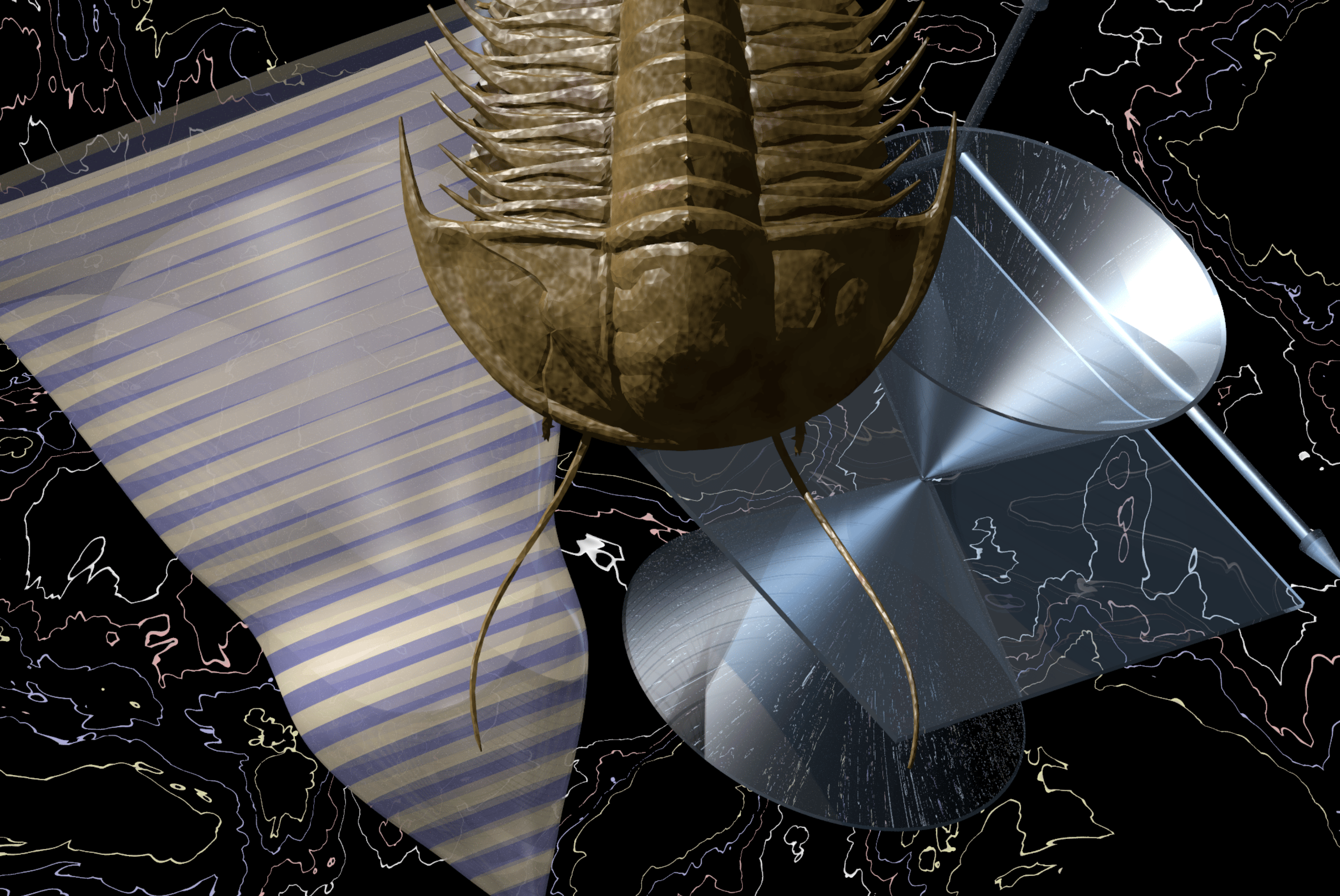
Metrics and Tetrads and Trilobites $g_{\mu\nu}$

Introduction to GR & Cosmology



A Very Short Proto-book by

N. Booker

To my parents

Contents

Chapter 1

Instead of a foreword

1.1 How to use the book

To the layman, the physical meaning of general relativity is not immediately obvious. Many relativists use the term interchangeably with gravitation, and one can make some arguments for why the two are synonymous. However, strictly speaking, GR forms the theoretical framework relativists use to study gravitation.
GR makes two innovations compared to Galilean relativity in classical mechanics. The first is by virtue of its predecessor, special relativity, which states that both space and time lie within a 4D spacetime. This results from the fact that the speed of light is identical in all reference frames and kills the Galilean ideas of absolute time and a 'superior' reference frame that all other frames should be thought of as being in relative motion with. The second innovation is the introduction of spacetime curvature caused by mass. It is this second innovation that is of great interest to relativists, as it produces a fully self-contained and seemingly[1] fundamental theory of gravity. To date, GR remains the best-accepted theory of gravitation, and both SR and CM remain widely used as its effective theories.
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. The main reason for this increase in mathematical sophistication lies in the nature of GR as an expansion of the previously known SR into curved space. Indeed, the study SR without much differential geometry is only possible because the curvature is zero, and the effects associated with it vanish.
This book is structured in four parts:

- METRICS covers the fundamentals of GR. The first half is dedicated to differential geometry, classical mechanics and special relativity. Curvature and the field equations are then introduced, from which we make an introduction to gravitational wave physics and black hole solutions. It resembles a standard 'General Relativity I' course in most universities, but with a few useful additions. A reader with a degree in mathematics should have no problem reading this part like a novel.

- TETRADS covers more advanced topics, but with a corresponding payoff. We will introduce the tetrad formalism and the resulting 3+1 decomposition of spacetime, which forms the basis of numerical relativity, a method of approximation used by astrophysicists as a practical aspect of general relativity. We will also discuss the ADM formalism, which lays the groundwork for a future quantised theory of gravity.

- TRILOBITES covers the basics of cosmology with some applications and resembles a standard 'Cosmology' course in most universities, with the addition of an introduction to physical cosmology.

This book assumes very few mathematical prerequisites from the reader, and many sections will be redundant for more mathematically proficient readers. For this reason, this book is *not* recommended for students doing a 'mathematics and physics' degree or physics students from countries where physics degrees conventionally contain a significant portion of mathematics, like some countries in Central Europe. Instead, such readers are recommended to use more sophisticated notes, like *General Relativity: A*

[1] Today, GR is well-understood as an effective theory, but we will not discuss this for quite some time.

Mathematical Introduction by Christian Bär (Universität Potsdam).
For any comments, suggestions or typos, please e-mail the following address:

`neil(dot)booker(dot)21(at)ucl(dot)ac(dot)uk`

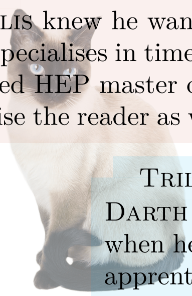
> **Quote 1.1** 🐈, aber sicher doch. Ich mache auch Fehler.
>
> *Felix Halbwedl, 8 February 2025*

1.2 Acknowledgements

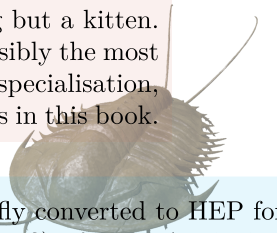
> **Quote 1.2** Credit for those who work hard, I just happen to know some things.
>
> *Felix Halbwedl, in his infinite humility, 22 December 2024*

Work on *Metrics and Tetrads and Trilobites: Introduction to General Relativity and Cosmology* started during the 2024-25 winter and summer semesters at University College London. The initial work was based on the lecture notes on General Relativity and Cosmology authored by Prof. Christian G. Böhmer and the Cosmology lecture notes by Prof. Betti Hartmann respectively. As the range of topics expanded, the book eventually acquired a range of references.
This book would not have been possible without the guidance of Prof. Christian G. Böhmer, whose brilliant leadership of GR and cosmology from 2007 to 2024 and 2021 respectively instilled in me a true love and a strong foundation in gravitation. I am especially grateful to him for the many stimulating mathematical and physical discussions we made throughout the past years, for taking me to various conferences and introducing me to the world of relativists, as well as being a true friend.
I would also like to thank Prof. Betti Hartmann, who took over GR and cosmology from Prof Böhmer, supervised my 3rd-year research internship and taught my cosmology course and Prof. Mitchell A. Berger, who taught GR and cosmology at UCL before leaving for Exeter in 2007 and replied to me when I e-mailed him about his 2004 lecture notes 20 years after their completion. Without them, the GR and cosmology courses at UCL would not have been nearly as interesting. I also thank Eissa Alnasrallah and Erik Jensko, my colleagues at the UCL Mathematical Physics Group.
I would like to thank Felix Halbwedl, Alex Lukov, Francisco Silva, Paulina Schlachter and Abhijeet Vats for physical discussions and advice in improving this book, Kirill Batrakov, Max Henderson, Alex Lukov, Paulina Schlachter and Francisco Silva for proofreading drafts of this book. I am also grateful 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.
I would like to thank Felix Halbwedl, Paul Kothgasser, Robert Schwarzl and other members of the Basisgruppe NAWI Physik (BaGru) at the Technische Universität Graz for 'adopting' me into their student community during the latter half of the writing of this book.
Lastly and most importantly, I dedicate this book to my parents, whose immense love and support throughout my life I would never be able to repay.

1.3 Cast of characters



Cat 1.1
Darth Felis knew he wanted to study HEP since he was nothing but a kitten. Today, he specialises in timelike quark propagators, and is quite possibly the most accomplished HEP master of our generation. Despite his different specialisation, he will advise the reader as well as the author on important concepts in this book.



Trilobite 1.2
Darth Arthropodus is a relativist, but briefly converted to HEP for one year when he was taken under the wing (or rather paw?) of Felis. As the latter's HEP apprentice, he studied QFTI, QFTII and the standard model before suffering a 'crisis of faith'. As we will see in this book, he is rather dim-witted.

References

- *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)
- *Cosmology* by Betti Hartmann (University College London)
- *Differential Geometry* by Mahir Hadžić (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)
- *Manifolds* by Georgios Papadopoulos (King's College London)
- *The General Theory of Relativity: A Mathematical Approach* by Farook Rahaman (Jadavpur University)

Part I

Metrics

Chapter 2

Manifolds

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

If one is to take it seriously, the study of manifolds deserves its own course. Being clumsy physicists, we do not do this here. Nonetheless, we acknowledge the fact that without manifolds and the many objects defined on them, it would be impossible to do GR beyond the bachelor level. In particular, we will emphasise how the abstract maps that define our objects ultimately reflect the objects' properties in physical calculations.

2.1 Minimal introduction to topology

While GR falls into the category of mathematical physics, it ultimately describes physical phenomena. This has two implications. While one can go on reasonably well in the bachelor's stage without looking into the mathematics behind GR, they will inevitably struggle if they still refuse to do so after having decided to specialise in it.

As you are reading this book, we assume that you are a future relativist aspirant who is nonetheless too timid to study mathematical relativity, which makes things rather simple. Having identified why we are here, the objective is hence to cover all the preliminaries within one section, so that you will not make mistakes that are too obvious later in your career.

Right now, we are still a long way from GR, or indeed, even SR, for the simple reason that many assumptions we have taken for granted ever since we were in school must now be formulated rigorously. This process is twofold:

- Introducing topology allows us to define a continuous map[1].
- Introducing differential geometry allows us to define smooth manifolds, which is essential for differentiable maps.

This section involves itself with both objectives. We start with the simplest, most harmless geometry possible, that being our innocent *Euclidean space* or *real coordinate n-space*[2] $\mathbb{R}^n$.

Definition 2.1 (n-tuple) Generally speaking, a *tuple* is an ordered list of elements, which are mathematical objects. Specifically, an n-tuple is a tuple of n non-negative integers (1, 2, 3, etc.).

Definition 2.2 (Real coordinate space) The so-called real coordinate space is a set of all n-tuples of real numbers:

$$\mathbb{R}^n = (x_1, x_2, \cdots, x_n)|x_i \in \mathbb{R} \tag{2.1}$$

It naturally comes with the following properties:

[1] A map can be viewed as a fancy word for a function.

[2] Strictly speaking, not every Euclidean space is a real coordinate space, but this distinction is usually ignored. In this book, we treat the terms as interchangable.

- Vector addition and scalar multiplication, such that it spans all the real numbers.
- Inner product or *dot product*:
$$\langle x, y\rangle = \sum_{i=1}^{n} x_i y_i \tag{2.2}$$
- Euclidean norm:
$$|x| = \sqrt{\langle x, x\rangle} \tag{2.3}$$

Derivation 2.1 (Open set) Let us now consider the concept of an *open set*. The simplest example would be the *Euclidean line* or the set of real numbers $\mathbb{R}$ (effectively $\mathbb{R}^1$). This is nothing but an *open interval*.

Definition 2.3 (Open interval) An open interval is the set of all points $x \in \mathbb{R}$ such that
$$I = (a, b) = \{x \in \mathbb{R} | a < x < b\} \quad \text{for} \quad a \in \mathbb{R} \quad \text{and} \quad b \in \mathbb{R} \tag{2.4}$$

This definition can be expanded to an Euclidean space of arbitrary dimension $\mathbb{R}^n$, where open intervals evolve into *open balls*:

Definition 2.4 (Open ball) An *open ball* is the set of all points x in $\mathbb{R}^n$ such that:
$$B_r(y) = \{x \in \mathbb{R}^n || x - y| < r\} \quad \text{for} \quad y \in \mathbb{R}^n \quad \text{and} \quad r \in \mathbb{R} \tag{2.5}$$

Remark 2.1 More intuitively, an open ball is the interior of an n-sphere of radius r centred at y. From here, we can define the notion of a *neighbourhood*

Definition 2.5 (Neighbourhood) A set $U \subseteq \mathbb{R}^n$ is known as a neighbourhood of $x \in \mathbb{R}^n$ if there exists an $\epsilon > 0$ such that the open ball $B_\epsilon(y)$ is a subset of U:
$$B_\epsilon(y) \subset U \quad \text{for} \quad \epsilon > 0 \tag{2.6}$$

Remark 2.2 More intuitively, the neighbourhood of an n-dimensional open ball surrounds it by ϵ.

Definition 2.6 (Open set) An *open set* V is then a set that is a neighbourhood of every point in it. Several equivalent, but more intuitive statements exist:

- $V \in \mathbb{R}^n$ is an open set if every point in V can be 'thickened out'[a] to an open ball within V.
- An open set can always be constructed from an arbitrary (maybe infinite) union of open balls.
- A set $V \subset \mathbb{R}^n$ is an open set if, for any $y \in V$, there is a open ball centred at y which is completely inside V.

The complement of an open set is then *closed*.

[a] Note that the criteria is $|x - y| < r$ instead of $|x - y| \leq r$, so any point can be technically thicked out without reaching the 'edge' defined by r.

Definition 2.7 (Topological space) The *topology* τ on a set X consists of a family of open sets O that satisfy:

- Both the empty set $\emptyset$ and the entire set (or rather space) X are open:
$$\{\emptyset, X\} \in O \tag{2.7}$$
- The intersection $U \cap V$ of two open sets U and V are also open:
$$U \in O \quad \text{and} \quad V \in O \to U \cap V \in O \tag{2.8}$$

- The union $\bigcup_{i\in I} V_i$ of any family of open sets $\{V_i | i \in I\}$ is open:

$$V_i \in O \to \bigcup_{i\in I} V_i \in O \tag{2.9}$$

The space X, often also written with its open sets as (X, O), is known as a *topological space*.

This is the central definition of a space in topology, one which we have taken for granted for so long since school.

Remark 2.3 The elements of a topological space are nothing but our good friends, the so-called *points*, which are elements of the set known as a space.

Definition 2.8 (Connected space) We say that X is *connected* if it is not a union of disjoint (i.e. non-intersecting) and nonempty open sets. That is to say, if the empty set $\emptyset$ and X itself are the only subsets that are both open and closed.

Definition 2.9 (Comparison of topologies) We can compare different topologies O_1 and O_2 for the same space X. We say that:

- O_1 is *finer* or *stronger* than O_2 if O_1 has more open sets than O_2. i.e. if

$$O_1 \supseteq O_2 \tag{2.10}$$

- O_1 is *coarser* or *weaker* than O_2 if O_1 has less open sets than O_2. i.e. if

$$O_1 \subseteq O_2 \tag{2.11}$$

All possible topologies for a space X lie between the two following extremals:

- The *indiscrete topology* or *trivial topology* is the empty set $\emptyset$ and the space X itself.
- The *discrete topology* consists of all the subsets of X. i.e. all possible combinations of the elements of X (power set).

Definition 2.10 (Continuous map) A map $f : X \to Y$ between two topological spaces x and y is known as *continuous* if, for any open subset $V \subset Y$, we have

$$f^{-1}(V) = \{x \in X | f(x) \in V\} \tag{2.12}$$

where $f^{-1}(V)$ is well-known as the *preimage*[a] of V.

[a]Roughly the inverse of a map.

Note 2.1 Conceptually, this is the central point of the section. In GR, the geometry we work with is spacetime manifolds, which are smooth and continuous spaces on which physical quantities vary. To describe how physical quantities change, we make use of continuous functions (or maps) between spaces.

Theorem 2.1 (Preimage of open set) If V is an open set, then the preimage $f^{-1}(V)$ is likewise open.

Definition 2.11 (Induced topology) For a topological space X, an arbitrary set Y and a map $f : Y \to X$, the weakest/coarsest topology (i.e. the *smolest* family of open sets) for which the map f is continuous is known as the *induced topology*, *pullback topology* or *initial topology* on Y by the map f. This induced topology is written as O_f, and can be mathematically written as

$$O_f = \{f^{-1}(U) | U \in O\} \tag{2.13}$$

where U is all the possible open sets within O.

Remark 2.4 Note that while Y is nothing but the mering of all the elements of O_f.

Definition 2.12 (Product topology) The *product topology* intuitively emerges from the induced topology. We start with:

- Two topological spaces (X, O_X) and (Y, O_Y).
- Two maps $\mathrm{pr}_1 : X \times Y \to X$ and $\mathrm{pr}_2 : X \times Y \to Y$. These are known as *natural projection maps* and are defined by

$$\mathrm{pr}_1(x, y) = x \quad \mathrm{pr}_2(x, y) = y \quad \text{for} \quad x \in X \quad y \in Y \tag{2.14}$$

The product topology is then the topology on the set $X \times Y$ induced by pr_1 and pr_2. Recalling the definition of an induced topology, it is the weakest/coarsest possible topology where the projection maps remain continuous.

So far, our idea of a so-called *distance* remains rather primitive, and it is desirable to formalise the concept in an arbitrary number of dimensions. Here we enjoy a first taste of the concept of a *metric*.

Definition 2.13 (Metric space) A so-called *metric space* is a set (again, or rather space) M that is equipped with a map called the *distance function*, the *metric function* or simply the 'metric'

$$d : M \times M \to \mathbb{R}^n \tag{2.15}$$

The distance function satisfies, for points $\{x, y, z\} \in M$:

- The distance must not be negative:
$$d(x, y) \geq 0 \tag{2.16}$$
- The distance is zero if and only if x and y are the same point:
$$d(x, y) = 0 \quad \text{if and only if} \quad x = y \tag{2.17}$$
- The order of arguments is irrelevant:
$$d(x, y) = d(y, x) \tag{2.18}$$
- Triangle inequality:
$$d(x, y) + d(y, z) \leq d(x, z) \tag{2.19}$$
i.e. the length of any side of a triangle is less than the sum of the other two. In the case where the two sides are equal, the three points are on the same line.

Here, there might be some confusion for certain readers who are revisiting GR after finishing a physics degree. The distance function is also called the metric function, so is it related to the metric tensor (also the 'metric') we know and love? But wait, I hear some clicking sounds in the background. It is Arthropodus, our trilobite relativist. Let's see what he has to say.

TRILOBITE 2.1

The distance function and the metric tensor are not the same objects. While the distance function defines the general notion of distances, the metric tensor only generates the line element ds^2, which is the infinitesimal distance squared. In the next chapter, we will see how the metric tensor can be used to 'recover' the metric function.

Well, not very enlightening, but this is what I meant when I said earlier that Arthropodus is rather dim-witted. Perhaps Felis can give some better comments.

CAT 2.2

The distance function tags any pair of spacetime points with a number; the number tagging has to follow some generic rules. The metric tensor Taylor-expands the distance function in the pair of spacetime points (x, y) up to second order and takes the lowest nonvanishing expansion coefficients.

Thank you, Felis. This is why we have more than one character around.

CAT 2.3
Miew.

TRILOBITE 2.4
Krrrk.

Thanks, guys. Now let us proceed to the concept of a *metric topology*.

Definition 2.14 (Metric topology) The set of all open balls $B_a(x) \subset M$ of a metric space M generates (or rather *is*) the metric topology of M.

The final concept in topology before proceeding to manifolds is the so-called *Hausdorff space*. To understand the motivation of Hausdorff spaces, it is necessary to introduce a bit of context. In some topologies, like the trivial topology (empty set $\emptyset$ and the space itself), it is impossible to 'separate' or 'distinguish' distinct points with disjoint open sets (i.e. non-intersecting neighbourhoods), simply because of how few open sets there are. This is problematic:

- If two distinct points always share exactly the same open neighbourhoods, then the topology cannot 'tell them apart' or 'distinguish' them. For this reason, the points are known as *topologically indistinguishable*.
- For example, again in the trivial topology, we only have the empty set $\emptyset$ and the space X itself. Hence, every point has the same neighbourhood, which again is X itself, and all points are topologically indistinguishable.

This is one motivation for the *Hausdorff condition*, which gives the constraints needed for two distinct points to exist. But somehow, this doesn't look right. Let us ask the great Felis for some guidance.

CAT 2.5
A stronger motivation for Hausdorff spaces is that the limit of a sequence, if it exists, is then unique with respect to that topology.

Definition 2.15 (Hausdorff condition) For any two *distinct* points $x, y \in X$ and $x \neq y$, there exist disjoint open sets (open neighbourhoods) U and V which satisfy:

$$x \in U \quad y \in V \quad U \cap V = \emptyset \tag{2.20}$$

Definition 2.16 (Hausdorff space) A topological space satisfying the Hausdorff condition is known as a Hausdorff space.

Remark 2.5 More intuitively, a Hausdorff space is a space where there are no distinct points that are 'arbitrarily close' to each other.

Theorem 2.2 (Hausdorff space properties)

- Every metric space is a Hausdorff space.
- If X and Y are topological spaces, Y is a Hausdorff space and there exists a one-to-one and continuous mapping $f : X \to Y$, then X is also a Hausdorff space.
- Any subspace of a Hausdorff space is also a Hausdorff space.
- From the above, the topological product $X \times Y$ of two Hausdorff spaces X and Y are Hausdorff spaces.

Remark 2.6 Often, when speaking of a Hausdorff space, we simply say that the space is *Hausdorff*.
We end with two definitions that will become useful later: *compact* and *connected spaces*.

Definition 2.17 (Covering) A collection of sets U_i with $i \in I$ is a *covering* of a subset A of a topological space X if for every point $x \in A$, we also have $x \in U_i$. Several related definitions exist:

- A covering is an *open covering* if all U_is are open sets.

- A subset U_i' of a covering that also covers A is a *subcovering* of the original.
- If the index i a subcovering U_i' is finite, U_i' is a *finite subcovering*.

Definition 2.18 (Compact space) A topological space X is compact if every open covering of X contains a finite subcovering.

2.2 Differentialble manifolds

As it turns out, merely having functions that are continuous is insufficient. We are now in a position to formulate differential geometry. The immediate benefit of topology (aside from the conceptually important continuous maps) is that we can work towards the definition of a *smooth manifold*, which is one of the two key takeaways of differential geometry.

Definition 2.19 (Chart) An *chart* is a ordered pair (U, ϕ) that consists of:

- An open subset U of a set M[a].
- A one-to-one map (or a homeomorphism[b]) $\phi : U \to \mathbb{R}^n$ such that the image $\phi(U)$ is open in $\mathbb{R}^n$.

While $\phi(U) \in \mathbb{R}^n$, $\phi(U)$ is *not* a subset of M.

[a]As we will find out later, M is the (spacetime) manifold in the context of this book. U_is are then open regions of the manifold on which a particular coordinate system is valid.

[b]Or, as we will see, diffeomorphism if M is a manifold.

Derivation 2.2 (Coordinate system) Charts are useful in that they allow us to formally define the notion of a *coordinate system*. Suppose that we have some point $p \in U$, ϕ acts on p as follows:

$$\phi(p) = (x^1(p), x^2(p), \ldots, x^n(p)) \in \mathbb{R}^n \tag{2.21}$$

That is to say, the ϕ on the point gives us a series of n numbers. Each of them is then the component of an n-dimensional *coordinate function* or *coordinate representation* x^i associated with the chart where $i = \{1 \cdots n\}$, defined as follows:

$$x^i = \mathrm{pr}_i \circ \phi \tag{2.22}$$

where pr_i projects onto the i-th coordinate. This might still look rather alien to some readers, but if one omits the point p and denotes ϕ as x^i, they will find

$$x^i = (x^1, x^2, \ldots, x^n) \in \mathbb{R}^n \tag{2.23}$$

which is exactly a coordinate system as we would write it as physicists. Here, as we have an arbitrary number of n dimensions, we have eschewed the familiar (t, x, y, z) notation. Instead, we use the superscript to denote which component or 'axis' we are talking about. When $i = 1$, we have $x^i = x^1$, and so on. This is known as the *index notation*, and we will discuss this towards the end of this chapter. The introduction of (local) coordinate systems also allows us to formalise a distinction we will encounter again and again throughout the rest of this book. Mathematical objects, when defined abstractly, often provide little practical information. It is through their incarnation in local coordinate frames that we truly understand their significance:

- The abstract definitions of mathematical objects provide little practical information. This is because an object, when defined abstractly, is almost always a map between one category of objects to another.
- It is only when the object is put into a coordinate frame that the object's components, values, expressions and any other information of physical interest appear.

As physicists, we are in a funny position in that we need to at least understand the abstract definitions to some extent such that our research is not mathematically incorrect, yet for any object, we are almost always interested in its manifestation in local coordinate systems. This distinction will be made again and again throughout this book, especially after the introduction of the index notation.

Here, a problem emerges. Given some local chart (U, ϕ), it is already possible to take the derivative of its coordinate function $x^i = \mathrm{pr}_i \circ \phi$. However, a local chart is merely one of the many coordinate systems of the same geometry that we can assume, and we want our partial derivatives to be (globally) meaningful, irrespective of our coordinate system. The solution is smooth manifolds, which we now work towards.

Definition 2.20 (Smooth function) A function $f : \mathbb{R}^n \to \mathbb{R}^k$ defined in $\mathbb{R}^n$ is said to be C^k if all partial derivatives of order k exist and are continuous. i.e. if

$$D^k f = \frac{\partial^{\sum k} f}{\partial x_1^{k_1} \cdots \partial x_n^{k_n}} \tag{2.24}$$

where:

- The subscripts $1, \cdots, n$ are the components of the coordinate system x.
- The superscripts $k_1, \cdots, k_n$ are the *number of times* each partial derivative $\partial x_1, \cdots, \partial x_n$ is performed.
- Naturally, $\sum k$ is then the total number of partial derivatives we have performed:

$$\sum k = k_1 + \cdots + k_n \tag{2.25}$$

If the function is C^k for all ms up to infinity, it is said to be *smooth*, *infinitely differentiable*, or C^∞.

This ensures that the notion of differentiability does not depend on which chart you use — it's a globally consistent structure.

> **Quote 2.2** Smooth Operator.
>
> *Felix Halbwedl, 2 November 2024*

Definition 2.21 (Atlas) An *atlas* of class C^k (where m depends on our choice) is an indexed collection of charts $\{(U_i, \phi_i)\}$, where $i \in I$ where:

- The union of the U_i is equal to M. That is, the U_i cover M:

$$M = \bigcup_{i \in I} U_i \tag{2.26}$$

- For any two overlapping charts $U_i \cap U_j \neq \emptyset$, the map

$$(\phi_i \circ \phi_j^{-1}) : \phi_i(U_i \cap U_j) \to \phi_\alpha \left(U_\alpha \cap U_\beta\right) \tag{2.27}$$

  must be C^m. That is to say, partial derivatives for ϕ to order m must exist and be continuous. This property has two names:

  - The charts are C^k*-compatible*.
  - The charts are smoothly sewn together.

Definition 2.22 (Smooth atlas) A *smooth atlas* is an atlas of class C^∞.

Definition 2.23 (Maximal atlas) A *differentiable structure* or a *maxima atlas* of class C^k is an atlas that has *all the charts that are* C^k-compatible.

Now, we can finally introduce a differentiable manifold.

Definition 2.24 (Differentiable manifold) An n-dimensional *differentiable manifold* M of class C^k is a topological space that is:

1. Connected.

2. Hausdorff.

3. Equipped with an n-dimensional maximal atlas.

Definition 2.25 (Smooth manifold) A *smooth manifold* is simply a manifold that is of class C^∞. In other words, all charts of the atlas equipped on the manifold must have smooth functions.

The important takeaway from smooth manifolds is that, when a manifold is smooth, all of the functions encoded in its atlas are C^∞ or, in other words, infinitely differentiable. That is to say, no matter what reference frame we choose, taking partial derivatives of coordinate functions will always be meaningful. From a physical point of view, it can then be seen why only a smooth manifold is sufficient to describe the geometries of interest in GR.
Having introduced the smooth manifold, we also need a differential geometry analogue of the continuous functions in topology. This is so that we can deform one space into another continuously, but in a way that allows derivatives to exist everywhere. The solution, which is already hinted at when we looked at smooth functions, is to make our smooth function f defined in Definition 2.27 such that it is a map between manifolds. To do this, both manifolds must also be smooth[3] to satisfy the definition of a smooth map.

Definition 2.26 (Smooth map between manifolds) A function $f : M \to N$ between two smooth manifolds M with maximal atlas $(U_i\phi_i)$ where $i \in I$ and N with maximal atlas $(V_a\psi_a)$ where $a \in A$ is smooth (or C^∞) if the function

$$\psi_a \circ f \circ \phi_i^{-1} : \phi_i(U_i) \subset \mathbb{R}^m \to \mathbb{R}^n \tag{2.28}$$

is smooth (or C^∞).

Remark 2.7 Previously, the motivation for making the smooth map $f : \mathbb{R}^m \to \mathbb{R}^n$ was not immediately obvious. Now we see that this can actually give rise to a (smooth) map between two smooth manifolds. In summary:

$$\begin{array}{ccc} U \subset M & \xrightarrow{f} & V \subset N \\ \downarrow \phi & & \downarrow \psi \\ \phi(U) \subset \mathbb{R}^n & \xrightarrow{\psi \circ f \circ \phi^{-1}} & \psi(V) \subset \mathbb{R}^m \end{array} \tag{2.29}$$

As teased earlier in Definition 2.19, we introduce the notion of *diffeomorphism*.

Definition 2.27 (Smooth map between manifolds) A function $f : M \to N$ is a diffeomorphism if

- f is a bijection (f is one-to-one).
- Both f and f^{-1} are smooth.

Theorem 2.3 Given (smooth) manifolds M, N and P and smooth (C^∞) maps $\phi : M \to N$ and $\psi : N \to P$, then the function $\phi \circ \psi : M \to P$ is also smooth.

2.3 Things that live on manifolds

TRILOBITE 2.6
This section title is borrowed from Mitchell A. Berger's 2004 GR notes.

Interesting fact, Arthropodus. Well, there he goes again, waggling his many legs. One wonders how he knows this in the first place?
Having introduced the geometrical framework we will be working with, it is now possible to introduce (or rather reintroduce with mathematical rigour) the many geometrical objects we know and love. As you have almost certainly learned most of them in your previous studies, this should not be hard to do. So, where to begin? The final goal of this chapter is tensors, so everything we do should move us towards that. In differential geometry, the definition of scalars is not changed, but the concept of vectors is superseded by that of tangent vectors, which are vectors that represent the direction along the tangent

[3]If the manifolds are not smooth, we could talk about continuous maps, but not smooth ones.

of a curve or a surface.
Hence, curves seem like a natural place to begin. Again, we recall that a curve γ is formally defined with respect to its parameterisation, which effectively imposes a 'constraint' on it.

Definition 2.28 (Parametrised curve) From the above, it is intuitive to see that the curve or *parametrised curve* γ is a continuous map from a number to a point within a 1D interval on the real line $I = (a, b) \subseteq \mathbb{R}$ to a point on a manifold M.

$$\gamma : I \to M \tag{2.30}$$

In some texts, the result of the mapping is written as the n-dimensional coordinate space $\mathbb{R}^n$ that the manifold M lives on:

$$\gamma : I \to \mathbb{R}^n \tag{2.31}$$

If γ is smooth in addition to being continuous, then it is known as a *smooth parameterised curve.*

Here, nothing seems to make sense at first glance. The definition seemingly implies that a curve turns an interval in real numbers into a manifold, which is not what immediately comes to mind when one considers curves. Yet not all is hopeless, as long as one learns to make a correspondence between this abstract definition and the 'simplified' definition we physicists know and love:

- As noted before, the abstract definition of a mathematical object is almost always a map between one category of objects to another. In our case, this is a map from a real interval to the manifold.
- However, when we actually utilise the object, we only map a specific element of one category of objects to a specific element of another. In our case, this is a map from a real number to a point on the manifold.
- As it turns out, this happens to be the definition we are familiar with, which says that a curve takes a specific value of the parameter τ and returns a point on the coordinate system.

The takeaway here, which will be invoked in the rest of the chapter and beyond again and again, is then that the definition of an object we use when making physical calculations can be recovered when one considers the map defined by its abstract definition and takes a single element from the two categories the map is concerned with.

Theorem 2.4 (Pass through) For some point $p \in M$ and some arbitrary time $t = t_0$ where $a < t_0 < b$. The curve γ is said to *pass through* p at t_0 if

$$\gamma(t_0) = p \tag{2.32}$$

Notably, a curve does *not* transform as a tensor[4]. To derive its transformational properties and those of its derivatives, we need to investigate something that *does* transform tensorially. This leads us to *tangent vectors.*

Definition 2.29 (Tangent vector) The tangent vector X_p to a curve γ at some point $p \in M$ of a differentiable manifold M is a linear map to real numbers:

$$X_p : C^\infty(M) \to \mathbb{R} \tag{2.33}$$

where $C^\infty(M)$ is a shorthand denoting the set of all smooth scalar fields[a] on the manifold M

$$C^\infty(M) = \{f : M \to \mathbb{R} | f \text{ is smooth}\} \tag{2.34}$$

X_p satisfies, for arbitrary constants a and b and arbitrary functions f and g:

- Linearity in both arguments:
$$X_p(af + bg) = aX_pf + bX_pg \tag{2.35}$$
- Leibniz rule:
$$X_p(fg) = f(p)X_pg + g(p)X_pf \tag{2.36}$$

[a]See Definition 2.32.

[4]We will cover this in the next chapter.

Derivation 2.3 (Tangent vectors in a local coordinate system) The tangent vector is a good opportunity to discuss the two points we have noted previously:

- The abstract definition is coordinate-free, whereas we must take a coordinate system in physical calculations.
- The abstract definition maps one whole category of objects to another, whereas in real calculations, the map is between single elements of these categories.

First, we reduce the two categories in the definition of the tangent vector to one element from each. This is simple, as the elements of $C^\infty(M)$ and $\mathbb{R}$ are nothing but a single function (or scalar field) and a real number. The map itself hence reads

$$X_p(f) := \frac{d}{dt}(f(\gamma(\tau)))\bigg|_{\tau=0} \tag{2.37}$$

which one can prove to be linear in both arguments and satisfy the Leibniz rule.
However, instead of a vector, as one would expect, the output remains a number. This is because the definition so far is still coordinate-free. This is because the actual tangent vector in physical calculations is not the abstract X_p, but rather a X^i defined as

$$X^i := X_p(x^i) \tag{2.38}$$

where $X_p(x^i)$ is the tangent vector map of the *coordinate function* x^i.
If we take an arbitrary coordinate system, the function then reads $f = f(x^1, \ldots, x^n)$. Using the chain rule and inserting (2.38), we then have

$$X_p(f) = X_p(f(x^1, \ldots, x^n)) = \sum_{i=1}^{n} \frac{\partial f}{\partial x^i}(p) X_p(x^i) = \sum_{i=1}^{n} X^i \frac{\partial f}{\partial x^i}(p) = \sum_{i=1}^{n} X^i \left.\frac{\partial}{\partial x^i}\right|_p (f) \tag{2.39}$$

where $i \left.\frac{\partial}{\partial x^i}\right|_p$ is known a *derivation*, a directional derivative acting on smooth functions. Removing f on both sides for clarity, we have

Theorem 2.5 (Tangent vector in a local coordinate system)

$$X_p = \sum_{i=1}^{n} X^i \left.\frac{\partial}{\partial x^i}\right|_p \tag{2.40}$$

We then see that the abstract X_p *sums over*[a] all the individual components, which X^i holds.

[a] This concept is the key point of index notation.

Remark 2.8 A tangent vector is but a single vector. Hence, it is distinct from a vector *field*, which, as we will see, assigns a tangent vector to each point.
Remark 2.9 Physically, a tangent vector represents the change of a function along tiny displacements at a point. Hence, the tangent vector of a constant function always yields zero:

$$X_p(\text{const.}) = 0 \tag{2.41}$$

Definition 2.30 (Tangent space) The *tangent space* T_pM to a manifold M at some point $p \in M$ is the set of tangent vectors at p.

Definition 2.31 (Tangent bundle) The *tangent bundle* TM of a manifold M is the union of all tangent spaces to M:

$$TM = \bigcup_{p\in M} T_pM \tag{2.42}$$

Quote 2.3 You've been using at least two bundles already.

Paulina Schlachter, 7 November 2024

The final loose end to discuss is the so-called *pushforward* of tangent vectors.

Derivation 2.4 (Pushforward) We define:

- Smooth manifolds M and N.
- A smooth map ϕ between manifolds $\phi : M \to N$.
- A point $p \in M$ and a tangent space T_pM to M.

It should be clear now that the map ϕ alone is insufficient, because we now have the extra objects p and T_pM. What are the maps that transform these two objects to their equivalent in N?

- As points are nothing but elements in manifolds, the equivalent of the point p in manifold N can be easily found as $\phi(p)$. So here ϕ suffices.
- The tangent space T_pM to M then becomes $T_{\phi(p)}N$ to N. As tangent spaces are not elements *in* the manifold, we will need some map other than ϕ.

Let us label this map as $\phi_{*,p}$ and call it the pushforward. While we have not defined it yet, we do know that it should take the form

$$\phi_{*,p} : T_pM \to T_{\phi(p)}N \tag{2.43}$$

The important thing now is to realise that the tangent space is a set of tangent vectors. As such, in accordance with Definition 2.29, $\phi_{*,p}$ must be linear and observe the Leibniz rule. As such:

Theorem 2.6 (Pushforward) For some smooth function $g \in C^\infty(N)$, the pushforward $\phi_{*,p}$ is a map abstractly defined as

$$\phi_{*,p}X_{\phi(p)} : C^\infty(\phi(p)) \to \mathbb{R} \tag{2.44}$$

In practice:

$$(\phi_{*,p}X_p)(g) := X_p(g \circ \phi) \tag{2.45}$$

Sometimes, the first bracket is omitted, and we write the LHS as $\phi_{*,p}X_p(g)$.

While the tensor field will elude us for some time, we are now in a position to consider scalar and vector fields. Starting with the *scalar field*, one should again note that the scalars (or rather real numbers) we have worked with are insufficient, as each of them is but a single scalar. From years of learning various field theories, we know at this point that a scalar field should assign a point on the manifold to a (real) number.

Definition 2.32 (Scalar field) A scalar field is a smooth map f from an n-dimensional manifold M to the (1D) set of real numbers $\mathbb{R}$:

$$\phi : M \to \mathbb{R} \tag{2.46}$$

Immediately, we see that it is exactly the opposite of a curve γ. If we compose the two, we get

$$(\phi \circ \gamma)(t) = \phi(\gamma(t)) : \mathbb{R} \to \mathbb{R} \tag{2.47}$$

At first glance, this does not seem very meaningful, as we have seemingly mapped a real number to a real number, but we should remind ourselves that this is nothing but the abstract mathematical definition of composing the two operations. Let us, instead, consider a physical scenario in 3D space (i.e. not 4D spacetime). Suppose that

- The scalar field $\phi := T(x)$ is the temperature of a particle at some position x.
- The curve $\gamma := x(t)$ is the trajectory (i.e. position) of a particle at time t.

Now we compose the two. The end result is $T(x(t))$, which gives us the temperature of a particle at some time t. While both the input t and the output T are scalars, they are two distinctly meaningful physical quantities. Two pedagogical comments are in order here:

- The source and the target of a map, as seen in its abstract definition, represent the abstract 'widest possible category' that the actual sources and targets lie. For example, most non-abstract instances of the scalar field $\phi : M \to \mathbb{R}$ do not literally map an n-dimensional manifold to real numbers, but rather map a single set of n-dimensional coordinates to a single real number.

- Generalising the last comment, we can see that it is always meaningful for physicists[5] to consider physical examples of abstract mathematical objects.

Having (hopefully) demystified scalar fields, we now move on to *vector fields*.

Definition 2.33 (Vector field) A vector field X is the natural development of a scalar field. Like a scalar field, the source of the map is a manifold M. But instead of the real numbers, the target of a vector field is a tangent bundle[a] TM:

$$X : M \to TM \tag{2.48}$$

An additional condition exists. The action of a vector field on a point must live in the point's corresponding tangent space:

$$X(p) \in T_pM \tag{2.49}$$

For any function $f \in C^\infty(M)$, the composition of the vector field and f maps a point to its tangent vector:

$$(X \circ f)(p) = X_p f \tag{2.50}$$

[a] Note that the target is not a single tangent space, as that would only cover a single point.

Theorem 2.7 (Smooth vector field) A vector field is *smooth* if $X \circ f$ is smooth for all fs. i.e. if

$$(X \circ f) \in C^\infty(M) \tag{2.51}$$

Expectedly from Definition 2.29, a smooth vector field is linear and satisfies the Leibniz rule. For arbitrary constants a and b and arbitrary functions f and g:

- Linearity:
$$X(af + bg) = aXf + bXg \tag{2.52}$$

- Leibniz rule:
$$X(fg) = fXg + gXf \tag{2.53}$$

Remark 2.10 Acknowledging our previous comments, we note that specific vector fields do not map whole manifolds to whole tangent bundles (or indeed, even whole tangent spaces). Rather, a physical vector field maps a point p to a tangent vector X_p.
Here there might be some confusion stemming from the name of the tangent vector:

- The tangent vector $X_p : C^\infty(M) \to \mathbb{R}$ maps a scalar field to a real number.
- Yet the vector field $X : M \to TM$ maps a point to a tangent vector.

That is to say, a vector field is a collection of tangent vectors at every possible point, while a tangent vector is the manifestation of a vector field at *one specific point*. This is identical to good ol' linear algebra, where a vector field is a collection of possible vectors.
Naively, we can then say that the concept of a vector in linear algebra has been 'superseded' by that of a tangent vector in differential geometry. This is conceptually correct, but with one distinction:

- In linear algebra, all possible vectors belong to a single vector space V, which is nothing but the flat space $\mathbb{R}^n$.
- In differential geometry, for each point p, we have a *distinct* tangent space T_pM, which itself is a vector space.

This holds a rather disturbing implication. In flat space, all tangent spaces are identical to each other and look like $\mathbb{R}^n$. It is when we introduce *curvature* in the next chapter that the tangent spaces become different from each other.

Derivation 2.5 (Lie bracket) We are now in a position to make a rather interesting discussion. It is intuitive that both addition and multiplication of two scalar fields yield a scalar field, but what about vector fields? Again, it is intuitive that addition holds, so all we have to do is formulate a mechanism for multiplication.

[5] And possibly even mathematicians!

Naively, this would be $X \circ Y$. Let us try this for the expression $(X \circ Y)(fg)$, where we have vector fields X and Y and smooth functions f and g. From inspection, we see that this expression is clearly linear in both arguments, but does it satisfy the Leibniz rule?

$$(X \circ Y)(fg) = X(Y(fg)) = X(gY(f) + fY(g)) = X(gY(f)) + X(fY(g)) \tag{2.54}$$

Using the Leibniz rule on $X(gY(f))$ and $X(fY(g))$ yields

$$(X \circ Y)(fg) = X(gY(f) + fY(g)) = \underbrace{g(X \circ Y)(f) + f(X \circ Y)(g)}_{①} + \underbrace{X(g)Y(f) + X(f)Y(g)}_{②} \tag{2.55}$$

Again from the Leibniz rule, if $X \circ Y$ is indeed a vector field, the RHS should only consist of ①. However, ② generally does not vanish. Hence, $(X \circ Y)$ fails the Leibniz rule and is not a vector field.

Note 2.2 In fact, there is no meaningful analogue for multiplication for vector fields.

That being said, our derivation so far is not entirely worthless, as it actually gives us a hint as to how we can formulate an operation on two vector fields that results in another vector field. Let us first consider swapping the order of X and Y. This gives us

$$(Y \circ X)(fg) = g(Y \circ X)(f) + f(Y \circ X)(g) + X(g)Y(f) + X(f)Y(g) \tag{2.56}$$

Now calculate $(X \circ Y)(fg) - (Y \circ X)(fg)$:

$$(X \circ Y)(fg) - (Y \circ X)(fg) = g(X \circ Y)(f) + f(X \circ Y)(g) - g(Y \circ X)(f) + f(Y \circ X)(g) \tag{2.57}$$

This operation, which satisfies linearity in both arguments and the Leibniz rule, is known as the *commutator* or the *Lie bracket* depending on the context. It is denoted as

Definition 2.34 (Lie bracket) The Lie bracket can be viewed as the measure of the noncommutativity of two vector fields:

$$[X, Y] = (X \circ Y)(fg) - (Y \circ X)(fg) \tag{2.58}$$

The Lie bracket satisfies the following identities:

Theorem 2.8 (Lie bracket properties)

- Antisymmetry:
$$[X, Y] = -[Y, X] \tag{2.59}$$
- Bilinearity:
$$[X, aY + bZ] = a[X, Y] + b[X, Z] \quad [X, fY] = f[X, Y] + XfY \tag{2.60}$$
- Cyclic property (Jacobi identity):
$$[[X, Y], Z] + [[Y, Z], X] + [[Z, X], Y] = 0 \tag{2.61}$$

2.4 Integral curves, flows and Lie derivatives

A specific type of curve is the so-called *integral curve*, which is a curve passing through a point p of interest, and whose tangent vectors form a vector field X of interest.

Definition 2.35 (Integral curve) Consider a vector field X and curve $\gamma(t)$ parameterised by some parameter t, both living in some manifold M. $\gamma(t)$ is said to be an integral curve of X if, at any specific point $p = \gamma(t)|_{t\in\mathbb{R}}$,

$$\frac{d\gamma(t)}{dt} = X_{\gamma(t)} \tag{2.62}$$

for $t \subset \mathbb{R}$. That is to say, the tangent vector of $\gamma(t)$ at any p is exactly the vector $X_{\gamma(t)}$, the *incarnation* of the vector field X at p.

The pushforward of the integral curve is nothing but the tangent vector itself:

$$\gamma_{*,p}\left(\frac{d}{dt}\right)(f) = \frac{d}{dt}(f(\gamma(t))) = T_p(\gamma(t))(f) \tag{2.63}$$

Now we consider the physical aspects. By acknowledging the fact that (2.62) is nothing but a simple set of ODEs, it is not hard for us to work out the physical meaning of these curves. If X are the electric or magnetic field E or B, then $\gamma(t)$ are the field lines. If X is the phase space of a dynamical system, then $\gamma(t)$ are the orbits, and so on.
As we know, a group of field lines then represent the flow of the electric or magnetic field. Likewise, in the theory of dynamical systems, a group of orbits represent the flow of the dynamical system in its area. In fact, the definition of the *flow* stems directly from that of the integral curve:

Definition 2.36 (Flow) The flow, or rather *flow curve* $\sigma(t,p)$, is an integral curve. It is a group of transformations that makes the map

$$\sigma : \mathbb{R} \times M \to M \tag{2.64}$$

which satisfies the additional property

$$\sigma(t+s,p) = \sigma(t,\sigma(s,p)) \tag{2.65}$$

for $\{t, s, t+s\} \in I$.

Two comments are in order:

- The vector field X is then the vector field *induced* by the flow.
- Every vector field X corresponds to a local, one-parameter group of transformations (or rather diffeomorphisms), which it is said to *generate*.

Remark 2.11 Note that as $\mathbb{R}$ is a real number, $\sigma : \mathbb{R} \times M \to M$ is essentially $\sigma : M \to M$, which is a diffeomorphism. Collecting all flow curves, one can say that a vector field generates infinitesimal diffeomorphisms through its flows.

Definition 2.37 (Completeness) A vector field X on a manifold M is *complete* if all of its flow curves exist no matter what value their parameter t takes.

Theorem 2.9 (Compact manifolds and completeness) If M is compact, then all vector fields that live on M are complete.

Finally, we concern ourselves with the derivative on a manifold. Naively, it is intuitive to first try out luck with the partial derivative:

$$\frac{\partial f}{\partial p} = \lim_{\epsilon \to 0} \frac{f(p+\epsilon) - f(p)}{\epsilon} \tag{2.66}$$

This may look legal, but there is one problem.

Trilobite 2.7
But we *still* haven't figured out how we can add two points (e.g. $p+\epsilon$) on a manifold!

Yeah. Sure. Easy for you to say, Arthropodus! Again, let us take the question to a more sophisticated animal.

Cat 2.8
In fact, there is *no* definition of adding two points on a manifold whatsoever. However, we do not necessarily need addition to take us from one point to another. Rather, we can employ the integral flow we introduced in the last section. At any point p, there exists a unique flow generated by some vector field X of interest. It is then exactly this flow which will take us to the nearby point.

This gives rise to the so-called *Lie derivative*, which allows us to take derivatives on a manifold.

Definition 2.38 (Lie derivative) In a manifold M, the Lie derivative $\mathcal{L}_X$ of a function $f \in C^\infty(M)$ along a vector field $X \in M$ is defined as

$$\mathcal{L}_X f(p) = \lim_{\epsilon \to 0} \frac{f(\sigma(\epsilon, p)) - f(p)}{\epsilon} \tag{2.67}$$

by exploiting the flow $\sigma(\epsilon, p)$ generated by X at p.
We can also define the Lie derivative of a vector field Y along a second vector field X.

$$\mathcal{L}_X Y_p = \lim_{\epsilon \to 0} \frac{\sigma_{\epsilon *} Y_p - Y_p}{\epsilon} \tag{2.68}$$

where the dependence on the point p can be done away with altogether.

Theorem 2.10 (Lie derivative properties)

- Reduction to vector field:
$$\mathcal{L}_X f = X(f) \tag{2.69}$$
- Reduction to Lie bracket/commutator:
$$\mathcal{L}_X Y = [X, Y] \tag{2.70}$$
This will be utilised again in the distant Part III.

While the concept of a Lie derivative naturally emerges from the meaninglessness of adding two points on the manifold, our discussions above has unfortunately given rise to a *smol* lie by omission. The Lie derivative does not replace the partial derivative, and the partial derivative does not die in differential geometry either. To clarify this distinction, we should note what we have done has essentially been a detour, as we have actually not recovered the well-definedness of the partial derivative.
The addition of two points remains a meaningless concept in differential geometry, and we have to look for an alternative that accomplishes the same. Our salvation, as it turns out, lies in tangent vectors.

Definition 2.39 (Partial derivative) For a point p in a smooth manifold M, a smooth function $f : M \to \mathbb{R}$) and tangent vector $v \in T_pM$, we can choose choose any smooth curve

$$\gamma : (-\delta, \delta) \to M, \quad \gamma(0) = p, \quad \gamma'(0) = v \tag{2.71}$$

Then the derivative of f at p in the direction v is defined as

$$df_p(v) = v(f) = \left.\frac{d}{dt}\right|_{t=0} f(\gamma(t)) \tag{2.72}$$

where t is the parameter of the curve. This is well-defined because any two curves with the same initial tangent vector give the same derivative.

This seems a bit awkward, chiefly because one does not see the partial symbol ∂ we know and love. To recover it, we will need the index notation which we will introduce in Definition 2.45. To avoid splitting this discussion into two parts, we assume that you are back from the future, with the toolkit of indices in hand.

Derivation 2.6 (Partial derivative in a coordinate system) In a coordinate system $x : U \subset M \to \mathbb{R}^n$, the function f and the vector v can be written as

$$f = \tilde{f} \circ x \quad v = v^i \left.\frac{\partial}{\partial x^i}\right|_p \tag{2.73}$$

(2.72) then reads

$$df_p(v) = v^i \frac{\partial \tilde{f}}{\partial x^i}(x(p)) \tag{2.74}$$

where $\frac{\partial \tilde{f}}{\partial x^i}(x(p))$ is the ordinary Euclidean partial derivative.

We are then in a position to compare the new Lie derivative and the good ol' partial derivatives we know and love:

- Partial derivatives measure the infinitesimal rate of change of a scalar field in a direction at a point. This means that only the internal variation of the field is accounted for.

- Lie derivatives measure the infinitesimal rate of change of a field *as it is dragged along the flow of a full vector field*. This means that in addition of the internal variation of the field, we also account for its *geometric deformation* as it is dragged along the frame.

2.5 Cotangent vectors

There remains one final piece to the puzzle before we can truly define what a tensor is, and that is the *cotangent space* that corresponds to every tangent space. Before we begin, it makes sense to find an analogy in what we know already.

Derivation 2.7 (Covectors) From our previous studies, we know that for every vector space V, there exists a corresponding, so-called *dual (vector) space* or *covector space* V^*. Crudely speaking, a vector space is a collection of column vectors, while a dual space is a collection of row vectors or *covectors*. It is well-known that a vector is defined as nothing but a collection of coordinates (i.e. scalars) that undergo vector addition and scalar multiplication.
Yet the same description can be said for covectors, so our task is then to define covectors in a way that their nature as a 'transposed' version of vectors is reflected.

Definition 2.40 (Covector) A covector is a map from a vector space to real numbers:

$$f : V \to \mathbb{R} \tag{2.75}$$

Thus concludes the abstract definition. In practice, this means that when we apply a specific covector to a vector, we get a scalar. Now where have we seen this before?
As it turns out, this is nothing but our good friend, the dot product. As the dot product between a row vector and a column vector yields a scalar, it is only natural that we distinguish vectors and covectors by defining the latter in relation to it[a].
The so-called cotangent space therefore naturally emerges as a replacement of the dual space in differential geometry. Just like how the dual space is the 'transposed' counterpart of the vector space, so is the cotangent space the 'transposed' counterpart of the tangent space.

[a] A much less common convention is the reverse, where we define covectors as vectors and define vectors as related to covectors via the dot product.

Exercise 2.1 Show that the dual space of a dual space is just the original space itself.

Hence, we can define the cotangent space by analogy:

Definition 2.41 (Cotangent space) The so-called cotangent space T_p^*M of a manifold M at some point $p \in M$ is nothing but the dual vector space to the tangent space T_pM.
Again, we note that in physics, we are more concerned with specific elements of this abstract cotangent space. These elements are cotangent vectors, which correspond to the tangent vectors we mentioned previously. Each cotangent vector is a linear map $\omega_p \in T_p^*M$ from its corresponding tangent space to real numbers:

$$\omega_p : T_pM \to \mathbb{R} \tag{2.76}$$

In reality, this map merely converts a single tangent vector to one real number.
By convention, the action of ω_p on some vector $X_p \in T_pM$ is denoted by the map

$$\omega_p(X_p) = \langle \omega_p, X_p \rangle \tag{2.77}$$

where the operation $\langle , \rangle$ is linear in both arguments.

Definition 2.42 (Cotangent bundle) Analogous to the tangent bundle, the *cotangent bundle* T^*M of a manifold M is the union of all cotangent spaces to M:

$$T^*M = \bigcup_{p \in M} T_p^*M \tag{2.78}$$

Just like how we developed the concept of a vector field to assign a vector to each point on the manifold, we also need to define a *covector field* to assign a cotangent vector to each point on the manifold.

Definition 2.43 (Covector field) The so-called covector field, rather confusingly also labelled by $\omega_p \in T_p^*M$, is a map

$$\omega : M \to T_p^*M \tag{2.79}$$

In reality, each point is assigned a cotangent vector.
We also define an operation $\omega(X)(p) = \omega_p(X_p)$ such that:

- It is $C^\infty M$ for all smooth vector fields X.
- It satisfies

$$\omega(X) : M \to \mathbb{R} \tag{2.80}$$

2.6 Index notation

Now is a good time to clarify the physical meaning of the vector and covector spaces. Indeed, it is nice and easy to say that vectors and covectors are 'column' and 'row' vectors, but these are nothing but ways we write quantities on a computer screen[6]. A better explaination can be found in 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' lower index V_ν. The positioning of the index is not arbitrary: An upper index denotes that the vector is *covariant*, and a lower index denotes that the vector is *contravariant*.

Definition 2.44 (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. This corresponds to a dual vector/row vector.
- When projected *along* the axes, they are said to be contravariant. This corresponds to a (column) vector.

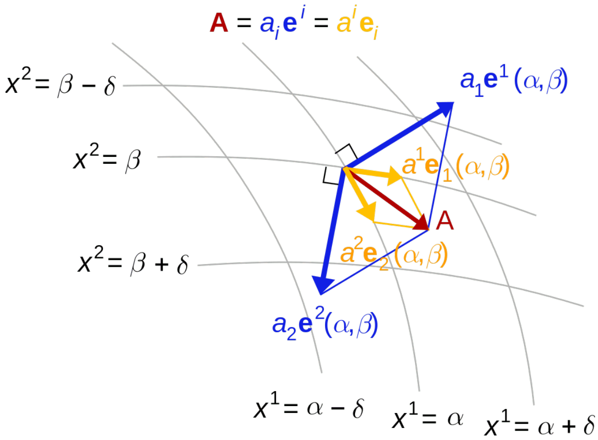


Figure 2.1: Covariant and contravariant components of the same vector A.

[6] Or rather, if you are still living in the 20th century, a piece of paper.

Remark 2.12 In flat space, the covariant and contravariant projections yield identical results, which is why you have most likely not come across this distinction beforehand. In fact, this will remain a non-concern for us until two chapters later.
Remark 2.13 Usually, convention dictates that the index is in Latin letters if we are working with 3D space, and Greek letters if we are working with 4D spacetime[7].

Definition 2.45 (Index notation) In the index notation, each upper and lower index symbol represents a collection of n components of our quantity in the horizontal and vertical direction respectively. Consider the a manifold of arbitrary dimensions with indices $i = 1, \cdots, m$ and $j = 1, \cdots, n$:

- Given an arbitrary number k, the k^{th} components of arbitrary row and column vectors V^i and V_j are V^k and V_k respectively.
- A *matrix*[a] A^{ij} has all elements projected covariantly. A_{ij} has all elements projected contravariantly. A^i_j has mixed projection. All three matrices have $m \times n$ components.
- Higher-dimensional arrays can also be so-represented. Consider a $r+s$-dimensional array with covariant indices a, b, c, etc. going up to d, e, f, etc. dimensions and contravariant i, j, k, etc. going up to l, m, n, etc. dimensions. We then have $d \times e \times f \times l \times m \times n \times \cdots$ components.

[a] Note that we have not yet introduced tensors, and are still working with a toy model of an arbitrary number of dimensions.

Remark 2.14 In most physical scenarios, each index has the same number of dimensions. This is true in the case of GR: every index has 4 dimensions to represent 4-spacetime. However, in abstract mathematics, this is not guaranteed.
We can now discuss some more shorthands of...

Cat 2.9
If you could find a package which automatically colours your indices, that would be great.

Trilobite 2.10
I am sceptical. This is not GR convention. What is the motivation for this?

Cat 2.11
Pedagogical reasons, so 'GR newbies' without the routine could grasp the index structure on an earlier stage. Of course, the colouring becomes unnecessary (and even cumbersome) if you are advanced enough.

Okay, calm down, you two. This is a GR class, not a zoo, you know. Let us compromise. I will implement this for the first chapter, but the first chapter only.

Cat 2.12
Purr!

Trilobite 2.13
Krrrk!

Note 2.3 (Additional conventions) One also has 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.
- When there is only one coordinate system, partial derivatives are often written in a shorthand that omits the denominator or the symbol ∂ altogether:

$$\frac{\partial f}{\partial X^\mu} = \partial_\mu f = f_{,\mu} \tag{2.81}$$

[7] This convention breaks down slightly when we discuss tetrads in Part II.

> **Trilobite 2.14**
> This is all rather interesting, but shouldn't we enlighten the readers as to how the objects in index notation, say, like a vector V^μ, are related to the abstract objects like the abstract vectors $V(p)$ we have previously defined?

Ah, thank you, my good arthropod friend. I almost forgot about that. Oh, and now he goes again. As always, all the writing is left to me...

Derivation 2.8 (Vectors in a local coordinate system) This is analogous to Derivation 2.3. Consider a manifold M with the coordinate system (U_i, ϕ_i), where $U_i \in M$ are open regions within the manifold and $\phi_i(p) = (x^1(p), \cdots, x^n(p))$ are the *actual 'axes'* that assigns coordinates to every point in the i^{th} region. This coordinate system forms a basis $\frac{\partial}{\partial x^\mu}\big|_p$ of T_pM, where $p \in U_i$.
It is then intuitive to create a coordinate system for the cotangent space too. As it turns out, this is not hard, and only requires the cotangent space operation ω_p. This alternate basis, denoted by the Greek index ν instead, is naturally denoted by $\frac{\partial}{\partial x^\nu}\big|_p$. Importantly, from the definition of a cotangent space, we know that the following is satisfied:

$$\left\langle dx^{\mu}|_p, \left.\frac{\partial}{\partial x^{\nu}}\right|_p \right\rangle = \delta^{\mu}_{\nu} \tag{2.82}$$

where δ^μ_ν is the Kronecker delta that we will introduce in Definition 3.7, and $dx^\mu|_p$ is a linear map from the tangent space to real numbers:

$$dx^{\mu}|_p : T_pM \to \mathbb{R} \tag{2.83}$$

Down to specific elements of the mapping, we have

$$dx^{\mu}|_p : \sum_{\nu=1}^{n} V^{\nu}(p) \left.\frac{\partial}{\partial x^{\nu}}\right|_p \to V^{\mu} \tag{2.84}$$

where we can write

$$V(p) = \sum_{\nu=1}^{n} V^{\nu}(p) \left.\frac{\partial}{\partial x^{\nu}}\right|_p \tag{2.85}$$

This relates the abstract, coordinate-free vector $V(p)$ and its local coordinate system form $V^\nu(p)$. As expected[a] relation is exactly the same as (eq:localtangentvector).

[a] Since a tangent vector is, above all, a vector.

Derivation 2.9 (Contraction) If we define an additional tangent vector in the same coordinate frame, but with a different index:

$$\omega(p) = \sum_{\mu=1}^{n} \omega_{\mu}(p) \left.\frac{\partial}{\partial x_{\mu}}\right|_p \tag{2.86}$$

We can then prove that

$$\langle \omega(p), V(p) \rangle = \sum_{\mu=1}^{m} \omega_{\mu}(p) V^{\mu}(p) \tag{2.87}$$

This is effectively the *inner product*, which, in index notation, is known as *summation over indices* or simply *contraction*[a].

[a] Because the indices, which disappear, can be seen as 'contracted'.

8

[8] Notationwise, the contraction is made by removing the identical upper and lower indices. However, it is not always possible to make contractions, as we will see in the next chapter.

Chapter 3

Tensors

With the introduction of index notation, we can now leave abstract mathematics and discuss everything in local coordinate systems. For this reason, and as we will see throughout the rest of this book, the index notation is the fundamental formalism we will work with in all of gravitation and cosmology. The first, and most immediate of its many benefits is that we can now perform *general coordinate transformations*, which is simply the transformation of objects from one coordinate system to another. This forms the basis for the a more practical definition of vectors and tensors nonetheless used by relativists. From here, it is then possible to introduce the almighty metric, from which almost all objects in GR emerge.

3.1 General coordinate transformations

Derivation 3.1 (Vector transformations) Let us consider the same abstract covector $\omega(p)$ in two different coordinate systems x^μ and y^μ:

$$\omega(p) = \sum_{\mu=1}^{n} A_\mu dx^\mu|_p = \sum_{\mu=1}^{n} B_\mu dy^\mu|_p \tag{3.1}$$

From (2.84) and the definition of $\omega(p)$ in the first coordinate system, we see easily that

$$\omega(p) \left.\frac{\partial}{\partial x^\mu}\right|_p = A_\mu \tag{3.2}$$

But from the definition of the second coordinate system, we also have

$$\omega(p) \left.\frac{\partial}{\partial x^\mu}\right|_p = \sum_{\nu=1}^{n} B_\nu \left\langle dy^\nu|_p, \left.\frac{\partial}{\partial x^\mu}\right|_p \right\rangle \tag{3.3}$$

Also, note that

$$\left.\frac{\partial}{\partial x^\mu}\right|_p = \sum_{\lambda} \left(\left.\frac{\partial}{\partial x^\mu}\right|_p (y^\lambda) \right) \left.\frac{\partial}{\partial y^\lambda}\right|_p \tag{3.4}$$

Hence, the RHS ultimately reduces to

$$\begin{aligned}
\omega_p \left(\left.\frac{\partial}{\partial x^\mu}\right|_p \right) &= \sum_{\nu=1}^{n} B_\nu \left\langle dy^\nu|_p, \sum_{\lambda} \left(\left.\frac{\partial}{\partial x^\mu}\right|_p (y^\lambda) \right) \left.\frac{\partial}{\partial y^\lambda}\right|_p \right\rangle \\
&= \sum_{\nu=1}^{n} B_\nu \left(\left.\frac{\partial}{\partial x^\mu}\right|_p (y^\lambda) \right) \left\langle dy^\nu|_p, \sum_{\lambda} \left.\frac{\partial}{\partial x^\lambda}\right|_p \right\rangle \\
&= \sum_{\nu=1}^{n} B_\nu \left.\frac{\partial}{\partial x^\mu}\right|_p y^\nu A_\mu = \sum_{\nu=1}^{n} B_\nu \left.\frac{\partial y^\nu}{\partial x^\mu}\right|_p
\end{aligned} \tag{3.5}$$

This concludes the proof for the general coordinate transformations of a contravariant vector. Note that:

- A_μ and B_ν are also often written as A'_μ and A_ν.
- The general coordinate transformation for a covariant vector follows a similar proof.

In summary:

Theorem 3.1 (Vector transformations) A covariant vector undergoes general coordinate transformations as

$$V'^{\nu} = \frac{\partial X'^{\nu}}{\partial X^{\mu}} V^{\mu} \tag{3.6}$$

and a contravariant vector transforms as

$$V'_{\nu} = \frac{\partial X^{\mu}}{\partial X'^{\nu}} V_{\mu} \tag{3.7}$$

One can denote the partial derivative term as the *Jacobian*, which will be familiar to anyone who has studied dynamical systems theory:

$$\frac{\partial X'^{\mu}}{\partial X^{\nu}} = J^{\mu}_{\nu} \tag{3.8}$$

Remark 3.1 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 in our derivation. 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.

3.2 Rise of tensors

Now that we have grasped both the concept of a vector field and a covector field, we are in a sublime position to define tensors. As you are already reading this book, we presume that you have had a first (but possibly incomplete) taste of tensors at some point. You may have observed that a 0-dimensional tensor *looks like* a scalar, a 1-dimensional tensor *looks like* a vector, a 2-dimensional tensor *looks like* a matrix, and in general, n-dimensional tensors *look like* n-dimensional arrays.

This statement is misleading and not very meaningful in the long run, but it does provide us with a hint as to what a tensor could look like. So far, even without defining tensors, we know that an array-like object can go up to an arbitrary number of indices. Intuitively, it would be convenient to think of them in terms of *rank*.

Definition 3.1 (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 3.1 A significant distinction must be made here. The rank, while corresponding to the 'dimensionality' of the array-like object, has no relation to 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} all the way to g_{33}.

From hindsight, we know that we are not satisfied with arbitrary objects 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* of these objects can be classified as *tensors*. A scalar is thus a rank-0 tensor, a vector a rank-1 tensor, and so on[1].

Note 3.2 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 this section.

[1]Those which do not are often informally called pseudoscalars, pseudovectors, and so on.

So, with the final roadblocks out of the way, we can finally define our beloved objects, the tensors. With the epitome of mathematical perfection in our grasp, we are now truly on the road to greatness, and nothing will stop us from this point on!

Definition 3.2 (Tensor) A *tensor* is something that transforms as a tensor.

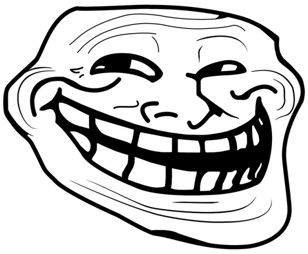


Figure 3.1: Ain't I a stinker?

Well, that didn't quite go as expected. Again, we need some animalistic help here. Let us invite our two Smart Lords of Gravitation. Arthorpodus will recall the age-old aphorism on tensors, and Felis will give some comments.

TRILOBITE 3.1
To computer scientists, a tensor is an n-dimensional array.

CAT 3.2
This definition is wrong.

TRILOBITE 3.3
To physicists, a tensor is something that transforms as a tensor.

CAT 3.4
This definition is incomplete.

TRILOBITE 3.5
To mathematicians, a tensor is an element of a tensor algebra.

CAT 3.6
This is the correct definition.

Thanks, guys. And there they go, riding into the particle horizon again! As it turns out, we can use Felis's comments as a guideline for this section.

While we can haphazardly define tensors at this point purely by their transformation rules, it helps, even for physicists, to acknowledge the fact that the creation of tensors involves the *tensor product*. To understand the tensor product, we consider what operations can be performed on vector spaces. The sum of two vector spaces is simply the *direct sum*, which you have probably seen before. We will not discuss the concept in general, but only how it is performed on two vector spaces.

Definition 3.3 (Direct sum of vector spaces) Given an m-dimensional vector space V and an n-dimensional vector space W with bases $\{\mathbf{v}_i : i = 1, \cdots, m\}$ and $\{\mathbf{w}_j : j = 1, \cdots, n\}$, the vector space sum is a $m+n$-dimensional vector space with the basis

$$V \oplus W = \mathrm{Span}_{i,j}\{\mathbf{v}_i, \mathbf{w}_j : i = 1, \cdots, m; j = 1, \cdots, n\} \tag{3.9}$$

This process is the so-called direct sum $V \oplus W$, which is spanned by the basis

$$V \oplus W = \mathrm{Span}_{i,j}\{\mathbf{v}_i, \mathbf{w}_j\} \tag{3.10}$$

An element in $V \oplus W$ should have the general form

$$\sum_{i=1}^{m} a^i \mathbf{v}_i + \sum_{j=1}^{n} b^j \mathbf{w}^j \tag{3.11}$$

where a^i and b^j are series of constants.

The product of two vector spaces is then the tensor product. Again, if you have taken an undergraduate course on quantum mechanics, this should be familiar to you. Like the direct sum, we only consider its application to vector spaces.

Definition 3.4 (Tensor product of vector spaces) Again, we consider an m-dimensional vector space V and an n-dimensional vector space W with bases $\{\mathbf{v}_i : i = 1, \cdots, m\}$ and $\{\mathbf{w}_j : j = 1, \cdots, n\}$. The tensor product $V \otimes W$ is spanned by the basis

$$V \otimes W = \mathrm{Span}_{i,j}\{\mathbf{v}_i \otimes \mathbf{w}_j\} \tag{3.12}$$

and an element in $V \otimes W$ should have the general form

$$\sum_{i=1}^{m} \sum_{j=1}^{n} c^{ij} \mathbf{v}_i \otimes \mathbf{w}_j \tag{3.13}$$

where c^{ij} is a matrix of constants.
Finally, the tensor product should observe:

- Linearity in both arguments:

$$(v_1 + v_2) \otimes w = (v_1 \otimes w) + (v_2 \otimes w) \quad v \otimes (w_1 + w_2) = (v \otimes w_1) + (v \otimes w_2) \tag{3.14}$$

- Associativity:

$$(\lambda v) \otimes w = v \otimes (\lambda w) = \lambda(v \otimes w) \tag{3.15}$$

Remark 3.2 In practice, when the tensor product is obivous, which is almost always the case, the $\otimes$ symbol is often omitted, leaving $C_{ijkl} = A_{ij} B_{kl}$.
A numerical example is the tensor product of two matrices:

$$\begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \otimes \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & 0 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \\ 0 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} & -1 \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & -1 & 0 \end{bmatrix} \tag{3.16}$$

Remark 3.3 Just like how a series of sums and products can be written as

$$\sum_{i=1}^{n} \text{ and } \prod_{i=1}^{n}$$

respectively, a series of direct sums and tensor products can be written as

$$\bigoplus_{i=1}^{n} \text{ and } \bigotimes_{i=1}^{n}$$

respectively.
Now consider an element of the direct sum vector space in Definition 3.4. This would be a 2D array with $m \times n$ components, all projected contravariantly. This object known as a *tensor*.
However, this is far from the end of the story. In principle, one can take the element of the tensor product of an arbitrary number of tangent spaces and cotangent spaces and still call it a tensor.

Definition 3.5 (Tensor, take 2) The tensor product of r tangent spaces and s cotangent spaces is a

(r, s) *tensor space*[a] given by

$$T_p^{(r,s)}M = \bigotimes^r T_pM \otimes \bigotimes^s T_p^*M \tag{3.17}$$

An element of any such $T_p^{(r,s)}M$ is then known as a (r, s) *tensor*. A *tensor field* is then a map from a manifold to the tensor space

$$T : M \to \bigotimes^r T_pM \otimes \bigotimes^s T_p^*M \tag{3.18}$$

which, in practice, is really a map from a specific point on the manifold to a specific tensor within the tensor space.

[a]This name is rarely used, partly because it is still a vector space, and the name 'tensor space' might give rise to confusion.

This is the closest we will go to the 'tensor algebra' definition mathematicians use. Alternatively, as physicists, we are more interested in how a tensor undergoes general coordinate transformations.

Theorem 3.2 (Tensor transformations) A tensor with rank $p+q$ undergoes general coordinate transformations as

$$T'^{\mu_1\cdots\mu_p}_{\nu_1\cdots\nu_q} = \frac{\partial X'^{\mu_1}}{\partial X^{\lambda_1}} \cdots \frac{\partial X'^{\mu_p}}{\partial X^{\lambda_p}} \frac{\partial X^{\sigma_1}}{\partial X'^{\nu_1}} \cdots \frac{\partial X^{\lambda_q}}{\partial X'^{\nu_q}} T^{\sigma_1\cdots\sigma_p}_{\lambda_1\cdots\lambda_q} \tag{3.19}$$

Remark 3.4 Here we see the significance of the ostensibly trollish Definition 3.2 and Note 3.2. If an object does not transform like the above, we can know for certain that it is *not* a tensor.

Exercise 3.1 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 how the object transforms.

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

$$\frac{d}{d\lambda}\frac{\partial X'^\mu}{\partial X^\nu} = \frac{\partial^2 X'^\mu}{\partial X^\nu \partial X'^\sigma}\frac{dX^\sigma}{d\lambda} \tag{3.20}$$

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

Note 3.3 (Indices) Now that we must work with multiple indices, several points must be noted:

- *Problem with contractions.* When an expression has identical upper and lower indices, it might be tempting to blindly contract them. However, this is only allowed when the end result is well-defined. That is to say, contractions like the well-known contraction of the Riemann tensor to the Ricci tensor[a]

  $$R^\mu_{\mu\lambda\sigma} = R_{\lambda\sigma} \tag{3.21}$$

  is allowed since the RHS, the Ricci tensor, is a well-defined object. In the meantime, contractions like

  $$g_{\mu\nu}R^{\mu\lambda} = g_\nu R^\lambda \tag{3.22}$$

  is disallowed since the rank-1 objects g_ν and R^λ are not well-defined.
- **Conservation of indices:** Indices are always 'conserved' in a given equivalence in the sense that, after all possible contractions, the total composition of upper indices and lower indices in the LHS should be identical to its RHS counterpart.
- **Free indices:** Some indices are arbitrary, and they are known as *free indices*. Consider the

following expression:

$$W^{\mu\nu} = T^{\mu\nu\lambda}_{\lambda} + T^{\mu\nu\sigma}_{\sigma} \tag{3.23}$$

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

$$W^{\mu\nu} = T^{\mu\nu\lambda}_{\lambda} + T^{\mu\nu\lambda}_{\lambda} = 2T^{\mu\nu} \tag{3.24}$$

While this may seem like cheating, it is legitimate, if not essential in many questions.

[a] We will cover this three chapters later.

One important quality of tensors is symmetry or anti-symmetry, which we will now investigate.

Definition 3.6 (Symmetry and anti-symmetry) A tensor is *symmetric* if

$$T^{\mu\nu} = T^{\nu\mu} \tag{3.25}$$

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

$$T^{\mu\nu} = -T^{\nu\mu} \tag{3.26}$$

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

$$T^{\mu\nu} = \underbrace{T^{(\mu\nu)}}_{\text{symmetric part}} + \underbrace{T^{[\mu\nu]}}_{\text{anti-symmetric parts}} \tag{3.27}$$

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:

$$g_{\nu[\lambda}R_{\sigma]\mu} = \frac{1}{2}(g_{\nu\lambda}R_{\sigma\mu} - g_{\nu\sigma}R_{\lambda\mu}) \tag{3.28}$$

- **Symmetrisation:** Also known as the *index anticommutator*, this operation returns a symmetric tensor:

$$g_{\nu(\lambda}R_{\sigma)\mu} = \frac{1}{2}(g_{\nu\lambda}R_{\sigma\mu} + g_{\nu\sigma}R_{\lambda\mu}) \tag{3.29}$$

- **Traceless part:** The traceless (or rather symmetric) part of a tensor $T_{\mu\nu}$ is expressed as $T_{\langle\mu\nu\rangle}$.

Note 3.4 In (3.28) and (3.29), the prefactors are not always $\frac{1}{2}$. While the numerator is always 1, the denominator corresponds to the number of terms that symmetrisation/antisymmetrisation yields. For example:

$$M^{\{\lambda\mu\nu\}} = \frac{1}{6}\left(M^{\lambda\mu\nu} + M^{\lambda\nu\mu} + M^{\mu\lambda\nu} + M^{\mu\nu\lambda} + M^{\nu\lambda\mu} + M^{\nu\mu\lambda}\right) \tag{3.30}$$

Sometimes one would encounter these notations imposed on something more complex like $\epsilon_{\lambda\sigma(\mu}D^{\lambda}S^{\sigma}_{\nu)}$. This simply denotes the part of $\epsilon_{\lambda\sigma\mu}D^{\lambda}S^{\sigma}_{\nu}$ that is symmetric with respect to μ and ν. The same principle applies to the other two notations.

3.3 Kronecker, Levi-Civita and the metric

With this ends our flirtations with abstract mathematics in Part I. While we make several brief returns to it in the next chapters, one is expected to be more comfortable manipulating tensors in index notation than abstract objects. Having finally conceptualised the principal concept in GR, the tensors, we introduce a few tensors with a few basic tensors which you may have seen before.

Definition 3.7 (Kronecker delta) The *Kronecker delta* is defined as

$$\delta^\mu_\nu = \begin{cases} 1 & \mu = \nu \\ 0 & \mu \neq \nu \end{cases} \tag{3.31}$$

It is invariant under transformations.

Fun fact 3.1 On Wikipedia, the title *Kronecker tensor* exists as a redirect of the Kronecker delta.

Note 3.5 The coordinate system and primes become important when we see that later on

$$\frac{\partial X'^\mu}{\partial X'^\nu} = \delta^\mu_\nu \quad \text{but} \quad \frac{\partial X^\mu}{\partial X'^\nu} \neq \delta^\mu_\nu \tag{3.32}$$

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 3.8 (Levi-Civita symbol) The *Levi-Civita symbol* is defined as

$$\epsilon_{\mu\nu\lambda} = \begin{cases} 0 & \mu = \lambda \text{ or } \nu = \lambda \text{ or } \lambda = \mu \\ +1 & \mu, \nu, \lambda \in \text{ permutation of } (1,2,3) \\ -1 & \mu, \nu, \lambda \notin \text{ permutation of } (1,2,3) \end{cases} \tag{3.33}$$

Remark 3.5 Physically, the Levi-Civita symbol defines the outer product. This is seen in the following:

$$(a \times b)_\mu = \epsilon_{\mu\nu\lambda} a_\nu b_\lambda \tag{3.34}$$

In the previous chapter, we discussed the concept of distances via the concept of the metric function. A tensor conceptually associated with it is the almighty *metric*[2], which we will now define abstractly.

Definition 3.9 (Metric) On a manifold M, the almighty metric g is a $(0,2)$ tensor that defines a map at each point p which takes two tangent spaces and maps them to the real numbers:

$$g_p : T_pM \otimes T_pM \to \mathbb{R} \tag{3.35}$$

In practice, it simply maps two tangent vectors to a real number. Given two vectors $X, Y, Z \in TM$ and a function $f \in C^\infty(M)$, the following properties are satisfied:

- Symmetry of arguments:
$$g(X,Y) = g(Y,X) \tag{3.36}$$
- Associativity:
$$g(X, Y + fZ) = g(X,Y) + fg(X,Z) \tag{3.37}$$

The definition above already gives us a good hint of what a metric does. For example, one can see that the metric relates to the concept of distance as it returns, from two tangent vectors, a scalar which we can surmise as the distance. But like any good relativist, let us move our discussion to a local coordinate system. The local coordinate form of the metric relates to its abstract incarnation as follows:

$$g_{\mu\nu} = g\left(\frac{\partial}{\partial x^\mu}, \frac{\partial}{\partial x^\nu}\right) \tag{3.38}$$

where $\frac{\partial}{\partial x^\mu}$ and $\frac{\partial}{\partial x^\nu}$ are derivations, differentials of smooth functions we identified by name in Derivation[3] 2.3. Several comments can immediately be made:

- As a metric has rank 2, it is a 2D matrix on local coordinates. Hence, in n dimensions, a metric has n^2 components. However, only $\frac{n(n+1)}{2}$ components are independent due to symmetry.
- From the symmetry property, we know that $g_{\mu\nu}$ is symmetric in its indices.

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

[3]Ha ha ha

The metric also has an important property which only emerges in a local coordinate system: A metric is always said to be *non-degenerate*. That is, it satisfies the non-degeneracy condition, which actually sheds some light on what it looks like in local coordinates.

Definition 3.10 (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 3.6 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.

Theorem 3.3 (Non-degeneracy condition) $g_{\mu\nu}$ has a non-zero determinant.

Derivation 3.2 (Line element) So far, we are still no closer to the physical meaning of the metric, but hints already exist here and there. From (3.35), we recall that a metric takes tangent vectors as inputs. We also recall from (2.40) that any two tangent vectors v and w can be written in the coordinate basis as

$$v = v^\mu \left.\frac{\partial}{\partial x^\mu}\right|_p \qquad w = w^\nu \left.\frac{\partial}{\partial x^\nu}\right|_p , \tag{3.39}$$

Multiplying v, w and (3.38), one finds

$$g_{\mu\nu}(p) v^\mu w^\nu = g_p(v, w) \tag{3.40}$$

Now consider a smooth curve $\gamma(\lambda)$ in M with parameter λ, with coordinates

$$X^\mu(\lambda) = x^\mu(\gamma(\lambda)) \tag{3.41}$$

The tangent vector to the curve at parameter value λ is

$$\dot{\gamma}(\lambda) = \frac{d\gamma}{d\lambda} \in T_{\gamma(\lambda)}M \tag{3.42}$$

which in the coordinate basis, has the form

$$\dot{\gamma}(\lambda) = \frac{dX^\mu}{d\lambda} \left.\frac{\partial}{\partial x^\mu}\right|_{\gamma(\lambda)} . \tag{3.43}$$

If we look at an infinitesimal step along the curve, $\lambda \to \lambda + d\lambda$, the corresponding infinitesimal coordinate change is

$$dX^\mu = \frac{dX^\mu}{d\lambda} d\lambda \tag{3.44}$$

and the associated tangent vector is

$$v = \dot{\gamma}(\lambda) d\lambda = \left(\frac{dX^\mu}{d\lambda} d\lambda\right) \frac{\partial}{\partial x^\mu} = dX^\mu \frac{\partial}{\partial x^\mu} \tag{3.45}$$

So the infinitesimal displacement along the curve is encoded in the tangent vector as

$$v = dX^\mu \frac{\partial}{\partial x^\mu} \tag{3.46}$$

Inserting this into (3.40), one finds

$$g(v, v) = g_{\mu\nu} dX^\mu dX^\nu \tag{3.47}$$

But we also note that the line element, by definition, is the metric evaluated on this displacement vector with itself:

Definition 3.11 (Line element) The line element represents the infinitesimal distance squared along the curve:

$$ds^2 := g_{\gamma(\lambda)}(v, v) \tag{3.48}$$

Thus, one relates the metric the line element by

$$ds^2 = g_{\mu\nu} dX^\mu dX^\nu \tag{3.49}$$

It is hence possible to write the line element and the metric interchangeably. For example:

$$ds^2 = dx^2 + dy^2 \leftrightarrow g_{\mu\nu} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.50}$$

Definition 3.12 (Metric signature) The *signature* of a metric denotes its signs and has two common forms. Using the *Minkowski metric*, which we will define in (5.2), as an example:

- The number of positive and negative metric components. i.e. $(3, 1)$[a].
- The signs of the metric components. i.e. $(-, +, +, +)$

A second example is the 4D Euclidian metric, whose signature is$(4, 0)$ or $(+, +, +, +)$.

[a]Not to be confused with the notation (p, q) in the context of 'a (p, q) tensor'.

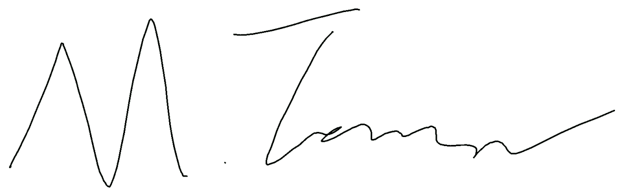


Figure 3.2: Signature of the metric.

Definition 3.13 (Riemannian and pseudo-Riemannian manifolds) A manifold and its corresponding metric are said to be *Riemannian* if the metric is *positive-definite*. That is, if

$$g(X, X) \geq 0 \tag{3.51}$$

and *pseudo-Riemannian* otherwise (i.e. $g(X, X) < 0$).

A few comments can be made:

- Riemannian metrics have signature $(+, +, \cdots, +)$, while pseudo-Riemannian metrics have signature $(-, +, \cdots, +)$.
- In differential geometry, one is mostly interested in Riemannian manifolds, while in gravitation and cosmology, one is *always* interested in pseudo-Riemannian manifolds.
- One example of a Riemannian metric is the 4D Euclidean metric, while one example of a pseudo-Riemannian metric is the Minkowski metric.

Definition 3.14 (Norm) The so-called *norm* of a tensor generalises the concept of *magnitude* in vectors. For a rank-1 tensor:

$$||V|| = \sqrt{g_{\mu\nu} V^\mu V^\nu} \tag{3.52}$$

For a rank-2 tensor:

$$||T|| = \sqrt{g_{\mu\lambda} g_{\nu\sigma} T^{\mu\nu} T^{\lambda\sigma}} \tag{3.53}$$

and so on.

Remark 3.7 A manifold can admit more than one metric. Yet at the fundamental level, the spacetime structure of the universe can only be governed by a single metric. The multiplicity of known metrics

(Schwarzschild, Kerr, etc.) represents different possible solutions or approximations, not competing structures on the same physical spacetime.

Definition 3.15 (Isometry) For two manifolds M and M' with metrics g and g', a diffeomorphism $f : M \to M'$ is a so-called *isometry* if the metric is invariant under its corresponding pushforward f_* of tangent vectors:

$$g(v, v) = g'(f^*v, f^*, v) \tag{3.54}$$

If f is also a local diffeomorphism, then it is called a *local isometry*.

Derivation 3.3 (Example isometries) The vector V generates an isometry if the Lie derivative of the metric with respect to it is zero:

$$\mathcal{L}_V g_{\mu\nu} = 0 \tag{3.55}$$

Rotational invariance:

$$V^\mu_\phi = \delta^\mu_\phi \tag{3.56}$$

Aside from giving birth to a vast number of mathematical objects we are interested in as relativists, the most prominent use of the metric is perhaps shifting indices up and down:

Theorem 3.4 (Raising and lowering indices) One of the most common uses of metrics is to raise or lower indices:

$$T^\mu = g^{\mu\nu} T_\nu \quad T_\mu = g_{\mu\nu} T^\nu \tag{3.57}$$

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

Definition 3.16 (Inverse metric) By definition, the inverse metric $g^{\mu\nu}$ satisfies

$$g_{\mu\nu} g^{\nu\lambda} = \delta^\lambda_\mu \quad g_{\mu\nu} g^{\mu\nu} = \delta^\mu_\mu = n \tag{3.58}$$

where n is the dimension of the metric.

Remark 3.8 As a metric has rank 2, one can solve for the 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.

3.4 Covariant derivative

Quote 3.1 Tell you what: forget covariant; let's just call it the shmovariant derivative.

Leonard Susskind, to Andre Cabannes, in General Relativity: The Theoretical Minimum

Unfortunately, we are not done with basic differential geometry yet, because we have *still* not properly defined how one can take the derivative of a tensor in a physically meaningful way. While we have indeed defined partial derivatives in Derivation 2.6, they cannot be used blindly for the simple reason that we do not know for certain whether the partial derivative of a tensor remains a tensor.

Exercise 3.2 Show that the Lie derivative of a tensor remains a tensor.

Up to this point, we have ignored this problem as we are living in Euclidean or Minkowski space, where the partial derivative of a tensor is always a tensor, but if one introduces curvature, this is not the case, which makes it a sad day for the trilobites. This is seen easily in the following example, where flatness is not assumed.

Derivation 3.4 (Transformation of partial derivatives) We take a simple example and transform the partial derivative of a simple vector A^m:

$$\begin{aligned}\frac{\partial A^\mu}{\partial X^\mu} &= \frac{\partial X^\lambda}{\partial X'^\nu}\frac{\partial}{\partial X^\lambda}\left(\frac{\partial X^\mu}{\partial X^\sigma}A^\sigma\right)\\ &= \frac{\partial X^\lambda}{\partial X'^\nu}\left(\frac{\partial^2 X'^\mu}{\partial X^\sigma \partial X^\lambda}A^\sigma + \frac{\partial X'^\mu}{\partial X^\sigma}\frac{\partial A^\sigma}{\partial X^\lambda}\right)\\ &= \frac{\partial X^\lambda}{\partial X'^\nu}\frac{\partial^2 X'^\mu}{\partial X^\sigma \partial X^\lambda}A^\sigma + \frac{\partial X^\lambda}{\partial X'^\nu}\frac{\partial X'^\mu}{\partial X^\sigma}\frac{\partial A^\sigma}{\partial X^\lambda}\end{aligned} \tag{3.59}$$

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

This motivates the invention of a new derivative that returns a tensor when given one. For more seasoned readers, this concept will not be alien. We now give a few examples of what physicists often call *extended derivatives*:

- The infamous, so-called, *Navier-Stokes equation*[4] reads

$$\frac{D\vec{V}}{Dt} = \nabla p + v\nabla^2\vec{V} \tag{3.60}$$

  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 field) covariant derivative* $\left(\nabla - \frac{iq}{\hbar c}\vec{A}\right)^2$ where $\vec{A}$ is the vector potential.

Definition 3.17 (Affine connection) Given a manifold M and a vector field X on M, a *connection* or an *affine connection* to the metric D on a manifold M is an operator that assigns, for each X, the mapping

$$D_X : TM \to TM \tag{3.61}$$

such that, for all vector fields $X, Y, Z \in TM$ and functions $f \in C^\infty M$, D satisfies:

$$D_X(Y+Z) = D_XY + D_XZ \tag{3.62}$$

$$D_{X+Y}Z = D_XZ + D_YZ \tag{3.63}$$

$$D_{fX}Y = fD_XY \tag{3.64}$$

$$D_X(fY) = X(f)Y + fD_XY \tag{3.65}$$

D_XY, known as the *covariant derivative* of Y along X, is the extended derivative we primarily make use of in gravitation.

Remark 3.9 The Lie derivative/commutator obeys all but the third condition.

Derivation 3.5 (Christoffel symbol) Now let us see how this works in the index notation we physicists know and love. First of all, we note that in gravitation, the covariant derivative is conventionally denoted by the symbol ∇ instead of D. We also introduce, given the local coordinate system $x^\mu = (x^1, x^2, \ldots, x^n)$, the notation

$$\nabla_\mu = \nabla_{\frac{\partial}{\partial x^\mu}}\Big|_p \tag{3.66}$$

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

For given vector fields X^μ and Y^μ in the coordinate frame x^μ, we hence have

$$\begin{aligned} D_X Y &= \nabla_{\sum X^\mu \frac{\partial}{\partial x^\mu}\big|_p} \sum_\nu Y^\nu \frac{\partial}{\partial x^\nu}\bigg|_p \\ &= \sum X^\mu \nabla_\mu \sum_\nu Y^\nu \frac{\partial}{\partial x^\nu}\bigg|_p \\ &= \sum X^\mu \left(\frac{\partial}{\partial x^\mu}\bigg|_p (Y^\nu) \frac{\partial}{\partial x^\nu}\bigg|_p + Y^\nu \nabla_\mu \frac{\partial}{\partial x^\nu}\bigg|_p \right) \end{aligned} \tag{3.67}$$

If define an additional object $\Gamma^\lambda_{\mu\nu}$ known as a *connection coefficient*, satisfying

$$\nabla_\mu \frac{\partial}{\partial x^\mu}\bigg|_p = \sum_{\lambda=1}^{n} \Gamma^\lambda_{\mu\nu} \frac{\partial}{\partial x^\lambda}\bigg|_p \tag{3.68}$$

Hence, (3.67) becomes

$$D_X Y = \sum \left(X^\mu \partial_\mu Y^\lambda + \Gamma^\lambda_{\mu\nu} X^\mu Y^\nu \right) \frac{\partial}{\partial x^\lambda}\bigg|_p \tag{3.69}$$

> **Quote 3.2** 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*

$\Gamma^\lambda_{\mu\nu}$ is then our good friend, the *Christoffel symbol*. Sometimes, they are called *Christoffel connections* due to their nature as an affine connection.

Remark 3.10 Aren't you a little qt 3.14? Yes you are! Yes you are!
In terms of the metric, the Christoffel reads

Definition 3.18 (Christoffel symbol)

$$\Gamma^\lambda_{\mu\nu} = \frac{1}{2} g^{\lambda\sigma} \left(g_{\sigma\mu,\nu} + g_{\nu\sigma,\mu} - g_{\mu\nu,\sigma} \right) \tag{3.70}$$

> **Quote 3.3** 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*

Fun fact 3.2 The Christoffel symbol we commonly 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:

$$\Gamma_{\mu\nu\lambda} = g_{\mu\sigma} \Gamma^\sigma_{\mu\nu} \tag{3.71}$$

Exercise 3.3 Show that the Christoffel symbol transforms under general coordinate transformations as

$$\Gamma'^\mu_{\nu\lambda} = \frac{\partial X'^\mu}{\partial X^\lambda} \frac{\partial X^\sigma}{\partial X'^\nu} \frac{\partial X^\rho}{\partial X'^\lambda} \Gamma^\lambda_{\sigma\rho} + \frac{\partial^2 X^\rho}{\partial X'^\nu \partial X'^\lambda} \frac{\partial X'^\mu}{\partial X^\rho} \tag{3.72}$$

It is then possible to finally define the covariant derivative ∇:

Definition 3.19 (Covariant derivative)

$$\nabla_\mu T^\nu = \partial_\mu T^\nu + \Gamma^\nu_{\mu\lambda} T^\lambda \tag{3.73}$$

Exercise 3.4 Show that, in Euclidian space, the covariant derivative just reduces to the good ol' partial derivative.

Theorem 3.5 (Covariant derivative properties) The covariant derivative has the following properties:

1. Linearity: for all $\alpha, \beta \in \mathbb{R}$
$$\nabla_\mu(\alpha A + \beta B) = \alpha \nabla_\mu A + \beta \nabla_\mu B \tag{3.74}$$

2. Leibnitz rule:
$$\nabla_\mu(AB) = B\nabla_\mu A + A\nabla_\mu B \tag{3.75}$$

3. Commutativity with contraction:
$$\nabla_\mu A^{\nu_1 \cdots \lambda \cdots \nu_m}_{\sigma_1 \cdots \lambda \cdots \sigma_n} = \nabla_\mu A^{\nu_1 \cdots \nu_m}_{\sigma_1 \cdots \sigma_n} \tag{3.76}$$

4. Torsion-free: for all smooth functions $f \in C^\infty(M)$
$$\nabla_\mu \nabla_\nu f = \nabla_\nu \nabla_\mu f \tag{3.77}$$

Remark 3.11 This property holds in standard GR, and more generally modified gravity theories that assume zero torsion[a], but not modified gravity theories where the torsion tensor is non-zero.

[a] We will discuss torsion in depth in the next section.

Note 3.6 The covariant derivatives of the Kronecker delta and the Levi-Civita symbol are always zero.

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

Derivation 3.6 (Covariant derivatives of higher-rank tensors) We observe, for example, $\nabla_a T_b$:

$$\nabla_\mu T_\nu = \partial_\mu T_\nu - \Gamma^\lambda_{\mu\nu} T_\lambda \tag{3.78}$$

Notice how the arbitrary index c went upwards? We therefore observe the almost slavish loyalty of our 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

$$\nabla_\lambda T^{\mu\nu} = \partial_\lambda T^{\mu\nu} + \Gamma^\mu_{\lambda\sigma} T^{\sigma\nu} + \Gamma^\nu_{\lambda\sigma} T^{\mu\sigma} \tag{3.79}$$

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

$$\nabla_\lambda T^\mu_\nu = \partial_\lambda T^\mu_\nu + \Gamma^\mu_{\lambda\sigma} T^\sigma_\nu - \Gamma^\sigma_{\lambda\mu} T^\mu_\sigma \tag{3.80}$$

$$\nabla_\lambda T_{\mu\nu} = \partial_\lambda T_{\mu\nu} - \Gamma^\sigma_{\lambda\mu} T_{\sigma\nu} - \Gamma^\sigma_{\lambda\nu} T_{\mu\sigma} \tag{3.81}$$

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 contravariant positions *on the Christoffel* are negative.

Chapter 4

Classical mechanics

Having grasped the mathematical toolkit, the goal now is to reinvent all the wheels. That is to say, to reformulate everything preceding GR in this formalism we have introduced in the last two chapters. This will not be too hard, for the simple reason that we already know well the result we want to reproduce, which, up to this point, we have blindly taken for granted. The key to this is Lagrangian and Hamiltonian mechanics, which live in the action formalism. Although strongly associated with classical mechanics, the principles they underline are used in possibly every single branch of physics, save for thermodynamics.

4.1 Action principle

Up to the mid-20[th] century, significant equations of motion had been usually derived by intuition, and this (understandably) carries over when these equations are introduced in undergraduate physics. However, with the benefit of hindsight, we are able to do this the proper way by using the so-called *action principle*, which is possibly the most important idea in the entirety of physics. The two elements leading to the action principle are the action itself and the concept of symmetries.
We begin by transitioning from classical mechanics to classical field theory and work slowly towards the action, Starting with the *variational formalism*. Two equivalent formulations of the variational formalism exist - Lagrangian and Hamiltonian mechanics. In classical mechanics, the central quantities are the 4-position x (or often q) and momentum p.

Note 4.1 (Reference frames) We can choose certain frames that simplify calculations:

- For a spacelike separation[a] $(x-y)^2 < 0$, one can always, without loss of generality, choose a frame to set $(x^0 - y^0) = 0$.
- For a timelike separation $(x-y)^2 > 0$, one can always, without loss of generality, choose a frame to set $(\vec{x} - \vec{y}) = 0$.

[a] We will define this by the end of this chapter.

We now take the monumental step of actually understanding what a field theory actually means. The central point is migrating from a coordinate-centric system we have seen up to this point to a field-centric system. In field theories, the 4-position x is replaced with a 4-field $\phi(x) = (\phi_0, \phi_1, \phi_2, \phi_3)$. ϕ_1, ϕ_2 and ϕ_3 are simply the spatial components of the corresponding 3-field, while ϕ_0 is a *scalar* or *time-like* component of the 4-field[1]. For example, we consider the *Lagrangian density*[2] with only one *kinetic term*. Previously, this would merely be the kinetic energy:

$$\mathcal{L} = \frac{mv^2}{2} \tag{4.1}$$

In a field theory, we turn the velocity to derivative over 4-coordinates or the *4-derivatives* of the field and absorb the coupling-like mass m to a normalised value of 1. This then constructs the simplest field theory Lagrangian, which is that of a free massless scalar field, which can be used to model particles like massless scalar bosons[3]:

[1] For example, in the electromagnetic 4-potential, ϕ_0 is the electric scalar potential.
[2] Often also simply called the *Lagrangian*, although you will be able to tell the difference by looking at the notation.
[3] We see it more often in approximate models as fundamental massless scalar particles with zero mass are rare.

Definition 4.1 (Free massless scalar field Lagrangian)

$$\mathcal{L} = \frac{1}{2}\partial_\mu\phi\partial^\mu\phi \tag{4.2}$$

The sole term is the *kinetic energy density*, which arises from the variation of the field ϕ over the 4-coordinates.

The only term in this Lagrangian is a so-called kinetic term, which we can now properly define as a term consisting of field derivatives. In analogy to classical mechanics, the rest of the terms make up the *potential*.
One way we can introduce a potential to the Lagrangian is to make the free field massive instead of massless. Note that this does not mean that the field itself is massive (which makes little sense) but rather that the particle that generates the field is massive. This gives rise to a mass coupling[4] term, and the Lagrangian becomes the standard Lagrangian of a free massive scalar field:

Definition 4.2 (Free massive scalar field Lagrangian)

$$\mathcal{L} = \frac{1}{2}\partial^\mu\phi\partial_\mu\phi - \frac{1}{2}m^2\phi^2 \tag{4.3}$$

where the second term is the *potential energy density*.

Remark 4.1 So far, the only scalar field observed in nature is the (massive) Higgs field. The hypothetical inflation field is also a scalar field.
Remark 4.2 This Lagrangian is used to derive the infamous Klein-Gordon equation, a failed candidate for a quantum equation of motion incorporating SR. Let us compare this CFT result with its classical mechanics analogue, which is the Lagrangian of a harmonic oscillator:

$$\mathcal{L} = \frac{1}{2}m\dot{x}^2 - \frac{1}{2}kx^2 \tag{4.4}$$

Essentially, $\frac{1}{2}m^2|\phi|^2$ represents the field-theory version of Hooke's law. It is a potential energy term that determines how the field oscillates about its vacuum. The mass term m^2 replaces the spring constant k and sets the curvature (strength) of the potential.
From the Lagrangian, we can formulate how quantities related to it are defined in CFT. In classical mechanics, the *canonical momentum* is defined as

$$p = \frac{dL}{d\dot{q}} \tag{4.5}$$

where q is the generalised 4-coordinates. In CFT, this becomes

Definition 4.3 (Canonical momentum)

$$\pi = \frac{\partial\mathcal{L}}{\partial(\partial_t\phi)} \tag{4.6}$$

The action in CFT is unchanged from its CM counterpart, but uses the CFT Lagrangian instead of the CM Lagrangian:

Definition 4.4 (Action) For a set of fields ϕ_i with the 4-position x^i, the *action* is defined as

$$S = \int \mathcal{L}(\phi, \partial_\mu\phi)d^4x = \int Ldx^0 = \int Ldt \tag{4.7}$$

where L is the Lagrangian and $\mathcal{L}$ is the Lagrangian density.

We are now in a position to discuss symmetries. Physicists use the word 'symmetry' as a more sophisticated way of saying that a physical quantity we are interested in stays invariant under the change of

[4]We will often see the word 'coupling'. Sometimes, it is the short form of the *coupling constant*, which denotes any physical constant that a term in the Lagrangian may have. We then speak of the field *coupling* to whatever the physical quantity this coupling constant represents. For example, in a mass term, the coupling constant is the mass, and we say that in the term, the field couples to mass.

some other 'background' quantity. If this physical quantity of interest instead changes with the background quantity, we then sadly say that symmetry is 'broken'. Broadly speaking, we are interested in the following symmetries:

- **Global and local gauge symmetries:**
  - *Global symmetries* have transformations that are identical everywhere in spacetime. They lead to conserved quantities (like energy, momentum, charge) and usually reflect real physical invariances.
  - *Local gauge symmetries* have transformations that can vary from point to point in spacetime. They are not symmetries of nature but symmetries (or rather redundancies) in our description.
- **External and internal symmetries:**
  - *External symmetries* are transformations that involve changes to the spacetime coordinates themselves. One example is Poincaré symmetry.
  - *Internal symmetries* are transformations that act on internal degrees of freedom of fields (e.g. charge, spin, etc.), leaving spacetime coordinates unchanged.
- **Discrete and continuous symmetries:**
  - *Continuous symmetries* are governed by transformation parameters that can admit a continuous range of values. One example is Poincaré symmetry.
  - *Discrete symmetries* involve transformations that take on only specific values. Examples are the C, P and T symmetries.

So far, we have heard about the layman's version of Noether's theorem 'All symmetries lead to conservation laws'. We have made a conceptual overview of symmetries, but what are their mathematical implications? As it turns out, symmetries are defined with respect to an action principle. Consider an *infinitesimal coordinate transformation*

$$x^\mu \to x'^\mu = x^\mu + \epsilon^\mu \tag{4.8}$$

Up to the first order expansion, a generic field in x^μ undergoes the corresponding transformation:

$$\phi(x) \to \phi'(x') = \phi(x) - \epsilon^\mu \partial_\mu \phi(x) \tag{4.9}$$

The variation of the field can then be written as

$$\delta\phi(x) = \phi'(x') - \phi(x) = -\epsilon^\mu \partial_\mu \phi(x) \tag{4.10}$$

An important point of note concerns the Lagrangian (density). While it has rank 0, the Lagrangian $\mathcal{L}$ in some theories might not transform as a scalar. Rather, its variation takes the general form of the total derivative of a current-like vector field K^μ. Without considering fields, the variation of the Lagrangian *arising from a symmetry transformation* has the general form

$$\delta\mathcal{L} = \partial_\mu K^\mu(\phi, \partial_\mu \phi) \tag{4.11}$$

where:

- Physically, $K^\mu(\phi, \partial_\mu\phi)$ is the measure of the failure of $\mathcal{L}$ to transform as a scalar.
- $\partial_\mu K^\mu(\phi, \partial_\mu\phi)$ is then a 'total derivative' of $K^\mu(\phi, \partial_\mu\phi)$, which depends on both the field ϕ and the field's 4-derivative $\partial_\mu\phi$.

Derivation 4.1 (Action principle) We can solve for the variation of the action by integrating the variation of the Lagrangian (4.11), which gives:

$$\delta S = \int d^4x \delta\mathcal{L} = \int d^4x \partial_\mu K^\mu(\phi, \partial_\mu\phi) \tag{4.12}$$

Using the divergence theorem, this integral can be converted into a surface integral over the boundary

$\mathcal{S} = \partial V$ of the spacetime region V:

$$\delta S = \int_{\partial V} d^3x K^\mu(\phi, \partial_\mu \phi) n_\mu \tag{4.13}$$

where n_μ is the normal vector to the boundary. We are left with a *boundary term* that is exactly $\partial_\mu K^\mu(\phi, \partial_\mu \phi) n_\mu$ where n_μ is a directional 4-vector.
There are two scenarios in which this boundary term can be ignored:

- The variation $\delta\phi$ (and with that, ϕ and its derivative) vanishes on the boundary[a].
- The boundary extends into infinity.

The important step now is to *assume the first point*, which can be justified if we impose boundary conditions. Depending on the physical scenario, we usually use one of the two main boundary conditions.

Definition 4.5 (Dirichlet boundary condition) The *Dirichlet boundary condition* or the *boundary condition of the first type* sets the position to be time-invariant at the boundary:

$$\dot{q}|_{\partial V} = 0 \quad \partial_\mu \phi|_{\partial V} = 0 \tag{4.14}$$

Definition 4.6 (Neumann boundary condition) The *Neumann boundary condition* or the *boundary condition of the second type* sets the momentum is time-invariant at the boundary:

$$\left.\frac{\partial L}{\partial \dot{q}}\right|_{\partial V} = 0 \quad \left.\frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)}\right|_{\partial V} = 0 \tag{4.15}$$

where we see that this boundary condition is realised by exploiting Hamilton's equations[a].

[a] We only derive Hamilton's equations later on. However, our current lack of knowledge of Hamilton's equations does not prevent us from imposing the Neumann boundary condition. Rather, it only temporarily prevents us from interpretating the LHS as the momentum time derivative - which we have revealed as a spoiler anyway.

Either way, the term $\partial_\mu K^\mu(\phi, \partial_\mu \phi) n_\mu$ then vanishes due to its dependency on both ϕ and $\partial_\mu \phi$:

$$\int_{\mathcal{S}} d^3x K^\mu(\phi, \partial_\mu \phi) n_\mu = 0 \tag{4.16}$$

In fact, assuming that we choose suitable boundary conditions, all total derivative terms of the form $\partial_\mu K^\mu(\phi, \partial_\mu \phi) n_\mu$ are boundary terms, and hence identically vanish.

Note 4.2 Boundary terms do not contribute to the equations of motion. As such, we can always add or subtract boundary terms to/from the Lagrangian arbitrarily[a].

[a] The same applies to zero terms for the simple reason that they are zero

From this, we see that the action is invariant under the symmetry:

Theorem 4.1 (Action principle)

$$\delta S = 0 \tag{4.17}$$

This is the almighty *action principle.*

[a] This is typically justified in physical field theories where fields and their variations vanish at spatial or temporal infinity.

Remark 4.3 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*[5]. This simply means that the time derivative of the action of an isolated system is zero. As the principle can be used for action generated by *any* field, it is often mentioned as '*an* action principle' instead of '*the* action principle'.

[5] 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.

4.2 Equations of motion

The physical significance of our previous derivation is not immediately obvious, especially with respect to how the boundary term vanishes by dint of the Dirichlet and Neumann boundary condition. One can lift this shrowd of confusion by realising that the action principle allows us to derive the equations of motion. Let us show this with the general example in classical mechanics.

Derivation 4.2 (Euler-Lagrange equations) Now that we have assumed the vanishing of the boundary term, let us evaluate the Lagrangian variation $\delta\mathcal{L}(\phi, \partial_\mu\phi)$ explicitly. We consider an infinitesimal variation of the field:

$$\phi(x) \to \phi(x) + \delta\phi(x) \tag{4.18}$$

The variation of the Lagrangian is exactly analogous to differentiating a function of two variables:

$$f(x, y) \to f(x + \delta x, y + \delta y) \to \delta f = \frac{\partial f}{\partial x}\delta x + \frac{\partial f}{\partial y}\delta y \tag{4.19}$$

As $\mathcal{L}$ has two variables, the field ϕ and the field derivative ∂_μ, its total variation is given by

$$\delta\mathcal{L} = \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi\right) \tag{4.20}$$

Using the product rule gives

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\phi}\delta\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta(\partial_\mu\phi) \tag{4.21}$$

Noting that $\delta(\partial_\mu\phi) = \partial_\mu(\delta\phi)$, we can write

$$\delta\mathcal{L} = \frac{\partial\mathcal{L}}{\partial\phi}\delta\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\partial_\mu(\delta\phi) \tag{4.22}$$

Inserting this result into (4.12) gives

$$\delta S = \int d^4x \left[\frac{\partial\mathcal{L}}{\partial\phi}\delta\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\partial_\mu(\delta\phi)\right] \tag{4.23}$$

We can apply integration by parts to the second term and apply Gauss's law[a]. This gives

$$\delta S = \int d^4x \frac{\partial\mathcal{L}}{\partial\phi}\delta\phi + \underbrace{\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi}_{\text{boundary term}}\Bigg|_{\partial V} - \int d^4x \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\right)\delta\phi \tag{4.24}$$

where we recall that ∂V denotes the boundary.
Through this process, we have explicitly exposed the so-called boundary term. But there is something very peculiar about this boundary term, isn't there? Let us inspect the two components:

- $\delta\phi$ is nothing but the time derivative (variation) of the position (field), which is zero at the boundary under the Dirichlet boundary condition (4.14).
- $\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}$ is nothing but the time derivative of the momentum, which is zero at the boundry under the Neumann boundary condition (4.15).

So you see, no matter which boundary condition we take, this boundary term is ultimately nothing but zero. The important conclusion you should yield from our discussion is the following:

Note 4.3 Due to its vanishing, the boundary term *does not contribute to the equations of motion.*

In contrast, the rest of the expression, which contribute to the equations of motion, is known as the *bulk term*[b].
The final step we have left is to simply apply the action principle, which then leads to the famous Euler-Lagrange equations:

Theorem 4.2 (Euler-Lagrange equations)

$$\frac{\partial \mathcal{L}}{\partial \phi} - \partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} = 0 \tag{4.25}$$

[a]As such, the boundary term is also called the *surface term*.
[b]This expression is rarely used, and almost always in gravitation.

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

- The bulk term is so-called because it integrates over the entire volume of spacetime. It is the term that contributes to the action. When an action principle is imposed, the integrand of the bulk term vanishes, as seen in (4.25).
- The boundary term reflects the influence of boundary conditions for the action. It is the term that does not contribute to the action. By imposing an action principle, we have also assumed that $\delta\phi = 0$ on the boundary - a boundary condition.

Like in CM, the *Hamiltonian* is essentially a Legendre transformation of the Lagrangian:

Definition 4.7 (Hamiltonian and Hamiltonian density) The *Hamiltonian* H is

$$H = \int \mathcal{H}(\phi, \pi, \partial_\mu \phi) d^3x \tag{4.26}$$

which is the volume integral of the *Hamilton density* $\mathcal{H}$. Also simply called the *Hamiltonian*, it is given by

$$\mathcal{H}(\phi_i, \pi_i, t) = \sum_i \pi_i \dot{\phi}_i(\phi_j, \pi_j) - \mathcal{L}(\dot{x}_k(x_j, p_j), x_k, t) \tag{4.27}$$

where ϕ is the field and π is the canonical momentum.

Remark 4.4 As it turns out, the quantity which we have been led to believe to be the Hamiltonian as undergrads is actually the Hamiltonian density $\mathcal{H}$.

Derivation 4.3 (Hamilton's equations) By taking the variation of (4.27), one finds

$$\delta\mathcal{H} = \sum_i \delta\pi_i \dot{\phi}^i(\phi_j, \pi_j) - \delta\pi^i \frac{\partial \mathcal{L}}{\partial \pi^i} = \sum_i \delta\pi_i \dot{\phi}^i(\phi_j, \pi_j) - \sum_i \delta\phi^i(\phi_j, \pi_j)\dot{\pi}_i \tag{4.28}$$

Now compare this against the general variation:

$$\delta\mathcal{H} = \delta\phi^i(\phi_j, \pi_j) \frac{\partial \mathcal{H}}{\partial \phi^i(\phi_j, \pi_j)} + \delta\pi_i \frac{\partial \mathcal{H}}{\partial \pi_i} \tag{4.29}$$

By equating the two expressions for $\delta\mathcal{H}$, we recover the so-called *Hamilton's equations*:

Theorem 4.3 (Hamilton's equations)

$$\frac{\partial \mathcal{H}}{\partial \phi_i} = -\dot{\pi}_i \quad \frac{\partial \mathcal{H}}{\partial \pi_i} = \dot{\phi}_i \tag{4.30}$$

Defining the Poisson bracket $\{f, g\}$ of some two quantities f and g as

Definition 4.8 (Poisson bracket)

$$\{f, g\} = \sum_i \left(\frac{\partial f}{\partial q_i} \frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i} \frac{\partial g}{\partial q_i} \right) \tag{4.31}$$

we can rewrite Hamilton's equations as

$$\dot{x}_i = \{x_i, H\} \quad \dot{p}_i = \{p_i, H\} \tag{4.32}$$

As we will see later, the quantum version of the first equation is simply the Heisenberg equation, or the Schrödinger equation in the Heisenberg picture.

4.3 Noether's theorem

We can now finally derive Noether's theorem. The first step is to realise that the variation of the Lagrangian has been written in two formulations (4.20) and (4.11) respectively. Combining them yields

$$\delta\mathcal{L} = \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi\right) = \partial_\mu K^\mu \to \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi\right) - \partial_\mu K^\mu = 0 \tag{4.33}$$

Now let us define another current-like quantity called the *Noether current*:

Definition 4.9 (Noether current) The 4-vector equivalent of the probability density ϱ is the so-called Noether current, *probability 4-current* or *conserved current* J^μ. Its zeroth component is simply the good ol' probability density, and its 3 other (spatial) components are the *probability (3-)current* J^i.

From the Noether current, one can derive a charge-like quantity representing the total probability called the *Noether charge*, the *probability charge* or the *conserved charge* as it is conserved with respect to time:

Definition 4.10 (Probability charge)

$$Q = \int d^3x J^0 \tag{4.34}$$

As probability is conserved, J^μ is Lorentz-invariant and satisfies the *continuity equation*:

Theorem 4.4 (Continuity equation)

$$\partial_\mu J^\mu = 0 \tag{4.35}$$

Rather cheatingly, we can now equate (4.33) and (4.35):

$$\partial_\mu J^\mu = \partial_\mu\left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi\right) - \partial_\mu K^\mu = 0 \tag{4.36}$$

By removing the partial derivatives, we recover the expression for the Noether current, known as Noether's theorem:

$$J^\mu = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi - K^\mu \tag{4.37}$$

In the most common case, the Lagrangian is strictly invariant, which leads to the vanishing of the vector field K^μ. We then have:

Theorem 4.5 (Noether's theorem)

$$J^\mu = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\delta\phi \tag{4.38}$$

The implications are twofold:

- In principle, this is used to calculate (or 'read off' if the reader is experienced) the 4-current (or Noether current) for any field that transforms under a global symmetry.
- In practice, these fields are, among others, the Klein-Gordon and Dirac scalar fields[6].

Quote 4.1 Yes, but it is the same sound

David Steiner, comparing the 'oe' in 'Noether' with 'ö', 21 November 2024

Note 4.4 Noether's theorem implies the conservation of the charge associated with the probability current:

$$\frac{dQ}{dt} = \int d^3x \partial_0 J^0 = 0 \tag{4.39}$$

Finally, we can directly relate the Noether current and the action, from (4.24) and (4.25), one can see that

$$\frac{\partial\mathcal{L}}{\partial\phi} = \partial_\mu\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)} = \frac{\delta S}{\delta\phi} \tag{4.40}$$

[6]This is covered in any standard text of QFT.

which, by insertion into (4.36), gives

$$\partial_\mu J^\mu = \delta\phi \frac{\delta S}{\delta \phi} = \frac{\delta \phi}{\delta \theta}\frac{\delta S}{\delta \phi} \tag{4.41}$$

where θ is the parameter for an infinitesimal transformation satisfying

$$x \to x' = x + \theta \delta x \tag{4.42}$$

4.4 Geodesic equation

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.
- On a sphere, this is a great circle.

In simpler words, it is the ‘straightest possible line’ of a geometry. More rigorously:

Definition 4.11 (Geodesic) Particles that travel on geodesics satisfy the action principle.

Definition 4.12 (Spacelike, timelike and null) In GR, geodesics can be classified into one of 3 types:

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

$$ds^2 \begin{cases} > 0 & \text{in the signature } (-,+,+,+) \\ < 0 & \text{in the signature } (+,-,-,-) \end{cases} \tag{4.43}$$

  For obvious reasons, spacelike geodesics are utterly unphysical and do not represent any physical objects.

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

$$ds^2 \begin{cases} < 0 & \text{in the signature } (-,+,+,+) \\ > 0 & \text{in the signature } (+,-,-,-) \end{cases} \tag{4.44}$$

  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

$$ds^2 = 0 \tag{4.45}$$

  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.

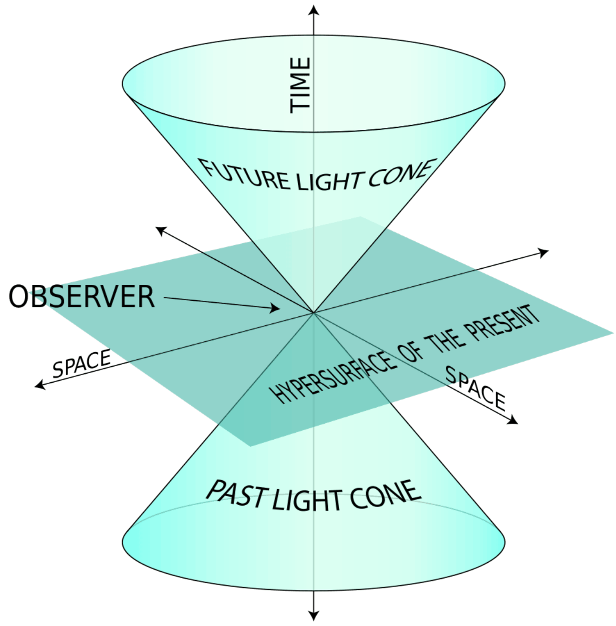


Figure 4.1: Visualisation of a light cone.

Remark 4.5 Spacelike geodesics are inside the light cone; timelike geodesics are outside the light cone; while lightlike geodesics are exactly on the light cone.
Remark 4.6 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.

Definition 4.13 (Worldline) The path of a particle in 4D spacetime is often known as its *worldline*.

TRILOBITE 4.1
This so-called worldline is nothing but a fancy word used by relativists to confuse people.

Indeed, Arthropodus. Now, shall we proceed to deriving the geodesic equation?

TRILOBITE 4.2
Krrrk.

Good!

Derivation 4.4 (Geodesic equation) The action for a free particle is given by the integral of the spacetime interval (or proper time) along the path:

$$S = \int \sqrt{-g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu} d\lambda \tag{4.46}$$

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

$$L = \sqrt{g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda}} \tag{4.47}$$

One can simplify this by replacing the action with:

$$S' = \int g_{\mu\nu}\dot{x}^\mu \dot{x}^\nu d\lambda \tag{4.48}$$

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

$$L' = g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} \tag{4.49}$$

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 equation of motions using this simplified scheme. We first transform the metric into the simplified Lagrangian in (4.49). For example:

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 \leftrightarrow g_{\mu\nu} = L = -1\dot{t}^2 + 1\dot{x}^2 + 1\dot{y}^2 + 1\dot{z}^2 \tag{4.50}$$

By inserting (4.49) into the Euler-Lagrange equations, we can derive the following expression

Theorem 4.6 (Geodesic equation) The *geodesic equation*, which some geodesic X^μ travels on, is

$$\frac{d^2X^\mu}{d\lambda^2} + \Gamma^\mu_{\nu\sigma}\frac{dX^\nu}{d\lambda}\frac{dX^\sigma}{d\lambda} = 0 \tag{4.51}$$

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

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

Remark 4.7 The geodesic equation reflects the maximisation of proper time in GR.

Chapter 5

Special relativity

Quote 5.1 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 their historical derivation. While this makes sense in introductions, we will find it far more intuitive to explain many underlying mechanisms using our toolkit of field theory and tensors.

5.1 Minkowski space

As is well known, special relativity was developed by Einstein in the years immediately preceding general relativity. While widely understood today to be nothing but an effective theory of GR under the limit where masses are *smol*, it remains very useful due to the failure of high energy physicists to incorporate gravitation into the field theory framework.

SR operates in Minkowski space, which unsurprisingly uses the aptly-named Minkowski metric. In the next chapter, where we will properly introduce curvature, we will see that the Minkowski metric yields zero curvature. But we will take this as granted for the time being. The Minkowski space line element reads

$$ds^2 = -dt^2 + dx^2 + dy^2 + dz^2 \tag{5.1}$$

which corresponds to the Minkowski metric, conventionally denoted by $\eta_{\mu\nu}$:

Definition 5.1 (Minkowski metric)

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{5.2}$$

Remark 5.1 For s to be proper time τ, the norm of the 4-position x^μ must observe $||X|| = 1$.

Quote 5.2 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

Note 5.1 (Metric signature) Unlike GR, the convention in HEP dictates that the Minkowski 4-metric in HEP has the signature $(+,-,-,-)$ instead. That is, the line element has the form

$$ds^2 = dt^2 - dx^2 - dy^2 - dz^2 \tag{5.3}$$

In the so-called United States, a country infested by string theorists, the $(+,-,-,-)$ signature is known as the 'West Coast' convention, while the $(-,+,+,+)$ is known as the 'East Coast' convention.

Fun fact 5.1 In fact, in some notations, t is the 4^{th} coordinate instead.

> **Quote 5.3** Good boy.
>
> *Alex Engertsberger, on a student making the correct choice of metric signatures*

There are two possible transformations in Minkowski space, known collectively as Poincaré transformations:

- Translations
- Lorentz transformations.

In this chapter, we will discuss both discussed using field theory. Let us start with translations:

Derivation 5.1 (Translation) Let us assume the same transformations as (4.8) and (4.9). The field variation is then shown in (4.10), which we substitute into the transformation of $\mathcal{L}$ in (4.20). We obtain

$$\delta\mathcal{L} = -\epsilon^\nu \left(\frac{\partial\mathcal{L}}{\partial\phi}\partial_\nu\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\partial_\mu\partial_\nu\phi \right) \tag{5.4}$$

Now substitute this into (4.12):

$$\delta S = -\epsilon^\nu \int d^4x \left(\frac{\partial\mathcal{L}}{\partial\phi}\partial_\nu\phi + \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\partial_\mu\partial_\nu\phi \right) \tag{5.5}$$

By inserting the equivalence from the Euler-Lagrange equations, we can rewrite the integral as

$$\delta S = -\epsilon^\nu \int d^4x \partial_\mu \left(\frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\partial_\nu\phi - \delta^\mu_\nu\mathcal{L} \right) \tag{5.6}$$

where the terms within the bracket is the canonical stress-energy tensor. A simple shift of indices gives its contravariant form:

Definition 5.2 (Stress-energy tensor)

$$T^{\mu\nu} = \frac{\partial\mathcal{L}}{\partial(\partial_\mu\phi)}\partial^\nu\phi - g^{\mu\nu}\mathcal{L} \tag{5.7}$$

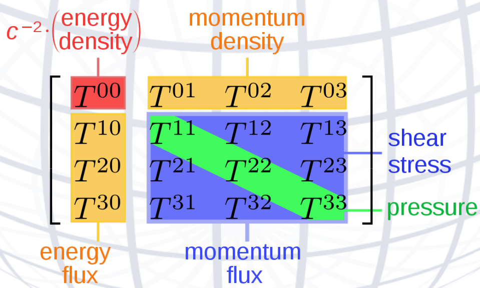


Figure 5.1: Components of the stress-energy tensor.

Importantly, by comparing (4.12) and (5.6), one can identify

$$\delta\mathcal{L} = \partial_\mu(\epsilon_\nu T^{\mu\nu}) = \partial_\mu K^\mu \tag{5.8}$$

In other words, our new friend $\epsilon_\nu T^{\mu\nu}$ resembles the previously seen boundary/surface term K^μ. Finally, if one takes a covariant derivative of the stress-energy tensor and applies the Euler-Lagrange equations, they will find that

$$\nabla_\mu T^{\mu\nu} = 0 \tag{5.9}$$

This shows the well-known conservation of matter-energy content, and follows directly from Noether's theorem applied to spacetime translations.

Exercise 5.1 Show that, for a generic massive scalar field, the stress-energy tensor reads

$$T_{\mu\nu} = -\partial_\mu\phi\partial_\nu\phi + \eta_{\mu\nu}\frac{1}{2}\left(\eta^{\lambda\sigma}\partial_\lambda\phi\partial_\sigma\phi - m^2\phi^2\right) \tag{5.10}$$

Note 5.2 (5.9) merely states that $T^{\mu\nu}$ is invariant when taken the covariant derivative over *4-coordinates*. As such, it describes the conservation of energy-momentum *density* as opposed to energy and momentum themselves. In cosmology, where inflation exists, the total energy of the universe is *not* conserved. Rather, as inflation gives rise to the expansion of spacetime, the total energy of the universe *increases* to preserve the invariance of energy-momentum density.

Remark 5.2 The 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.
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 5.1 (Stress-energy tensor for incompressible, viscous fluids)

$$T^{\mu\nu} = \rho u^\mu u^\nu - (p - \xi\theta)(\eta^{\mu\nu} - u^\mu u^\nu) - 2\eta\sigma^{\mu\nu} \tag{5.11}$$

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

$$\sigma^{\mu\nu} = \frac{1}{2}(u_{\mu,\lambda}h^\lambda_\nu + u_\nu h^\lambda_\mu) - \frac{1}{3}\theta h_{\mu\nu} \tag{5.12}$$

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

$$h_{\mu\nu} = \eta_{\mu\nu} - u_\mu u_\nu \tag{5.13}$$

Definition 5.3 (Perfect fluids) Perfect fluids are idealised fluids with no viscosity or heat conduction.

Remark 5.4 As they have no viscosity, perfect fluids do not have surface tension or shear stresses either, although they may be compressible *or* incompressible.
Remark 5.5 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 5.2 (Stress-energy tensor of perfect fluids) A perfect fluid has the following stress-energy tensor:

$$T_{\mu\nu} = \rho u_\mu u_\nu - p(\eta_{\mu\nu} - u_\mu u_\nu) \tag{5.14}$$

5.2 Lorentz transformation

TRILOBITE 5.1

For illustrative purposes, we will not not assume $c = 1$ in this section and the next.

Hmm, makes sense, but I'm starting to worry that soon it will be Arthropodus calling the shots in this book. One shrudders to imagine the prospective of trilobite world domination, but enought about that. As mentioned earlier, the other component of Poincaré transformations is *Lorentz transformations* denoted by Λ^{ν}_{μ}. Again, we first take the field theory approach.

Derivation 5.2 (Lorentz transformation) A Lorentz transformation can be either a rotation or a *Lorentz boost*. Assuming the parameterisation

$$\Lambda^{\mu}_{\nu} = \delta^{\mu}_{\nu} + \omega^{\mu}_{\nu} \quad \omega^{\mu\nu} = -\omega^{\nu\mu} \tag{5.15}$$

where $\omega^{\nu\mu}$ is some parameter, the coordinate and field transformations are

$$x^{\mu} \to x'^{\mu} = x^{\mu} + \Lambda^{\mu}_{\nu} x^{\nu} \quad \phi(x) \to \phi'(x) = \phi(x) + \frac{1}{2}\omega^{\rho\sigma}\Sigma_{\rho\sigma}\phi \tag{5.16}$$

where $\Sigma_{\rho\sigma}$ are the generators of the representations of the Lorentz group abstract elements corresponding to ϕ (e.g., for scalars $\Sigma_{\rho\sigma} = 0$, for vectors $\Sigma_{\rho\sigma}$ corresponds to antisymmetric tensors, and so on). Again, by using Noether's theorem, we can find that the contribution to the Noether current is the total angular momentum, which includes the orbital angular momentum and spin:

$$M^{\mu\rho\sigma} = x^{\rho}T^{\mu\sigma} - x^{\sigma}T^{\mu\rho} + \frac{\partial\mathcal{L}}{\partial(\partial_{\mu}\phi)}\Sigma^{\rho\sigma}\phi \tag{5.17}$$

For Lorentz transformations, $K^{\mu} = \partial_{\nu}(x^{\nu}J^{\mu} - x^{\mu}J^{\nu})$, and the Noether current becomes:

$$J^{\mu} = \omega_{\rho\sigma}M^{\mu\rho\sigma} \tag{5.18}$$

where the nature of $\omega_{\rho\sigma}M^{\mu\rho\sigma}$ as a boundary term is easily seen.

Hence, combining both types of symmetries, the most general form of the Noether current under Poincaré transformations can be written as:

$$J^{\mu} = \epsilon_{\nu}T^{\mu\nu} + \frac{1}{2}\omega_{\rho\sigma}M^{\mu\rho\sigma} \tag{5.19}$$

where $T^{\mu\nu}$ represents energy-momentum contributions, and $M^{\mu\rho\sigma}$ represents both orbital and intrinsic angular momentum contributions.

Theorem 5.3 (Preservation of the Minkowski metric) The defining relation of the Lorentz transformation is the preservation of the Minkowski metric:

$$\Lambda^{\mu}_{\nu}\Lambda^{\lambda}_{\sigma}\eta_{\mu\lambda} = \eta_{\nu\sigma} \tag{5.20}$$

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

Derivation 5.3 (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

$$\begin{pmatrix} t \\ \vec{x} \end{pmatrix}_S = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \begin{pmatrix} t \\ \vec{x} \end{pmatrix}_E \tag{5.21}$$

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

$$\begin{pmatrix} t \\ \vec{x} \end{pmatrix} = \begin{cases} \begin{pmatrix} t \\ t \end{pmatrix} & \text{moving to the right} \\ \begin{pmatrix} t \\ -t \end{pmatrix} & \text{moving to the left} \end{cases} \tag{5.22}$$

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

$$\alpha + \beta = \gamma + \delta \tag{5.23}$$

for the right-moving photon and

$$\alpha + \beta = \gamma - \delta \tag{5.24}$$

for the left-moving photon. Combining the equations yields

$$\alpha = \delta \quad \beta = \gamma \to \begin{pmatrix} t \\ \vec{x} \end{pmatrix}_S = \begin{pmatrix} \gamma & \delta \\ \delta & \gamma \end{pmatrix} \begin{pmatrix} t \\ \vec{x} \end{pmatrix}_E \tag{5.25}$$

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

$$\begin{pmatrix} t \\ \vec{x} \end{pmatrix}_S = \begin{pmatrix} t \\ 0 \end{pmatrix}_S \tag{5.26}$$

Quote 5.4 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

$$\begin{pmatrix} t \\ \vec{x} \end{pmatrix}_E = \begin{pmatrix} t \\ \vec{v}t \end{pmatrix}_E \tag{5.27}$$

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

$$\gamma = -\delta\vec{v} \tag{5.28}$$

Thus, we have

$$\begin{pmatrix} t \\ \vec{x} \end{pmatrix}_S = \begin{pmatrix} \gamma & -\vec{v}\gamma \\ -\vec{v}\gamma & \gamma \end{pmatrix} \begin{pmatrix} t \\ \vec{x} \end{pmatrix}_E \tag{5.29}$$

and the transformation matrix is

Definition 5.4 (Lorentz transformation in matrix form)

$$\lambda^{\mu}_{\nu} = \begin{pmatrix} \gamma & -\vec{v}\gamma \\ -\vec{v}\gamma & \gamma \end{pmatrix} \tag{5.30}$$

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

$$(\lambda^{\mu}_{\nu})^{-1} = \begin{pmatrix} \gamma & \vec{v}\gamma \\ \vec{v}\gamma & \gamma \end{pmatrix} \tag{5.31}$$

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 5.5 (Lorentz factor)

$$\gamma = \frac{1}{\sqrt{1 - \frac{|\vec{v}|^2}{c^2}}} \tag{5.32}$$

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 5.5** 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*

5.3 Relativistic dynamics

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

Definition 5.6 (Relativistic energy) Taking into account mass-energy, the *relativistic energy* observes

$$E^2 = \underbrace{p^2c^2}_{\text{kinetic energy}} + \underbrace{m^2c^4}_{\text{rest energy}} \tag{5.33}$$

Remark 5.6 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:

$$\Lambda^\mu_\nu = \begin{pmatrix} \gamma & -\beta\gamma \\ -\beta\gamma & \gamma \end{pmatrix} \tag{5.34}$$

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

Definition 5.7 (4-momentum) The *4-momentum* is

$$P^\mu = (E, p^i) = (E/c, p_x, p_y, p_z) \tag{5.35}$$

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:

$$P' = (\gamma mc, \beta\gamma mc) \tag{5.36}$$

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

Theorem 5.4 (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.

5.4 Maxwell's equations

It is well-known that electromagnetism is a gauge field theory of the gauge (vector) field A^μ, also known as the 4-potential. It consists of the electric scalar potential and the magnetic vector potential[1]:

[1] We can also use an alternative 4-potential made up of the magnetic scalar potential and the electric vector potential. However, this is rarely used due to the absence of observed magnetic monopoles.

Definition 5.8 (4-potential)

$$A^\mu = (\phi, A_i) = (\phi, A_x, A_y, A_z) \tag{5.37}$$

As any student of field theory knows, one can construct the field strength tensor for any vector (gauge) field A^μ by

Definition 5.9 (Field strength tensor)

$$F^{\mu\nu} = i[D^\mu, D^\nu] = \partial^\mu A^\nu - \partial^\nu A^\mu - i[A_\mu, A_\nu] \tag{5.38}$$

In our case, the field is the 4-potential (5.37). However, before blindly making the substitution, we also note that electromagnetism is a gauge theory governed by U(1) symmetry. This gauge symmetry is so-called as it is defined with respect to the Lie group U(1)[2], which is Abelian. From any standard Lie theory text, we know that representations of abstract elements of an Abelian Lie group (in our case, the gauge fields themselves) commute:

$$[A_\mu, A_\nu] = 0 \tag{5.39}$$

Inserting this into (5.38) gives us the field strength tensor for electromagnetism, which is the well-known *Faraday tensor* or *electromagnetic (field) tensor*:

Definition 5.10 (Faraday tensor)

$$F^{\mu\nu} = \partial^\mu A^\nu - \partial^\nu A^\mu \tag{5.40}$$

Written in our familiar electric and magnetic 3-fields E and B, it reads

$$F^{\mu\nu} = \partial^\mu a^\nu - \partial_\nu a^\mu = \begin{pmatrix} 0 & E_x & E_y & E_z \\ -E_x & 0 & B_z & -B_y \\ -E_y & -B_z & 0 & B_x \\ -E_z & B_y & -B_x & 0 \end{pmatrix} \tag{5.41}$$

The definition of the electric and magnetic 3-fields then follow intuitively:

Definition 5.11 (Electric and magnetic fields)

$$E_i = F^{0i} = -\partial_i A^0 + \partial_0 A^i \quad B^i = -\frac{1}{2}\epsilon^{ijk} F_{jk} = \epsilon^{ijk}\partial_j A_k \tag{5.42}$$

As we recall from our previous studies, the equations of motion for electromagnetism are Maxwell's equations. With the toolkit we have developed in previous chapters, we can derive them using the standard method of constructing the Lagrangian and insert it into the Euler-Lagrange equations. The starting point is the Lagrangian arising from the Faraday tensor.

Definition 5.12 (EM field Lagrangian)

$$\mathcal{L} = -\frac{1}{4} F_{\mu\nu} F^{\mu\nu} \tag{5.43}$$

From this, one can write the stress-energy tensor as

Definition 5.13 (Stress-energy tensor in an EM field)

$$T_{\mu\nu} = \frac{1}{\mu_0}\left(F_{\mu\lambda} F^\lambda_\nu - \frac{1}{4}\eta_{\mu\nu} F_{\sigma\rho} F^{\sigma\rho} \right) \tag{5.44}$$

where, like c, μ_0 is generally regarded as 1.

Remark 5.7 Physically, an EM field-dominated stress-energy tensor corresponds to conserved neutral currents in the 4-directions.

Exercise 5.2 Show that the trace of the stress-energy tensor of the electromagnetic field is zero.

[2] We will investigate its gauge transformations very soon.

Derivation 5.4 (Recovery of Maxwell's equations) We can write down the full form of (5.43) by inserting (5.40):

$$\begin{aligned}\mathcal{L} &= -\frac{1}{4}(\partial^\mu A^\nu - \partial^\nu A^\mu)(\partial_\mu A_\nu - \partial_\nu A_\mu) \\ &= -\frac{1}{4}(\partial^\mu A^\nu \partial_\mu A_\nu - \partial^\nu A^\mu \partial_\mu A_\nu - \partial^\mu A^\nu \partial_\nu A_\mu + \partial^\nu A^\mu \partial_\nu A_\mu)\end{aligned} \tag{5.45}$$

Noting that all indices are free, we relabel some of them and find that

$$\mathcal{L} = \frac{1}{2}(\partial_\mu A_\nu)(\partial^\nu A^\mu) - \frac{1}{2}(\partial_\mu A_\nu)(\partial^\mu A^\nu) \tag{5.46}$$

Now let us take a *smol* detour and consider a 'total derivative' term of the form $\partial_\mu(A_\nu \partial^\mu A^\nu)$. As per the product rule, we immediately see that one can expand it as

$$\partial_\mu(A_\nu \partial^\mu A^\nu) = (\partial_\mu A_\nu)(\partial^\mu A^\nu) + A_\nu(\partial_\mu \partial^\mu A^\nu) \tag{5.47}$$

Using this relation, one can rewrite the two terms in (5.46) as

$$\frac{1}{2}(\partial_\mu A_\nu)(\partial^\nu A^\mu) = \frac{1}{2}\partial_\mu(A_\nu \partial^\nu A^\mu) - \frac{1}{2}A_\nu \partial^\nu \partial^\mu A_\mu \tag{5.48}$$

$$-\frac{1}{2}(\partial_\mu A_\nu)(\partial^\mu A^\nu) = \frac{1}{2}A_\nu \Box A^\nu - \frac{1}{2}\partial_\mu(A_\nu \partial^\mu A^\nu) \tag{5.49}$$

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

Definition 5.14 (D'Alembertian)

$$\Box = \partial^\mu \partial_\mu = \eta^{\mu\nu}\partial_\nu \partial_\mu = \underbrace{-\frac{\partial^2}{\partial t^2} + \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}}_{\text{in 4D spacetime with signature }(-,+,+,+)} \tag{5.50}$$

Fun fact 5.2 The reason why it is denoted as a square remains a historical mystery.

But at the same time, the term $\partial_\mu(A_\nu \partial^\mu A^\nu)$ looks a bit familiar, doesn't it? One can see that this so-called term is nothing but our good friend, the boundary term in (4.11). Recalling that zero terms and the boundary terms can be added or removed from the Lagrangian to our liking, (5.46) is then

$$\mathcal{L} = \frac{1}{2}A_\mu(\Box A^\mu - \partial^\mu \partial^\nu)A_\nu \tag{5.51}$$

By applying the Euler-Lagrange equations, we find that

$$\partial_\mu\left(\frac{\partial \mathcal{L}}{\partial(\partial_\mu A^\nu)}\right) - \frac{\partial \mathcal{L}}{\partial A^\nu} = 0 \tag{5.52}$$

where A^ν, which we ultimately recognise as a field variable, replaces ψ.

We then recover Maxwell's equations in index notation

Theorem 5.5 (Maxwell's equations)

$$\partial_\mu F^{\nu\mu} = J^\nu \tag{5.53}$$

Derivation 5.5 (Source and structure equations) By decomposing the various components of (??, we can write Maxwell's equations as the *structure equations* and *source equations*:

- Structure equations:

$$\partial_\mu F_{\nu\lambda} + \partial_\lambda F_{\mu\nu} + \partial_\nu F_{\lambda\mu} = 0 \tag{5.54}$$

- Source equations:

$$\partial_\nu F^{\mu\nu} = -4\pi J^\mu \tag{5.55}$$

Or, rewritten in classical physics notation:

- Structure equations:

$$\nabla \cdot \mathbf{B} = 0 \quad \nabla \times \mathbf{E} + \partial_t \mathbf{B} = 0 \tag{5.56}$$

- Source equations:

$$\nabla \cdot \mathbf{E} = 4\pi\rho \quad \nabla \times \mathbf{B} - \partial_t \mathbf{E} = \mathbf{J} \tag{5.57}$$

Exercise 5.3 Derive (5.56) and (5.57) by plugging the Faraday tensor into (5.54) and (5.57).

Now we investigate the gauge transformation brought about by our so-called U(1) symmetry. In gauge theory, there exist physical observables that are invariant under certain transformations of the potentials. In the case of electromagnetism, this manifests in the invariance of the Faraday tensor $F^{\nu\mu}$ (and hence, the EM fields E and B) under a set of 4-potential gauge transformations that is unsurprisingly called the U(1) transformations:

Definition 5.15 (U(1) transformations)

$$A_\mu(x) \to A_\mu(x) + \partial_\mu \Lambda(x) \tag{5.58}$$

where $\Lambda(x)$ is a scalar field and satisfies the wave equation.

The U(1) group has only a single generator, that being the 1×1 identity matrix $\mathbb{I}_1$. This sole generator has an 'eigenscalar' with only one component, from which electric charge arises.

Derivation 5.6 (Invarance of the Faraday tensor) One can verify the invariance of the Faraday tensor by noting that the following conditions simultaneously hold

$$\frac{\partial \mathcal{L}}{\partial(\partial_0 A_\mu)} = F^{\mu 0} \quad \frac{\partial \mathcal{L}}{\partial(\partial_0 A_\mu)} = 0 \tag{5.59}$$

where $F^{\mu 0}$ are the canonically conjugate EM fields.
Another way to see this is that, assuming zero 4-current, Maxwell's equations can be written as

$$\partial_\nu F^{\nu\mu} = \Box A^\mu - \partial^\mu(\partial_\nu A^\nu) = 0 \tag{5.60}$$

Immediately, we note that the equation does not depend on A^μ itself. Rather, the only dependence lies in $\partial_\nu A^\nu$.
Either way, we see that we can shift A_μ by any gradient $\partial_\mu \Lambda$, as we just did in our U(1) transformation, without affecting the Faraday tensor $F_{\mu\nu}$.

As it turns out, this is expected from the so-called *Noether's second theorem.*

Theorem 5.6 (Noether's second theorem) If a Lagrangian is invariant under an infinite-dimensional local symmetry group (like U(1) symmetry), there exist differential identities that the Euler-Lagrange equations are subject to, known as Noether identities, that make the equations to be not independent from each other. These Noether identities then reflect the redundancy in the equations of motion.

We can verify that such an identity exists for Maxwell's equations.

Derivation 5.7 (Noether identity in Maxwell's equations) The Lagrangian is invariant under our U(1) transformation, and it is possible to write the variation of the action as

$$\delta S = \int d^4x \partial_\mu \Lambda \cdot \left(\frac{\delta \mathcal{L}}{\delta A_\mu}\right) = \int d^4x \Lambda \cdot \partial_\mu \left(\frac{\delta \mathcal{L}}{\delta A_\mu}\right) = 0 \tag{5.61}$$

But since $\Lambda(x)$ is arbitrary, we know that, identically:

$$\partial_\mu \left(\frac{\delta \mathcal{L}}{\delta A_\mu}\right) = 0 \tag{5.62}$$

This is not immediately useful. Let us then take a detour and, for no reason whatsoever for the time being, introduce the conserved current, which emerges from the invariance of the Lagrangian under the global U(1) transformation (5.58).

Definition 5.16 (4-current)

$$J^\mu = (\rho, J_i) = (\rho, J_x, J_y, J_z) \tag{5.63}$$

This is nothing but our good friend, the so-called *4-current density vector* or simply the *4-current*, the last three components of which we have known since secondary school. Say hello to them!

> **Quote 5.6** Hello. My name is Inigo Montoya. You killed my father. Prepare to die.
>
> The Princess Bride, *1987*

From quantum electrodynamics, we have the relationship

$$J^\mu = -\frac{\partial \mathcal{L}}{\partial A_\mu} \tag{5.64}$$

This is a cute trick, but it quickly rises in significance. We recall that all this has been nothing but a giant detour from deriving the QED Noether identity. Recalling the tensorial form of Maxwell's equations (5.53), we can write

$$-\frac{\delta \mathcal{L}}{\delta A_\mu} = \partial_\nu F^{\nu\mu} \tag{5.65}$$

Now let us take a second 4-derivative on both sides. Recalling (5.62) which we derived earlier, one can conclude that the LHS is zero. This yields our Noether identity

Theorem 5.7 (Faraday tensor Noether identity)

$$\partial_\mu \partial_\nu F^{\nu\mu} = 0 \tag{5.66}$$

which holds identically due to the antisymmetry of $F^{\nu\mu}$.

The takeaway from this gauge invariance of $F^{\nu\mu}$ is that not all degrees of freedom in A_μ are physical. The unphysical degrees of freedom correspond to pure gauge and are thus called *gauge freedoms* which, if not eliminated, will lead us to mistakenly count multiple configurations of A^μs as distinct and give rise to erroneous results.

The standard procedure to eliminate gauge freedoms is *gauge fixing*. In classical EM, we attempt to eliminate this gauge freedom in (5.58) by the Lorentz-invariant Lorenz gauge.

> **Quote 5.7** Amazingly, the missing "t" is not a typo here.
>
> *Alessio Serafini*

Theorem 5.8 (Lorenz gauge)

$$\underbrace{\partial_\mu A^\mu = 0}_{\text{SR}} \quad \underbrace{\nabla_\mu A^\mu = 0}_{\text{GR}} \tag{5.67}$$

Under the Lorenz gauge, Maxwell's equations reduce to the d'Alembertian (wave) equation

$$\Box A^\mu = 0 \tag{5.68}$$

Chapter 6

Curvature

> **Quote 6.1** In [the] name of the metric, of the Riemann tensor, and the Weyl tensor.
>
> *Felix Halbwedl, 19 October 2025*

In this chapter, the geometric trinity of nonmetricity, torsion and curvature is finally introduced. After noting that only curvature survives in standard GR, we will investigate various curvature tensors and their properties, which will allow us to understand the geometrical mechanisms of curved spacetime.

6.1 Geometric trinity

> **Quote 6.2** The theory [Einstein] developed lends itself to an interpretation of the phenomena of gravity as the manifestation of the curvature of space-time. However, (...) there are three distinct and yet physically equivalent descriptions of gravity, which are rooted in the mathematical framework of metric-affine geometry. These formulations ascribe gravitational phenomena either to non-vanishing curvature, torsion, or non-metricity. These descriptions form the geometric trinity of General Relativity.
>
> *Lavinia Heisenberg, 2023*

At this point, most readers should know, either from popular science or other GR texts, that the central idea that distinguishes GR from classical mechanics and SR is the introduction of curvature and, with that, curved spacetime. Sadly, this is, as usual, a terrible simplification. There are, in fact, *three* so-called objects in GR that one can argue to carry intrinsic physical significance: nonmetricity, torsion and curvature. They are known in some literature in recent years as the so-called *geometric trinity*, and all of them conceptually relate to parallel transport.

Definition 6.1 (Parallel transport) Given a vector field X, a curve $\gamma(t)$ and a tangent vector $T_{\gamma(t)}$ to the curve, X is said to be parallel transported along $\gamma(t)$ if

$$\nabla_{T_{\gamma(t)}} X = 0 \tag{6.1}$$

In local coordinates, where we write the vector as V^μ and the tangent vector as T^μ, the condition reads

$$T^\mu \nabla_\mu V^\nu = 0 \tag{6.2}$$

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.

Definition 6.2 (Autoparallel) Parallel transport of tangent vectors is *autoparallel*. This gives rise to geodesics, on which tangent vectors are always parallel transported.

Definition 6.3 (Geometric trinity of GR)

- *Curvature* measures the failure of a vector to return to its original *orientation* after parallel transport.
- *Nonmetricity* measures the failure of parallel transport to preserve lengths and angles.
- *Torsion* measures the failure of multiple infinitesimal parallel transports to commute.

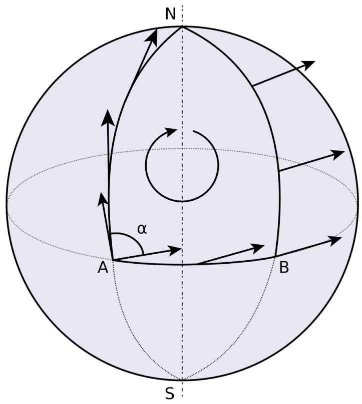


Figure 6.1: Tilting of the vector after parallel transport on a curved surface.

In standard GR, we are only concerned with curvature, while nonmetricity and torsion vanish. For this reason, curvature is studied far more by relativists than nonmetricty or torsion and thus deserves its own chapter. The goal of this section is hence twofold:

- Clarify how all three objects ultimately emerge from the Christoffels.
- Explain the physical meaning of nonmetricity and torsion and their demise in standard GR.

We first define the so-called *nonmetricity tensor*, which is the covariant derivative of the metric tensor:

Definition 6.4 (Nonmetricity tensor)

$$Q_{\mu\nu\lambda} = \nabla_\mu g_{\nu\lambda} \tag{6.3}$$

The nonmetricity tensor provides information on how the inner product of vectors changes under parallel transport. More specifically, nonmetricity is the measure of the 'deformation' of lengths and angles after undergoing parallel transport:

- A vector that is parallel transported can change its length. In a 4D spacetime, this means that the 'tick rate' of a clock can also change depending on the path on which it is parallel transported.
- Two vectors that start orthogonal may become non-orthogonal after parallel transport.

For example, let v^μ be parallel transported along a curve with tangent u^μ. Then, even though

$$\nabla_u v^\mu = 0 \tag{6.4}$$

its squared length changes as

$$\frac{d}{d\lambda}(g_{\mu\nu}v^\mu v^\nu) = Q_{\lambda\mu\nu}u^\lambda v^\mu v^\nu \tag{6.5}$$

Sadly, nonmetricity always vanishes in *Riemannian geometry*, which is what standard GR operates in. This is known as *metric compatibility*:

Definition 6.5 (Metric compatitibility) A covariant derivative ∇ is known to be *metric compatible* if its operation on the metric yields zero:

$$\nabla_\lambda g_{\mu\nu} = 0 \tag{6.6}$$

This ensures that no deformation of lengths and angles happens after parallel transport.

The next object is the *torsion tensor*:

Definition 6.6 (Torsion tensor) Given any vectors $X, Y \in TM$ and $\omega \in T^*M$, the so-called torsion tensor T is an object defined in relation to the covariant derivative:

$$T(X, Y, \omega) = \omega(\nabla_X Y - \nabla_Y X - [X, Y]) \tag{6.7}$$

where $[X, Y]$ is the Lie bracket.
In a local coordinate frame, this is simply the twice antisymmetrisation of the Christoffel symbol:

$$T^\lambda_{\mu\nu} = \Gamma^\lambda_{\mu\nu} - \Gamma^\lambda_{\nu\mu} = 2\Gamma^\lambda_{[\mu\nu]} \tag{6.8}$$

where T is antisymmetric in its lower indices.

Although sometimes shrouded in an air of mystery, the physical meaning of the torsion tensor can be quite easily understood. To quote a great relativist:

> **Quote 6.3** The torsion tensor measures the failure of a parallelogram to close.
>
> ---
>
> *Christian G. Böhmer*

Definition 6.7 (Torsion-free) A connection which is symmetric in its lower indices is known as 'torsion-free'.

Remark 6.1 In standard GR, we assume manifolds to be torsion-free. This explains why all Christoffel symbols we see in GR are symmetric in their lower indices. Let us discuss what this means. Suppose two infinitesimal displacement vectors $u, v \in M$, we can construct the two infinitesimal halves of a parallelogram. Starting from some point $p \in M$:

1. Move in direction u by a *smol* parameter ϵ. Then, move in direction v by ϵ. This is one half.
2. Move in direction v by a *smol* parameter ϵ. Then, move in direction u by ϵ. This is the other half.

In Riemannian geometry, which is the case of standard GR, M has no torsion, and the endpoints of these two paths would coincide up to order ϵ^2. Alternatively, if torsion is non-zero, these endpoints differ. To see this mathematically, we calculate the displacement from the original point p when one travels through each half of the parallelogram.

Derivation 6.1 (Effects of torsion) In the first half, we first parallel transport by u^μ, and then by v^μ. The strategy is as follows:

- First parallely transport the vector v^μ by an infinitesimal distance ϵ along u^μ. At this point, the distance from p is nothing but ϵu^ρ.
- However, the first transportation has distorted v^μ, so we cannot blindly add ϵv^μ. During the first transportation, v^μ is changed by δv^μ, and the post-transformation v'^μ is

$$v'^\rho = v^\rho + \delta v^\rho = v^\rho - \epsilon \Gamma^\rho_{\mu\nu} u^\mu v^\nu \tag{6.9}$$

- This, coupled by the *smol* ϵ, is what we have to actually add to u^μ. So the total displacement is

$$x^\rho_{(u\to v)} = \epsilon u^\rho + \epsilon v'^\rho = \epsilon u^\rho + \epsilon v^\rho + \epsilon^2 \Gamma^\rho{}_{\mu\nu} u^\mu v^\nu \tag{6.10}$$

In the second half, the strategy is similar, but with v^μ and u^μ swapped:

- First parallel transport the vector u^μ by an infinitesimal distance ϵ along v^μ. The distance so far is ϵv^ρ.
- The first change has distorted u^μ, resulting in

$$u'^\rho = u^\rho + \delta u^\rho = u^\rho - \epsilon \Gamma^\rho_{\mu\nu} v^\mu u^\nu \tag{6.11}$$

- We now couple this by the *smol* ϵ and add it to our v^μ. The total displacement reads

$$x^{\rho}_{(v\to u)} = \epsilon v^\rho + \epsilon u'^\rho = \epsilon v^\rho + \epsilon u^\rho + \epsilon^2 \Gamma^\rho{}_{\mu\nu} v^\mu u^\nu \tag{6.12}$$

The difference between the endpoints of transporting along each of the paths is then the failure of the parallelogram to close:

$$\begin{aligned} \Delta x^\rho &= \epsilon^2 (\Gamma^\rho_{\nu\mu} - \Gamma^\rho_{\mu\nu}) u^\mu v^\nu \\ &= \epsilon^2 T^\rho{}_{\mu\nu} u^\mu v^\nu \end{aligned} \tag{6.13}$$

Physically, it represents the twisting of the geometrical structure.

A final note then concerns the relationship between the nonmetricity and torsion tensors and the Christoffels. First, let us build two rank-3 objects from the torsion and nonmetricity tensors by permuting their indices and halving the result. For the nonmetricity tensor, this yields the *disformation tensor*[1]:

Definition 6.8 (Disformation tensor)

$$L_{\mu\nu\lambda} = \frac{1}{2}(Q_{\mu\nu\lambda} + Q_{\nu\lambda\mu} + Q_{\lambda\mu\nu}) \tag{6.14}$$

For the torsion tensor, this yields the *contorsion tensor*:

Definition 6.9 (Contorsion tensor)

$$K_{\mu\nu\lambda} = \frac{1}{2}(T_{\mu\nu\lambda} + T_{\nu\lambda\mu} + T_{\lambda\mu\nu}) \tag{6.15}$$

Remark 6.2 The convention is to write both tensors in all lower indices when defining them.
At first glance, one might be tempted to laugh at them, as the objects look funny and even sound funny. However, we note that the structure of their definitions looks very much like that of our beloved Christoffels in (3.70). So, the writing is now on the wall, and the question is, do they have anything to do with the Christoffels? As it turns out, the answer is yes.

Definition 6.10 (Levi-Civita connection) Any general affine connection can be decomposed uniquely as:

$$\tilde{\Gamma}^\lambda_{\mu\nu} = \Gamma^\lambda_{\mu\nu} + K^\lambda{}_{\mu\nu} + L^\lambda{}_{\mu\nu} \tag{6.16}$$

where $\Gamma^\lambda_{\mu\nu}$, the part of the Christoffel without any contribution from nonmetricity or torsion, is the so-called *Levi-Civita connection*[a].

[a]Despite their name, it has nothing to do with the Levi-Civita symbol.

Remark 6.3 In other words, the Christoffels can be called a Levi-Civita connection when it is both metric compatible and torsion-free. Hence, all Christoffels in GR are Levi-Civita connections. As such, we regrettably have to say goodbye for now to our dear friends, the nonmetricity and torsion tensors.
With the connection out of the way, we can now conclude this section and move on to... Wait, what is that, Arthropodus? You want to make some comments? But must we... Okay! Okay! You win. Go ahead! See if I care.

Trilobite 6.1

While both the nonmetricity and torsion tensors vanish in standard GR, the nontrivialness of nonmetricity and/or torsion is a staple in modified theories of gravity. For example, Einstein-Cartan theory assumes non-zero torsion emerging from spin. The assumption of vanishing nonmetricity and torsion is not without good reason, however, as even if they do not vanish, they contribute minimally in most cases to the extent that in the real-life limit, they make no discernible difference in physical predictions.

For the love of the metric, can't you at least wait until Part II? Some animals!

[1]This object often appears in literature without its own name.

Quote 6.4 We all are animals.

Thomas Hutter, 18 November 2025

Okay, all of this has been rather gruelling, but we can now discuss the most interesting implication of nonmetricity and torsion, no thanks to our good friend Arthropodus. As we saw in the last two sections, both nonmetricity and torsion are quantities that possess an inherent geometrical meaning, and both contribute to the Christoffels
Yet there is a third contribution, or rather the *only* contribution in standard GR, which is the Levi-Civita connection. The question then is, what is the object (other than the metric itself) associated with the Levi-Civita connection? To begin with, it is easy to prove that, in the (flat) Minkowski metric (5.2), the Christoffels are zero.

Exercise 6.1 Do this now.

The Minkowski metric is used for physical scenarios ignoring GR like special relativity and quantum field theory. Thus, in classical physics and even in SR, spacetime is completely flat with no nonmetricity, no torsion, and no Levi-Civita connection. This is why we have been able to do all of classical physics without doing any differential geometry whatsoever. Alternatively, in GR, which, as we know from popular science, lives in curved spacetime, the Levi-Civita connection becomes non-trivial. It is then intuitive to see that the object associated with the Levi-Civita connection *must* be curvature.

6.2 Riemann tensor, Ricci tensor and scalar

Now we can finally investigate the curvature tensors, which lie at 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.

Definition 6.11 (Riemann tensor) For $X, Y, Z \in TM$ and $\omega \in T*M$, the Riemann tensor is a $(1,3)$ tensor given by

$$R(X,Y,Z,\omega) = \omega(-\nabla_X(\nabla_Y Z) + \nabla_Y(\nabla_X Z) + \nabla_{[X,Y]}Z) \tag{6.17}$$

In local coordinates, it is defined by a series of Christoffel symbols:

$$R^{\sigma}_{\mu\nu\lambda} = \Gamma^{\sigma}_{\mu\lambda,\nu} - \Gamma^{\sigma}_{\nu\lambda,\mu} + \Gamma^{\sigma}_{\nu\sigma}\Gamma^{\rho}_{\rho\nu} - \Gamma^{\rho}_{\mu\sigma}\Gamma^{\sigma}_{\rho\mu} \tag{6.18}$$

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 6.1 (Riemann tensor properties) The Riemann tensor has the following properties:

1. **Antisymmetry I:**
$$R_{\mu\nu\lambda\sigma} = -R_{\nu\mu\lambda\sigma} \tag{6.19}$$

2. **Antisymmetry II:**
$$R_{\mu\nu\lambda\sigma} = -R_{\mu\nu\sigma\lambda} \tag{6.20}$$

3. **Cyclic identity:**
$$R_{\sigma\mu\nu\lambda} + R_{\lambda\sigma\mu\nu} + R_{\sigma\nu\lambda\mu} = 0 \tag{6.21}$$
This is the (in)famous *first Bianchi identity*.

4. **Symmetry:**
$$R_{\mu\nu\lambda\sigma} = R_{\lambda\sigma\mu\nu} \tag{6.22}$$

5. **Kretschmann scalar:** An invariant quality is the *Kretschmann scalar* K, defined by
$$K = R^{\mu\nu\lambda\sigma}R_{\mu\nu\lambda\sigma} \tag{6.23}$$

Remark 6.4 From (6.19), (6.20) and (6.21), we can prove (6.22) by permuting the indices of (6.21) four times and adding them together.

Exercise 6.2 Do it.

Remark 6.5 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 6.12 (Ricci tensor)

$$R_{\mu\nu} = R^{\lambda}_{\mu\lambda\nu} \tag{6.24}$$

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

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

Definition 6.13 (Ricci scalar)

$$R = g^{\mu\nu} R_{\mu\nu} \tag{6.25}$$

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 6.6 Despite their definitions, it is more convenient to solve for the Ricci tensor and the Ricci scalar by using the metric, like

$$R_{\mu\nu} = g_{\mu\mu} g^{\sigma\lambda} R^{\mu}_{\nu\sigma\lambda} \quad R = g^{\mu\nu} R_{\mu\nu} \tag{6.26}$$

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 can 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 6.2: 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 6.2 (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:

$$R_{\mu\nu\lambda\sigma} = K(g_{\mu\lambda}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\lambda}) \tag{6.27}$$

Remark 6.7 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

$$R_{\mu\nu} = \frac{1}{2}Kg_{\mu\nu} \tag{6.28}$$

Remark 6.8 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.

6.3 Weyl tensor and its demise

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 6.2 (Ricci decomposition)

$$R_{\mu\nu\lambda\sigma} = S_{\mu\nu\lambda\sigma} + E_{\mu\nu\lambda\sigma} + C_{\mu\nu\lambda\sigma} \tag{6.29}$$

The first 2 terms are objects with no name defined as

$$S_{\mu\nu\lambda\sigma} = \frac{R}{n(n-1)}(g_{\mu\sigma}g_{\nu\lambda} - g_{\mu\lambda}g_{\nu\sigma}) \tag{6.30}$$

$$E_{\mu\nu\lambda\sigma} = \frac{1}{n-2}(Z_{\mu\sigma}g_{\nu\lambda} - Z_{\nu\sigma}g_{\mu\lambda} - Z_{\mu\lambda}g_{\nu\sigma} + Z_{\nu\lambda}g_{\mu\sigma}) \tag{6.31}$$

where, for a manifold of dimension n:

$$Z_{\mu\nu} = R_{\mu\nu} - \frac{1}{n}Rg_{\mu\nu} \tag{6.32}$$

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

Definition 6.14 (Weyl tensor)

$$C_{\mu\nu\lambda\sigma} = R_{\mu\nu\lambda\sigma} + \frac{1}{n-2}(R_{\mu\sigma}g_{\nu\lambda} - R_{\mu\lambda}g_{\nu\sigma} + R_{\nu\lambda}g_{\mu\sigma} - R_{\nu\sigma}g_{\mu\lambda}) + \frac{1}{(n-1)(n-2)}R(g_{\mu\lambda}g_{\nu\sigma} - g_{\mu\sigma}g_{\nu\lambda}) \tag{6.33}$$

where n is the dimension of the manifold. One can simplify this using the index commutator seen in (3.28):

$$C_{\mu\nu\lambda\sigma} = R_{\mu\nu\lambda\sigma} - \frac{2}{n-2}\left(g_{\mu[\lambda}R_{\sigma]\nu} - g_{\nu[\lambda}R_{\sigma]\mu}\right) + \frac{2}{(n-1)(n-2)}Rg_{\mu[\lambda}g_{\sigma]\nu} \tag{6.34}$$

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

Quote 6.5 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 6.9 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 6.3: 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 6.15 (Schouten tensor)

$$P_{\mu\nu} = \frac{1}{n-2}\left(R_{\mu\nu} - \frac{R}{2(n-1)}g_{\mu\nu}\right) \tag{6.35}$$

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 6.16 (Bach tensor)

$$B_{\mu\nu} = P_{\lambda\sigma}W^{\lambda\sigma}_{\mu\nu} + \nabla^\lambda\nabla_\lambda P_{\mu\nu} - \nabla^\lambda\nabla_\mu P_{\nu\lambda} \tag{6.36}$$

[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.

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

6.4 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 6.3 (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

$$N^\mu = \frac{dX^\mu}{ds} \tag{6.37}$$

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

$$\frac{d^2X^\mu}{d\lambda ds} = \frac{d^2X^\mu}{dsd\lambda} \tag{6.38}$$

we can show that

$$N^\mu \frac{\partial T^\mu}{\partial x^\mu} = T^\mu \frac{\partial N^\mu}{\partial x^\mu} \tag{6.39}$$

and

$$N^\mu \nabla_\mu T^\nu = T^\mu \nabla_\mu N^\nu \tag{6.40}$$

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 (6.40):

$$T^\lambda \nabla_\lambda \left(T^\mu \nabla_\mu N^\nu\right) = T^\lambda \nabla_\lambda \left(N^\mu \nabla_\mu T^\nu\right) \tag{6.41}$$

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:

$$\begin{aligned} T^\lambda \nabla_\lambda \left(T^\mu \nabla_\mu N^\nu\right) &= T^\lambda \left(\nabla_\lambda N^\mu\right) \nabla_\mu T^\nu + T^\lambda N^\mu \nabla_\lambda \nabla_\mu T^\nu \\ &= \underbrace{T^\lambda \left(\nabla_\lambda N^\mu\right) \nabla_\mu T^\nu + T^\lambda N^\mu \nabla_\mu \nabla_\lambda T^\mu}_{\textcircled{1}} - T^\lambda N^\mu R^\lambda_{\nu\mu\sigma} T^\sigma \end{aligned} \tag{6.42}$$

We can show that ① vanishes:

$$\begin{aligned} T^\lambda \left(\nabla_\lambda N^\mu\right) \nabla_\mu T^\nu + T^\lambda N^\mu \nabla_\mu \nabla_\lambda T^\nu &= N^\lambda \left(\nabla_\lambda T^\mu\right) \nabla_\mu T^\mu + T^\lambda N^\mu \nabla_\mu \nabla_\lambda T^\nu \\ &= N^\lambda \left(\nabla_\lambda T^\mu\right) \nabla_\mu T^\nu + T^\mu N^\lambda \nabla_\lambda \nabla_\mu T^\nu \\ &= N^\lambda \left(\nabla_\lambda T^\mu \nabla_\mu T^\nu + T^\mu \nabla_\lambda \nabla_\mu T^a\right) \\ &= N^\lambda \nabla_\lambda \left(T^\mu \nabla_\mu T^\nu\right) \\ &= 0 \end{aligned} \tag{6.43}$$

which yields the final result

$$T^\lambda \nabla_\lambda \left(T^\mu \nabla_\mu N^\nu\right) = -T^\lambda N^\mu R^\nu_{\lambda\mu\sigma} T^\sigma = \left(R^\nu_{\mu\lambda\sigma} T^\lambda T^\sigma\right) N^\mu \tag{6.44}$$

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

$$\frac{D^2 N^\mu}{D\lambda^2} = \left(R^\nu_{\mu\lambda\sigma} T^\lambda T^\sigma\right) N^\mu \tag{6.45}$$

where, importantly, T observes

$$T_\mu T^\mu = 1 \quad g_{\mu\nu} T^\mu N^\nu = 0 \tag{6.46}$$

Remark 6.10 $T_\mu T^\mu = 1$ 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.

Quote 6.6 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 6.7 Those fishies are doing such a great job at tiling the hyperbolic plane.

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

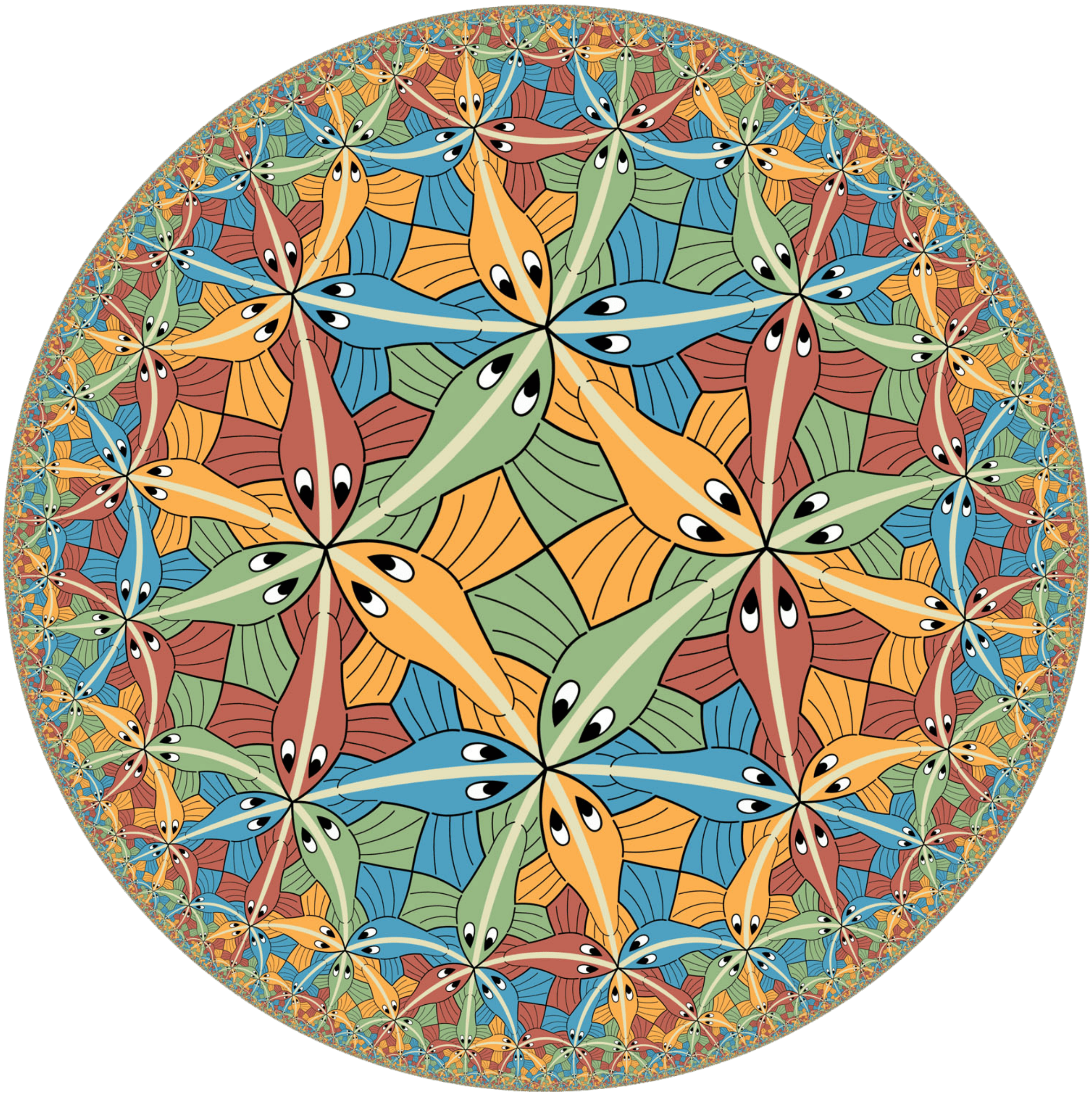


Figure 6.4: *Circle Limit III* (M. C. Escher, 1959)

Chapter 7

Field equations and gravitational waves

Quote 7.1 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 gravitational waves, which are small perturbations to an otherwise flat metric. Some more comments regarding the field theory approach are then made.

7.1 Einstein-Hilbert action

Quote 7.2 Imagine a modern version of Einstein who would have learned all the standard techniques of field theory description. We would have eventually constructed GR from the field theory perspective, just years later.

Lavinia Heisenberg, March 2019

We recall the action in terms of the Lagrangian density in (4.7). In GR, one can decompose the action into two terms:

Definition 7.1 (Einstein-Hilbert action)

$$S = \int_{\mathcal{V}} \mathcal{L} d^4x = S_H + S_M \tag{7.1}$$

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

Remark 7.1 An important plot twist, already foreshadowed, reveals itself here: (4.47) is a very limiting case (applying to only free particles).
The simplest gravitational action was proposed by Hilbert and Einstein in 1915:

Definition 7.2 (Hilbert term) The *Hilbert term* S_H is defined as

$$S_H = \frac{1}{16\pi} \int_{\mathcal{V}} R\sqrt{-g} d^4x \tag{7.2}$$

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

$$\mathcal{L}_H = R\sqrt{-g} \tag{7.3}$$

The only other contribution is the matter field contribution:

Definition 7.3 (Matter action) The *matter action* S_M is defined as

$$S_M = \int_{\mathcal{V}} \mathcal{L}_M(\phi, \partial_\mu \phi, g_{\mu\nu})\sqrt{-g}d^4x \tag{7.4}$$

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

Remark 7.2 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

$$\delta\psi = \left.\frac{d\psi_\lambda}{d\lambda}\right|_{\lambda=0} \tag{7.5}$$

where λ is a parameter, and we demand that

$$\delta\psi|_{\partial\mathcal{V}} = 0 \tag{7.6}$$

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

Derivation 7.1 (Proto-field equations) We begin with the Hilbert term. The variation of the Hilbert Lagrangian is

$$16\pi\delta\mathcal{L}_H = -\frac{\delta g}{2\sqrt{-g}}g^{\mu\nu}R_{\mu\nu} + (\delta g^{\mu\nu}R_{\mu\nu} + g^{\mu\nu}\delta R_{\mu\nu})\sqrt{-g} \tag{7.7}$$

We can use *Jacobi's formula*:

Theorem 7.1 (Jacobi's formula) For a matrix A,

$$\frac{d}{dt}\det A(t) = \operatorname{tr}\left(\operatorname{adj}(A(t))\frac{dA(t)}{dt}\right) = (\det A(t))\cdot\operatorname{tr}\left(A(t)^{-1}...\frac{dA(t)}{dt}\right) \tag{7.8}$$

By using this on the metric, one yields

$$\delta g = gg^{\mu\nu}\delta g_{\mu\nu} = -gg_{\mu\nu}\delta g^{\mu\nu} \tag{7.9}$$

Plugging this into the Hilbert term:

$$16\pi\delta\mathcal{L}_H = \left[\left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R\right)\delta g^{\mu\nu} + g^{\mu\nu}\delta R_{\mu\nu}\right]\sqrt{-g} \tag{7.10}$$

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

Theorem 7.2 (Paltini identity)

$$\delta R_{\mu\nu} = \nabla_\rho\delta\Gamma^\rho_{\nu\mu} - \nabla_\nu\delta\Gamma^\rho_{\rho\mu} \tag{7.11}$$

We then define a contravariant vector

$$\delta V^\rho \doteq g^{\mu\nu}\delta\Gamma^\rho_{\mu\nu} - g^{\rho\nu}\delta\Gamma^\mu_{\mu\nu} \tag{7.12}$$

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

$$\sqrt{-g}g^{\mu\nu}\delta R_{\mu\nu} = \partial_\rho(\sqrt{-g}V^\rho) \tag{7.13}$$

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

$$\delta S_H = \frac{1}{16\pi}\int_\Sigma\left(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R\right)\sqrt{-g}\delta g^{\mu\nu}d^4x + \frac{1}{16\pi}\oint_{\partial\nu} V^\mu d\sigma_\mu d^3x \tag{7.14}$$

where $d\sigma_\mu$ is the oriented volume element of the hypersurface $\partial\Sigma$.
The variation of the matter action is much easier:

$$\delta S_M = \int_\Sigma\left[\frac{\partial\mathcal{L}_M}{\delta g^{\mu\nu}} - \frac{1}{2}\mathcal{L}_M g_{\mu\nu}\right]\sqrt{-g}\delta g^{\mu\nu}d^4x \tag{7.15}$$

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

$$T_{\mu\nu} \doteq -2\frac{\delta \mathcal{L}_M}{\delta g^{\mu\nu}} + \mathcal{L}_M g_{\mu\nu} \tag{7.16}$$

The total variation of the action hence becomes

$$\delta S = \frac{1}{16\pi}\int_\sigma \left[R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi T_{\mu\nu}\right]\sqrt{-g}\delta g^{\mu\nu} d^4x \tag{7.17}$$

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

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi T_{\mu\nu} = \kappa T_{\mu\nu} \tag{7.18}$$

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.

7.2 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:

$$F = \begin{cases} \frac{GMm}{r^2} & \text{electromagnetism} \\ \frac{Q_1 Q_2}{kr^2} & \text{gravitation} \end{cases} \tag{7.19}$$

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:

$$a = \begin{cases} \frac{GM}{r^2} & \text{electromagnetism} \\ \frac{Q_1 Q_2}{kmr^2} & \text{gravitation} \end{cases} \tag{7.20}$$

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 7.3 Gravity is geometry

Various relativists

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

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}g_{\mu\nu}h \tag{7.21}$$

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

Theorem 7.3 (Trace reverse property)

$$g^{\mu\nu}\bar{h}_{\mu\nu} = \bar{h} = h - 2h = -h \tag{7.22}$$

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

Definition 7.5 (Einstein tensor)

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R \tag{7.23}$$

> **Quote 7.4** 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 7.4 (Einstein tensor properties) The Einstein tensor satisfies the following properties:

- **Symmetry:**
$$G_{\mu\nu} = G_{\nu\mu} \tag{7.24}$$
- **Contracted Bianchi identity:** The Einstein tensor has zero covariant divergence
$$\nabla^{\mu} G_{\mu\nu} = 0 \tag{7.25}$$
- **Trace:**
$$G = g^{\mu\nu} G_{\mu\nu} = -R \tag{7.26}$$
  This follows from (7.22).

We can now finally assemble the field equations.
Fun fact 7.1 (A bit of history) Einstein's original guess of the field equations in 1915 was

Theorem 7.5 (1915 guess of the field equations)

$$G_{\mu\nu} = \kappa T_{\mu\nu} \tag{7.27}$$

Remark 7.3 What could possibly go wrong?

7.3 Variants of the field equations

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

Theorem 7.6 (Field equations with the cosmological constant)

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = \kappa T_{\mu\nu} \tag{7.28}$$

where Λ is the cosmological constant.

Remark 7.4 We can likewise derive this from the Einstein-Hilbert action, whose Hilbert term is amended as

$$S_H = \frac{1}{16\pi} \int_{\mathcal{V}} (R - 2\Lambda)\sqrt{-g}\, d^4x \tag{7.29}$$

Note 7.1 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 7.7 (Vacuum field equations)

$$R_{\mu\nu} = 0 \tag{7.30}$$

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 (5.44) and replace the Minkowski metric $\eta_{\mu\nu}$ with a general metric $g_{\mu\nu}$. The stress-energy tensor then becomes

$$T_{\mu\nu} = F_{\mu\lambda}F^{\lambda}_{\nu} - \frac{1}{4}g_{\mu\nu}F_{\sigma\rho}F^{\sigma\rho} \tag{7.31}$$

The field equations hence become

Theorem 7.8 (Einstein-Maxwell equation)

$$G_{\mu\nu} = R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = \kappa\left(F_{\mu\lambda}F^{\lambda}_{\nu} - \frac{1}{4}g_{\mu\nu}F_{\sigma\rho}F^{\sigma\rho}\right) \tag{7.32}$$

7.4 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

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} + \mathcal{O}((h_{\mu\nu})^2) \quad \text{where} \quad |h_{\mu\nu}| \ll 1 \tag{7.33}$$

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

Remark 7.5 As both the metric $g_{\mu\nu}$ and the Minkowski metric $\eta_{\mu\nu}$ are symmetric, $h_{\mu\nu}$ is also symmetric.

Remark 7.6 One can also derive the inverse metric in linearised gravity as

$$g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu} \tag{7.34}$$

Derivation 7.2 (Linearised field equations) We begin by writing out the Christoffels under the linearised gravity regime.

$$\Gamma^{\mu}_{\alpha\beta} = \frac{1}{2}\eta^{\mu\nu}(h_{\alpha\mu,\beta} + h_{\beta\nu,\alpha} - h_{\alpha\beta,\nu}) = \frac{1}{2}(h^{\mu}_{\alpha,\beta} + h^{\mu}_{\beta,\alpha} - h_{\alpha\beta}{}^{,\mu}) \tag{7.35}$$

Remark 7.7 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

$$R_{\mu\nu} = \frac{1}{2}(h^{\alpha}_{\mu,\nu\alpha} + h^{\alpha}_{\nu,\mu\alpha} - h_{\mu\nu,\alpha}^{,\alpha} - h_{,\mu\nu}) \tag{7.36}$$

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

$$\bar{h}_{\mu\nu} = h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}h \tag{7.37}$$

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

$$\bar{h}^{\mu\alpha}{}_{,\alpha} = 0 \tag{7.38}$$

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

Theorem 7.9 (Linearised field equations)

$$G_{\mu\nu} = -\frac{1}{2}\Box\bar{h}_{\mu\nu} = -\frac{1}{2}\bar{h}_{\mu\nu}{}^{;\alpha}_{,\alpha} = \kappa T_{\mu\nu} \tag{7.39}$$

where $\Box$ is the d'Alembertian.

Derivation 7.3 (Plane wave solutions) We now return to the linearised field equations as seen in Theorem 7.9. As gravitational waves traverse through regions effectively devoid of matter, we also set the stress-energy tensor (i.e. matter content) as zero.

Theorem 7.10 (Linearised vacuum field equations)

$$\bar{h}_{\mu\nu}{}^{;\alpha}_{,\alpha} = 0 \tag{7.40}$$

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

$$\bar{h}_{\mu\nu} = A_{\mu\nu}\exp(ik_\alpha x^\alpha) \tag{7.41}$$

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

$$\bar{h}_{\mu\nu} = A_{\mu\nu}\cos(ik_\alpha x^\alpha) \tag{7.42}$$

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:

$$\underbrace{k_\alpha k^\alpha = 0}_{k_\alpha \text{ is a null vector}} \qquad \underbrace{A_{\mu\alpha}k^\alpha = 0}_{A_{\mu\alpha} \text{ is orthogonal to } k_\alpha} \tag{7.43}$$

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 7.6 (Gravitational wave)

$$\bar{h}_{\mu\nu} = A_{\mu\nu}\cos(\omega x - \omega t) \tag{7.44}$$

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

Derivation 7.4 (Transverse-traceless (TT) gauge) For a given 4-velocity u_ν, we impose the following gauge conditions

$$\underbrace{A^{0\nu} = 0 \to A^{\mu\nu}u_\nu = 0}_{\text{transverse wave}} \qquad \underbrace{A^\mu_\mu = 0}_{\text{traceless wave amplitude}} \tag{7.45}$$

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

$$A^{11} + A^{22} + A^{33} = 0 \tag{7.46}$$

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:

$$A^{\mu\nu} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & A_+ & A_\times & 0 \\ 0 & A_\times & -A_+ & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix} \tag{7.47}$$

$A_+ = 0$ 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

  $$x' = (x+y)/\sqrt{2} \quad y' = (x-y)/\sqrt{2} \tag{7.48}$$

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

Remark 7.8 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.

7.5 Qravity! Qravity! Qravity! [EMPTY]

Chapter 8

Schwarzschild solution

Quote 8.1 Every introductory course to GR should end with the Schwarzschild solution.

Christian G. Böhmer, 2025

8.1 Deriving the Schwarzschild solution

Definition 8.1 (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 8.1 The Schwarzschild metric describes a non-rotating, uncharged black hole. We will discuss the other three cases in the next chapter.

Figure 8.1: Faking his death after discovering the eternal tensor of youth $Y_{\mu\nu}$, Schwarzschild mangled his name and became a psychiatrist.

Derivation 8.1 (Schwarzschild solution) In the Minkowski space $X^\mu = (t, r, \theta, \phi)$, we first assume such a Lagrangian (and thus the metric):

$$L = -e^{\nu}\dot{t}^2 + e^{\lambda}\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2\theta\dot{\phi}^2 \tag{8.1}$$

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

$$\begin{aligned}
\Gamma^c_{tc} &= 0\\
\Gamma^c_{rc} &= \Gamma^t_{rt} + \Gamma^r_{rr} + \Gamma^\theta_{r\theta} + \Gamma^\phi_{r\phi} = \frac{2}{r} + \frac{1}{2}\left(\nu' + \lambda'\right)\\
\Gamma^c_{\theta c} &= \cot\theta\\
\Gamma^c_{\phi c} &= 0
\end{aligned} \tag{8.2}$$

and by that, the Ricci tensor components

$$\begin{aligned}
R_{tt} &= e^{\nu-\lambda}\left[\frac{1}{2}\nu'' + \frac{1}{4}\left(\nu'\right)^2 + \frac{1}{r}\nu' - \frac{1}{4}\nu'\lambda'\right]\\
R_{rr} &= -\frac{1}{2}\nu'' - \frac{1}{4}\left(\nu'\right)^2 + \frac{1}{4}\nu'\lambda' + \frac{1}{r}\lambda'\\
R_{\theta\theta} &= 1 - e^{-\lambda} + \frac{1}{2}r\lambda' e^{-\lambda} - \frac{1}{2}r\nu' e^{-\lambda}\\
R_{\phi\phi} &= \sin^2\theta R_{\theta\theta}
\end{aligned} \tag{8.3}$$

Due to the vacuum nature of Schwarzschild solutions, we can consider the vacuum field equations, as seen in (7.30). By simply plugging in the Ricci tensor, we find that

$$e^{\nu} = e^{-\lambda} = 1 - \frac{\mathcal{C}}{r} \tag{8.4}$$

Remark 8.2 The gravitational equations imply that the Ricci tensor reduces to zero. This is also called the *source-free Newtonian field equations.*
The metric becomes

$$ds^2 = -\left(1 - \frac{\mathcal{C}}{r}\right)dt^2 + \left(1 - \frac{\mathcal{C}}{r}\right)^{-1}dr^2 + r^2 d\Omega^2 \tag{8.5}$$

where we have the common simplified notation for the angular terms

$$d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2 \tag{8.6}$$

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

$$\lim_{c^2\to\infty}\Gamma^r_{tt} = \frac{G\mathcal{C}}{2r^2} \tag{8.7}$$

Compare this with the Newtonian gravitational potential

$$\nabla\Phi = \frac{GM}{r^2} \tag{8.8}$$

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

Definition 8.2 (Schwarzschild metric) As such, we arrive at the *Schwarzschild metric*:

$$ds^2 = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1}dr^2 + r^2 d\Omega^2 \tag{8.9}$$

Remark 8.3 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.

8.2 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 8.2 (Schwarzschild geodesic equations) Starting with the equation for θ:

$$\frac{d}{d\lambda}(2r^2\dot{\theta}) = 2r^2\ddot{\theta} + 4r\dot{r}\dot{\theta} = 2r^2 \sin\theta\cos\theta\dot{\phi}^2 \tag{8.10}$$

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:

$$\frac{d}{d\lambda}(-2f(r)\dot{t}) = 0 \rightarrow f(r)\dot{t} = E \tag{8.11}$$

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

$$\frac{d}{d\lambda}(2r^2\sin^2\theta\dot{\phi}) = 0 \rightarrow r^2\dot{\phi} = 2\ell \tag{8.12}$$

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:

$$Lf(r) = -E^2 + \dot{r}^2 + \frac{\ell^2}{r^2}f(r) \tag{8.13}$$

Hence

$$E^2 = \dot{r}^2 + \left(1 - \frac{2m}{r}\right)\left(\frac{\ell^2}{r^2} - L\right) \tag{8.14}$$

Dividing both sides by 2, we find that

$$\frac{1}{2}E^2 = \underbrace{\frac{1}{2}\dot{r}^2}_{\text{kinetic energy}} + \underbrace{\frac{1}{2}\left(1 - \frac{2M}{r}\right)\left(\frac{\ell^2}{r^2} - L\right)}_{\text{potential energy}} \tag{8.15}$$

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 8.3 (Effective potential)

$$V_{\text{eff}}(r) = \frac{1}{2}\left(1 - \frac{2M}{r}\right)\left(\frac{\ell^2}{r^2} - L\right) \tag{8.16}$$

Expanding yields

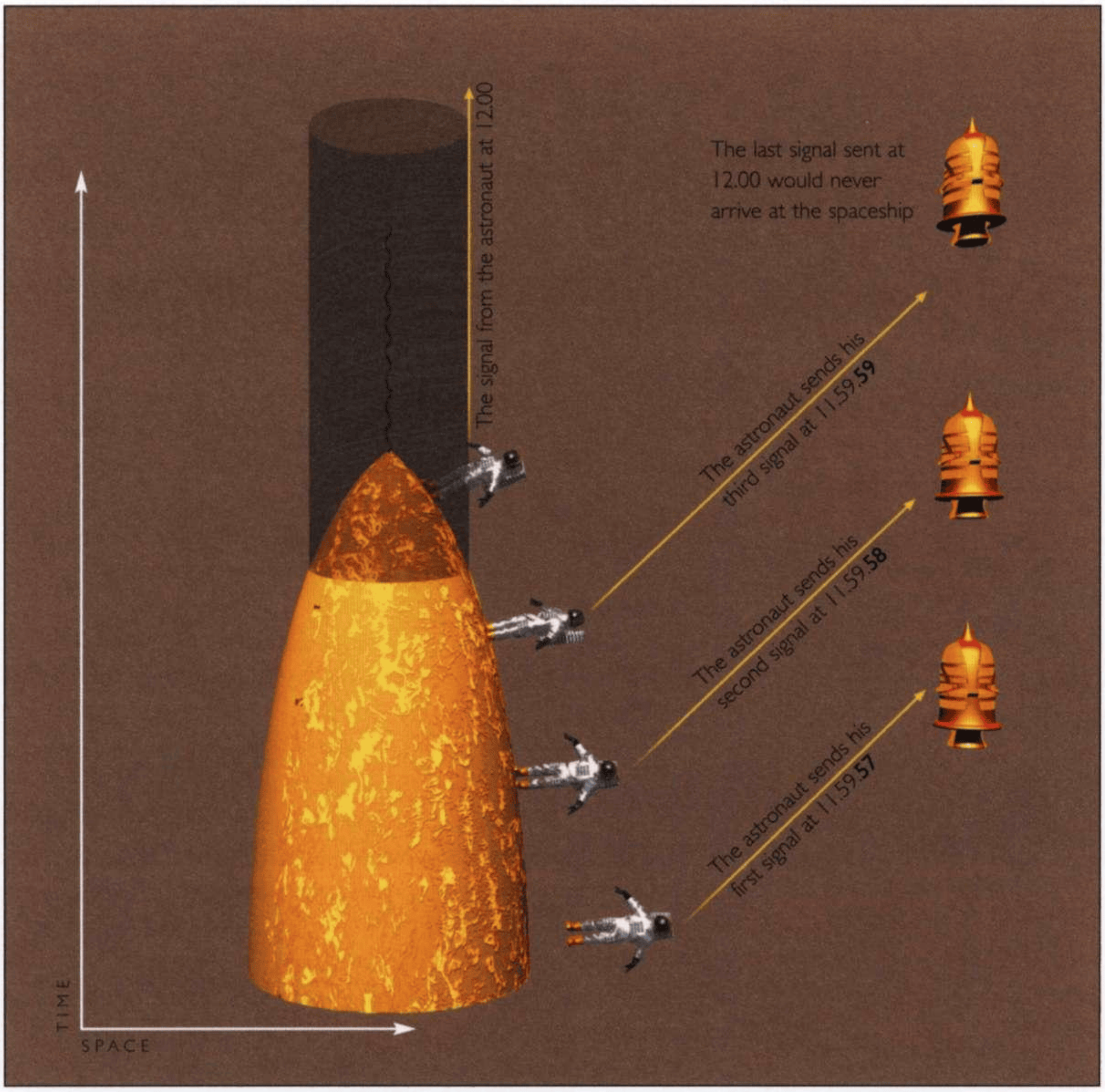
Theorem 8.1 (Schwarzschild geodesic equation)

$$V_{\text{eff}}(r) = \underbrace{\frac{LM}{r}}_{\text{Newtonian term}} + \underbrace{\frac{1}{2}\frac{\ell^2}{r^2}}_{\text{centrifugal barrier term}} - \underbrace{\frac{M\ell^2}{r^3}}_{\text{GR correction}} - \underbrace{\frac{1}{2}L}_{\text{insignificant constant}} \tag{8.17}$$

Remark 8.4 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 8.2 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



The illustration above shows an astronaut who lands on a collapsing star at 11.59.57 and joins the star as it shrinks below the critical radius where gravity is so strong that no signal can escape. He sends signals from his watch to a spaceship orbiting the star at regular intervals.

Someone watching the star at a distance will never see it cross the event horizon and enter the black hole. Instead, the star will appear to hover just outside the critical radius, and a clock on the surface of the star will seem to slow down and stop.

Figure 8.2: Astronaut entering the event horizon (*The Universe in a Nutshell*)

Exercise 8.1 Noting the conservation of mass and energy in relation to the geodesic equations, derive that

$$\frac{dt}{ds} = \frac{E}{M}\left(1 - \frac{2M}{r}\right)^{-1} \qquad \frac{d\phi}{dr} = \frac{\ell/M}{r^2 \sin^2\theta} \tag{8.18}$$

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 8.2 (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:

$$L = \begin{cases} -1 & \text{for massive particles} \\ 0 & \text{for massless particles (photons, gravitons, etc.)} \end{cases} \tag{8.19}$$

Remark 8.5 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 8.6 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 8.4 (Schwarzschild orbits) Orbits emerge on the radial stationary points of the effective potential. i.e. when

$$\frac{dV_{\text{eff}}}{dr} = 0 \to Lr^2 + \frac{\ell^2}{M}r - 3\ell^2 = 0 \tag{8.20}$$

Theorem 8.3 (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 8.3 (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

$$ds^2 = -\left(1 - \frac{2M}{r}\right)dt^2 + \left(1 - \frac{2M}{r}\right)^{-1}dr^2 + r^2 d\phi^2 \tag{8.21}$$

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

$$\frac{dr}{ds} = \sqrt{\frac{E^2}{M^2} - \frac{1}{M^2}\left(\frac{\ell^2}{r^2M^2} - L\right)\left(1 - \frac{2M}{r}\right)} \tag{8.22}$$

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

$$\frac{dr}{ds} = \frac{dr}{d\phi}\frac{\ell}{r^2M} \tag{8.23}$$

This then gives

Definition 8.5 (Orbit equation)

$$\left(\frac{dr}{d\phi}\right)^2 = \frac{E^2 - (1-\frac{2M}{r})\left(\frac{\ell^2}{r^2} - L\right)}{\ell^2/r^4} \tag{8.24}$$

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

$$\left(\frac{du}{d\phi}\right)^2 = \frac{E^2+L}{\ell^2} - \frac{2ML}{\ell^2}u - u^2 + 2Mu^3 \tag{8.25}$$

Remark 8.7 Ignoring the last term yields the Newtonian orbit equation.
Differentiating with respect to ϕ, we can simplify the equations even more.

Definition 8.6 (Simplified orbit equation)

$$\frac{d^2u}{d\phi^2} = -\frac{ML}{\ell^2} - u + 3Mu^2 \tag{8.26}$$

Remark 8.8 Again, ignoring the last term yields the Newtonian orbit equation.

[a] You will almost immediately see why.

Exercise 8.2 Use the Einstein and Newton orbit equations to investigate a planet close to the Sun. Prove that the relativistic correction term results in very small deviations from the Newtonian predictions.

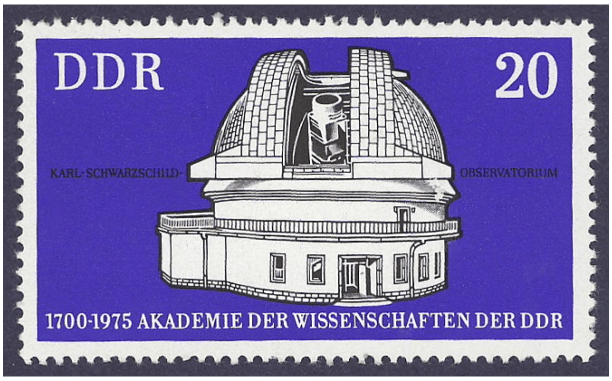


Figure 8.3: The GDR Academy of Sciences celebrates Schwarzschild.

8.3 Alternative 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 8.4 (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:

$$dr^* = \left(1 - \frac{2M}{r}\right)^{-1} dr \rightarrow r^* = r + 2M \ln\left|\frac{r}{2M} - 1\right| \tag{8.27}$$

Now the metric is of the form

$$ds^2 = \left(1 - \frac{2M}{r}\right)(-dt^2 + dr^{*2}) + r^2 d\Omega^2 \tag{8.28}$$

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

$$v = t + r^* \tag{8.29}$$

The metric hence becomes

Definition 8.7 (Schwarzschild metric in ingoing Eddington-Finkelstein coordinates)

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dv^2 + 2dvdr + r^2 d\Omega^2 \tag{8.30}$$

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*

$$u = t - r^* \tag{8.31}$$

which yields the metric

Definition 8.8 (Schwarzschild metric in outgoing Eddington-Finkelstein coordinates)

$$ds^2 = -\left(1 - \frac{2M}{r}\right) du^2 - 2dudr + r^2 d\Omega^2 \tag{8.32}$$

Remark 8.9 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 8.10 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.
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 8.5 (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:

$$ds^2 = -\left(1 - \frac{2M}{r}\right) dudv + r^2 d\Omega^2 \tag{8.33}$$

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

$$U = -4Me^{-u/4M} \quad V = 4Me^{v/4M} \tag{8.34}$$

and the metric becomes

$$ds^2 = -\frac{2M}{r} e^{-r/2M} dUdV + r^2 d\Omega^2 \tag{8.35}$$

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

$$T = \frac{1}{2}(V + U) \quad R = \frac{1}{2}(V - U) \quad \text{for} \quad r > 2M \tag{8.36}$$

[1] His son Peter Szekeres is responsible for one of the references of this book.

$$T = \frac{1}{2}(V - U) \quad R = \frac{1}{2}(V + U) \quad \text{for} \quad r > 2M \tag{8.37}$$

thus giving

Definition 8.9 (Kruskal-Szekeres coordinates)

$$ds^2 = \frac{2M}{r} e^{-r/2M}(-dT^2 + dR^2) + r^2 d\Omega^2 \tag{8.38}$$

where r is now implicitly defined as

$$16M^2 \left(\frac{r}{2M} - 1\right) e^{r/2M} = R^2 - T^2 \tag{8.39}$$

8.4 Related exact solutions

Quote 8.3 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.
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 interior solution*, which assumes the mass to be a spherical and uniform perfect fluid.

Derivation 8.6 (Schwarzschild interior solution) We recall that the stress-energy tensor of a perfect fluid is defined by (5.14), where u is the 4-velocity observing

$$u = (-e^{A(R)/2}, 0, 0, 0) \tag{8.40}$$

We can them derive the stress-energy tensor

$$T_{\mu\nu} = \begin{pmatrix} \rho e^{A(r)} & 0 & 0 & 0 \\ 0 & p e^{B(r)} & 0 & 0 \\ 0 & 0 & pr^2 & 0 \\ 0 & 0 & 0 & \sin^2\theta \end{pmatrix} \tag{8.41}$$

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

$$e^{-B} = 1\frac{2m(r)}{r} \tag{8.42}$$

As energy and momentum is conserved:

$$\nabla_\mu T^\mu_\nu = 0 \tag{8.43}$$

we can likewise determine A

$$A(r) = \log\left(\frac{C}{\rho_0 + p(r)}\right)^2 \tag{8.44}$$

and thus, the Schwarzschild interior metric

Definition 8.10 (Schwarzschild interior metric)

$$ds^2 = -\left(\frac{\rho_0 + p_c}{\rho_0 + p(r)}\right)^2 dt^2 + \frac{dr^2}{1 - (8\pi/3)\rho_0 \pi^2} + r^2 d\Omega^2 \tag{8.45}$$

A charged nonrotating mass can be represented by the Reissner-Nordström solution:

Definition 8.11 (Reissner-Nordström metric)

$$ds^2 = dt^2 = \left(1 - \frac{2M}{r} + \frac{r_q^2}{r^2}\right) dt^2 - \left(1 - \frac{2M}{r} + \frac{r_q^2}{r^2}\right)^{-1} dr^2 - r^2 d\theta^2 - r^2 \sin^2\theta d\varphi^2 \tag{8.46}$$

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

An uncharged rotating mass can be represented by the Kerr solution:

Definition 8.12 (Kerr metric)

$$ds^2 = -\frac{\Delta_r}{\varrho^2}\left(dt - a\sin^2\theta d\phi\right)^2 + \frac{\varrho^2}{\Delta_r} dr^2 + \varrho^2 d\theta^2 + \frac{\sin^2\theta}{\varrho^2}\left(adt - \left(r^2 + a^2\right) d\phi\right)^2 \tag{8.47}$$

where

$$\varrho^2 = r^2 + a^2\cos^2\theta \quad \Delta_r = r^2 - 2Mr + a^2 \tag{8.48}$$

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

Remark 8.11 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 8.3 Show that in the limit $a \to 0$, the Kerr metric becomes the Schwarzschild metric.

Remark 8.12 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:

$$r_\pm = M \pm \sqrt{M^2 - r^2} \tag{8.49}$$

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

$$ds^2 = -dt^2 + \left(r^2 + a^2\cos^2\theta\right)\left(\frac{dr^2}{r^2 + a^2} + d\theta^2\right) + \left(r^2 + a^2\right)\sin^2\theta d\phi^2 \tag{8.50}$$

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

$$r_{e\pm} = M \pm \sqrt{M^2 - a^2\cos^2\theta} \tag{8.51}$$

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.
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 8.13 (Mino time in the Kerr metric) In the Kerr metric, the Mino time λ observes

$$\frac{d\tau}{d\lambda} = \varrho^2 \tag{8.52}$$

Remark 8.13 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.
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.

Definition 8.14 (Kerr-Newman metric) At first glance, the Kerr-Newman metric is identical to the

[2] Aww!

Kerr metric

$$ds^2 = -\frac{\Delta_r}{\varrho^2}\left(dt - a\sin^2\theta d\phi\right)^2 + \frac{\varrho^2}{\Delta_r}dr^2 + \varrho^2 d\theta^2 + \frac{\sin^2\theta}{\varrho^2}\left(adt - \left(r^2 + a^2\right)d\phi\right)^2 \tag{8.53}$$

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

$$\Delta_r = r^2 - 2Mr + a^2 + r_q^2 \tag{8.54}$$

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

Figure 8.4: Schwarzschild takes a dirt nap (J. Giesen, 2004)

8.5 Penrose diagrams [EMPTY]

Figure 8.5: The great Gerard ‘t Hooft holding up a Penrose diagram, 30 April 2025.

Part II

Tetrads [DRAFT]

Chapter 9

Tetrads

Quote 9.1 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.

9.1 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 9.1 (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:

$$\gamma_a = (\gamma_0, \gamma_1, \gamma_2, \gamma_3) \tag{9.1}$$

The corresponding *tetrad metric* is

$$\gamma_{ab} = \gamma_a \gamma_b \tag{9.2}$$

Remark 9.1 In this book, the tetrad *basis* γ_a should not be confused with the tetrad *metric* γ_{ab}.

Note 9.1 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 9.2 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 9.2 (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 9.2 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 9.3 (Tetrad field) The tetrad field or *frame field* e_a^μ converts the coordinate basis X_μ to the tetrad basis γ_a:

$$\gamma_a = e_a^\mu X_\mu \tag{9.3}$$

Theorem 9.1 (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):

$$e_\mu'^a = \Lambda_b^a e_\nu^b \frac{\partial X^\nu}{\partial X'^\mu} \tag{9.4}$$

Remark 9.3 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 9.2 (Tetrad field properties) A tetrad field observes the following properties:

- Contractions:
$$e_i^\mu e_\mu^j = \delta_j^i \quad e_i^\mu e_\nu^i = \delta_\nu^\mu \tag{9.5}$$
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:
$$\gamma_{ab} = e_a^c e_d^b g_{cd} \tag{9.6}$$

Exercise 9.1 Suppose we encode a tetrad field as

$$e_\mu^0 dX^\mu = \left(1 - \frac{2M}{r}\right)^{1/2} dt \quad e_\mu^1 dX^\mu = \left(1 - \frac{2M}{r}\right)^{-1/2} dr \quad e_\mu^2 dX^\mu = r d\theta \quad e_\mu^3 dX^\mu = r \sin\theta d\phi \tag{9.7}$$

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

Remark 9.4 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.

9.2 Tetrad derivatives

Definition 9.4 (Tetrad derivative) We define the tetrad derivative with respect to a coordinate derivative:

$$\partial_m = \gamma_m \cdot \partial = \gamma_m \cdot e^\mu \frac{\partial}{\partial X^\mu} = e^\mu_m \frac{\partial}{\partial X^\mu} \tag{9.8}$$

Unlike coordinate derivatives, they do not commute:

Theorem 9.3 (Tetrad derivative commutation relation)

$$[\partial_m, \partial_n] = e^\nu_m \frac{\partial e^\nu_n}{\partial X^\nu} \frac{\partial}{\partial X^\mu} - e^\nu_n \frac{\partial e^\nu_m}{\partial X^\nu} \frac{\partial}{\partial X^\mu} \tag{9.9}$$

which is *not* a tetrad tensor.

We can shorten this as

$$[\partial_m, \partial_n] = (d^k_{mn} - d^k_{nm})\partial_k \tag{9.10}$$

where we have the *tetrad field derivative*:

Definition 9.5 (Tetrad field derivative)

$$d^k_{mn} = e^\mu_m e^\nu_n \frac{\partial e^k_\mu}{\partial X^\nu} \tag{9.11}$$

This is again *not* a tetrad tensor.

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

Derivation 9.1 (Tetrad covariant derivative) Consider a tetrad 4-vector A^m and its abstract 4-vector $A = \gamma_m A^m$. The directional derivative is

$$\partial_n A = \gamma_m \partial_n A^m + (\partial_n \gamma_m) A^m \tag{9.12}$$

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

$$\partial_n \gamma_m = e^\mu_n \frac{\partial \gamma_m}{\partial X^\mu} \tag{9.13}$$

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

$$\partial_n \gamma_m = \Gamma^k_{mn} \gamma_k \tag{9.14}$$

The original direction derivative of the abstract tensor then becomes

$$\partial_n A = \gamma_k (\nabla_n A^k) \tag{9.15}$$

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

$$\frac{\partial e_\mu}{X^\nu} = \Gamma^\kappa_{\mu\nu} e_\kappa = \frac{\partial e^m_\mu \gamma_m}{\partial X_\nu} \tag{9.16}$$

This becomes

$$\Gamma^\kappa_{\mu\nu} e_\kappa = e^m_\mu e^n_\nu (d^k_{mn} + \Gamma^k_{mn}) \gamma_k \tag{9.17}$$

which yields the relation

Theorem 9.4 (Tetrad Christoffel-coordinate Christoffel relation)

$$d_{lmn} + \Gamma_{lmn} = e_l^{\lambda} e_m^{\mu} e_n^{\nu} \Gamma_{\lambda_\mu \nu} \tag{9.18}$$

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

$$\Gamma_{lmn} = \gamma_{lk} \Gamma^{k}_{mn} \tag{9.19}$$

9.3 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 9.6 (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 9.5 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 9.7 (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

$$k = \frac{1}{\sqrt{2}}(t + z) \quad l = \frac{1}{\sqrt{2}}(t - z) \tag{9.20}$$

and the second two are a complex vector and its conjugate

$$m = \frac{1}{\sqrt{2}}(x - iy) \quad \bar{m} = \frac{1}{\sqrt{2}}(x + iy) \tag{9.21}$$

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

$$e_0{}^{\mu} = l^{\mu} \quad e_1{}^{\mu} = n^{\mu} \quad e_2{}^{\mu} = m^{\mu} \quad e_3{}^{\mu} = \bar{m}^{\mu} \tag{9.22}$$

The metric is then

$$g_{ab} = -l_a n_b - n_a l_b + m_a \bar{m}_b + \bar{m}_a m_b \tag{9.23}$$

Remark 9.6 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 9.5 (Null tetrad properties) The null tetrad satisfies the orthogonality and normalisation conditions:

$$l^a k_a = -1, \quad m^a \bar{m}_a = 1, \quad l^a m_a = l^a \bar{m}_a = k^a m_a = k^a \bar{m}_a = 0 \tag{9.24}$$

Theorem 9.6 (Null tetrad transformations) A null tetrad can undergo the following transformations. Together, these represent the six-parameter group of Lorentz transformations.

- **Boost transformation:**

$$k' = Bk, \quad l' = B^{-1} l, \quad m' = m, \quad \bar{m}' = \bar{m} \tag{9.25}$$

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

- **Spin transformation:**

$$k' = k, \quad l' = l, \quad m' = e^{i\Phi} m, \quad \bar{m}' = e^{-i\Phi} \bar{m} \tag{9.26}$$

where Φ is a real phase parameter.

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

$$k' = k, \quad l' = l + L\bar{m} + \bar{L}m + L\bar{L}k, \quad m' = m + Lk, \quad \bar{m}' = \bar{m} + \bar{L}k \tag{9.27}$$

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

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

$$k' = k + K\bar{m} + \bar{K}m + K\bar{K}l, \quad l' = l, \quad m' = m + Kl, \quad \bar{m}' = \bar{m} + \bar{K}l \tag{9.28}$$

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

9.4 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 9.8 (Newman-Penrose scalars) Real NP scalar components:

$$\Phi_{00} = \frac{1}{2}R_{ab}k^a k^b \quad \Phi_{11} = \frac{1}{2}R_{ab}(k^a l^b + m^a \bar{m}^b) \quad \Phi_{22} = \frac{1}{2}R_{ab}l^a l^b \quad \Lambda = \frac{R}{24} \tag{9.29}$$

Complex NP scalar components:

$$\Phi_{01} = \bar{\Phi}_{10} = \frac{1}{2}R_{ab}k^a m^b \quad \Phi_{02} = \bar{\Phi}_{20} = \frac{1}{2}R_{ab}m^a m^b \quad \Phi_{12} = \bar{\Phi}_{21} = \frac{1}{2}R_{ab}l^a m^b \tag{9.30}$$

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 9.9 (Weyl scalars)

$$\Psi_0 = C_{abcd}l^a m^b l^c m^d \tag{9.31}$$

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

$$\Psi_1 = C_{abcd}l^a k^b l^c m^d \tag{9.32}$$

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

$$\Psi_2 = C_{abcd}l^a m^b \bar{m}^c k^d \tag{9.33}$$

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.

$$\Psi_3 = C_{abcd}l^a k^b \bar{m}^c k^d \tag{9.34}$$

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

$$\Psi_4 = C_{abcd}k^a \bar{m}^b k^c \bar{m}^d \tag{9.35}$$

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

Remark 9.7 Remember that a gravitational wave is a wave. The transverse and longitudinal components are the same as in any other wave.

Remark 9.8 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 9.9 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 9.7 (Petrov classification) In the *Petrov classification*, also known as the *classification of gravitational fields*, we have the following possible types:

- **Type I:**
$$\Psi_0 = 0 \tag{9.36}$$

- **Type II:**
$$\Psi_0 = \Psi_1 = 0 \tag{9.37}$$
Type II regions combine the effects for types D, III, and N in a rather complicated nonlinear way.

- **Type III:**
$$\Psi_0 = \Psi_1 = \Psi_2 = 0 \tag{9.38}$$
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:**
$$\Psi_0 = \Psi_1 = \Psi_2 = \Psi_3 = 0 \tag{9.39}$$
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:**
$$\Psi_0 = \Psi_1 = \Psi_3 = \Psi_4 = 0 \tag{9.40}$$
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:**
$$\Psi_0 = \Psi_1 = \Psi_2 = \Psi_3 = \Psi_4 = 0 \tag{9.41}$$
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 10

ADM formalism

> **Quote 10.1** 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.

10.1 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 10.1 (Hamilton's equations)

$$\frac{\partial H}{\partial X_i} = -\dot{p}_i \quad \frac{\partial H}{\partial p_i} = \dot{X}_i \tag{10.1}$$

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

$$X^\mu = (t, x^\alpha) \quad \text{where} \quad \alpha = 1, 2, 3 \tag{10.2}$$

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

$$\gamma_m = (\xi_0, \xi_a) \quad \text{where} \quad a = 1, 2, 3 \tag{10.3}$$

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:

$$\xi_0 \cdot \xi_0 = -1 \quad \xi_0 \cdot \xi_a = 0 \tag{10.4}$$

[1] Here, ξ has nothing to do with Killing vectors.

The tetrad metric γ_{mn} becomes

$$\gamma_{mn} = \begin{pmatrix} -1 & 0 \\ 0 & \gamma_{ab} \end{pmatrix} \tag{10.5}$$

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

$$g_{\alpha\beta} = \gamma_{ab} e^a_\alpha e^b_\beta \tag{10.6}$$

The *spatial tetrad field* and the *inverse spatial tetrad* field are

Definition 10.1 (Spatial tetrad field and inverse spatial tetrad field)

$$e^a_\alpha = \begin{pmatrix} \alpha & 0 \\ -e^a_\alpha \beta^\alpha & e^a_\alpha \end{pmatrix} \quad e^\alpha_a = \begin{pmatrix} 1/\alpha & \beta^\alpha/\alpha \\ 0 & e^\alpha_a \end{pmatrix} \tag{10.7}$$

where we introduce the two following terms:

Definition 10.2 (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.

$$\alpha = \frac{d\tau}{dt} \tag{10.8}$$

Physically, this is the rate of advance of proper time between the constant time hypersurfaces.

Definition 10.3 (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.

$$\beta^\alpha = \frac{dX^\alpha}{dt} \tag{10.9}$$

Physically, this is the rate at which coordinates shift between the constant time hypersurfaces.

Remark 10.1 The *spatial tetrad field* e^a_α and its inverse, the *inverse tetrad field* e^α_a observe

$$e^0_\alpha = e^t_a = 0 \tag{10.10}$$

The original 4D metric and inverse metric are then

$$g_{\mu\nu} = \begin{pmatrix} -\alpha^2 + \beta_\mu \beta^\mu & \beta_\nu \\ \beta_\mu & \gamma_{\mu\nu} \end{pmatrix} \quad g^{\mu\nu} = \begin{pmatrix} -\frac{1}{\alpha^2} & \frac{\beta^\nu}{\alpha^2} \\ \frac{\beta^\mu}{\alpha^2} & \gamma_{\mu\nu} - \frac{\beta^\mu \beta^\nu}{\alpha^2} \end{pmatrix} \tag{10.11}$$

and the determinants observe

$$\sqrt{-g} = \alpha\sqrt{y} \tag{10.12}$$

Rewriting in terms of the line element gives

Definition 10.4 (Original 4D line element)

$$ds^2 = (-\alpha^2 + \beta_\mu \beta^\mu)dt^2 + 2\beta_\mu dt dx^\mu + \gamma_{\mu\nu} dx^\mu dx^\nu \tag{10.13}$$

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 10.5 (Normal vector)

$$n^\mu = (1/\alpha, -\beta^i/\alpha) \tag{10.14}$$

From this, we can express the (3D) spatial metric.

Definition 10.6 (Spatial metric)

$$\gamma_{\mu\nu} = g_{\mu\nu} + n_\mu n_\nu \tag{10.15}$$

In summation convention:

$$n^\mu = -\alpha \nabla^\mu t \tag{10.16}$$

where the minus sign is there to guarantee that n^μ is future pointing.

Remark 10.2 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 10.1 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.

10.2 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:

$$\gamma_{ab} = g_{\mu\nu} e^\mu_a e^\nu_b \quad ds^2 = g_{\mu\nu} dx^\mu dx^\nu = \gamma_{ab} dy^a dy^b \tag{10.17}$$

We then consider the decomposed 3+1D spacetime, which has the parametric equation

$$X^\alpha = X^\alpha(Y_a) \tag{10.18}$$

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

$$ds^2 = g_{\mu\nu} dX^\mu dX^\nu = h_{ab} dY^a dY^b \tag{10.19}$$

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 10.7 (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:

$$\mathrm{I} = ac\langle X_u, X_u\rangle + (ad+bc)\langle X_u, X_v\rangle + bd\langle X_v, X_v\rangle = Eac + F(ad+bc) + Gbd \tag{10.20}$$

One can write this in matrix notation and the form of the metric:

$$\mathrm{I} = x^T \begin{pmatrix} E & F \\ F & G \end{pmatrix} y \quad g_{\mu\nu} = \begin{pmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{pmatrix} = \begin{pmatrix} E & F \\ F & G \end{pmatrix} \tag{10.21}$$

Remark 10.3 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 10.8 (Projection operator) The projection operator projects objects on the manifold onto the hypersurface:

$$P^\alpha_\beta = \delta^\alpha_\beta + n^\alpha n_\beta \tag{10.22}$$

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 10.9 (Extended covariant derivative) The *extended covariant derivative* is

$$D_\mu = P^\alpha_\mu \nabla_\alpha \tag{10.23}$$

Remark 10.4 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 $\mathrm{I\!I}$. 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 10.2 (Unit normal on a hypersurface)

$$n_\mu = \epsilon \left| g^{\alpha\beta} \frac{\partial f}{\partial X^\mu} \frac{\partial f}{\partial X^\mu} \right|^{-\frac{1}{2}} \frac{\partial f}{\partial X^\mu} \tag{10.24}$$

where n_μ observes $n_\mu e^\mu_a$ and for convenience, we define an arbitrary scalar $\epsilon = n^\mu n_\nu$. ϵ satisfies

$$n^\mu n_\nu = \epsilon = \begin{cases} -1 & \text{if } \Sigma \text{ is spacelike} \\ 1 & \text{if } \Sigma \text{ is timelike} \end{cases} \tag{10.25}$$

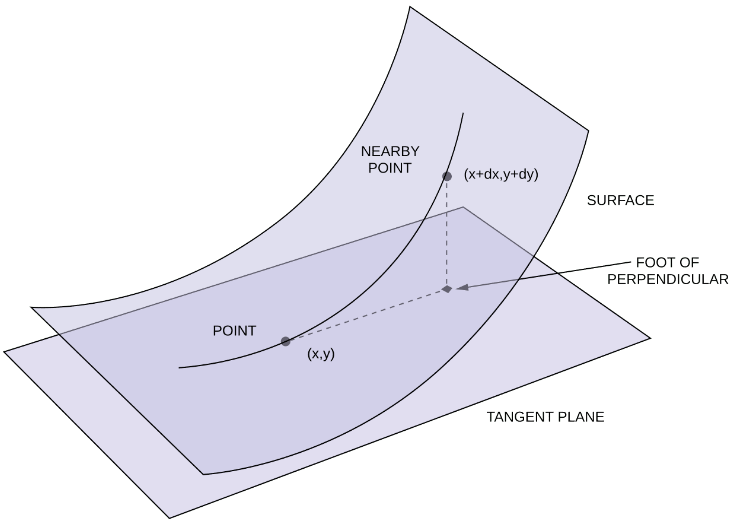


Figure 10.1: A surface and its tangent plane.

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 10.5 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 10.10 (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 defined as:

$$K_{ij} = -n_\mu \nabla_j e^\mu_i = -e^a_i e^b_j n_\mu \left(\partial_b e^\mu_a + \omega_{b\,a}{}^c e^\mu_c \right) \tag{10.26}$$

[2] Or rather go back?
[3] There is also the *third fundamental form*, although this is outside the scope of the book.

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 10.6 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 10.3 (Gauss-Codazzi equations)

- **Gauss equation**:
$$P^\rho_\alpha P^\beta_\sigma P^\gamma_\mu P^\delta_\nu \, {}^{(4)}R^\alpha_{\beta\gamma\delta} = \mathcal{R}^\rho_{\sigma\mu\nu} + K^\rho_\mu K_{\sigma\nu} - K^\rho_\nu K_{\sigma\mu} \tag{10.27}$$
- **Codazzi equation**:
$$P^\alpha_\rho P^\mu_\beta P^\nu_\gamma \, {}^{(4)}\mathcal{R}^\rho_{\sigma\mu\nu} n^\sigma = D_\gamma K^\alpha_\beta - D_\beta K^\alpha_\gamma \tag{10.28}$$

We can then derive the Ricci scalar from the extrinsic curvature:

Theorem 10.4 (Ricci scalar)

$${}^{(4)}R = R + K^2 + K^{ij}K_{ij} - 2\nabla_n K = \frac{2}{N} D^i D_i N \tag{10.29}$$

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 $T_{\alpha\beta}$:

$$T = S - E \tag{10.30}$$

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 10.11 (Stress tensor)
$$S_{\alpha\beta} = T_{\mu\nu} P^\nu_\alpha P^\nu_\beta \tag{10.31}$$

With these projected quantities we have the total projection of the Einstein field equations.

10.3 3+1 field equations

Theorem 10.5 (3+1 field equations)

- **Total projection onto Σ_t**:
$$\mathcal{L}_m K_{ij} = -D_i D_j N + N[R_{ij} - 2K_{il}K^l_j + KK_{ij} + 4\pi(\gamma_{ij}(S_E) - 2S_{ij})] \tag{10.32}$$
Note that $\mathcal{L}_m$ is the Lie derivative we have encountered earlier. The Lagrangian is L as usual.
- **Total projection along n^μ**:
$$R - K_{ij}K^{ij} + K^2 = 16\pi E \tag{10.33}$$
This is known as the *Hamiltonian constraint*.
- **Mixed projection onto Σ_t and along n^μ**:
$$D_j K^i_j - D_i K = 8\pi p_i \tag{10.34}$$
where $p_i = -T_{\mu\nu} n^\mu \gamma^\nu_i$ is the momentum density. This is known as the *momentum constraint*.

10.4 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

$$\pi = \frac{\partial L}{\partial \dot{q}} \tag{10.35}$$

In GR, the tensorial equivalent is defined as

Definition 10.12 (ADM canonical momentum)

$$\pi^{ij} = \frac{\partial L}{\partial \dot{\gamma}_{ij}} = -\sqrt{\gamma}(K^{ij} - \gamma^{ij}K) \tag{10.36}$$

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 10.6 (ADM Hamilton's equations)

$$\frac{\partial H}{\partial \pi^{ij}} = \dot{\gamma}_{ij} \quad \frac{\partial H}{\partial \gamma^{ij}} = -\dot{\pi}_{ij} \tag{10.37}$$

Theorem 10.7 (Evolution equations)

- **Spatial metric**:

$$\dot{\gamma}_{ij} = D_i\beta_j + D_j\beta_i - 2\alpha K_{ij} \tag{10.38}$$

- **Extrinsic curvature**:

$$\begin{aligned}\dot{K}_{ij} =& -D_iD_j\alpha + \alpha(R_{ij} - 2K_{il}K^l_j + KK_{ij}) + \beta^k\partial_k K_{ij} \\ &+ K_{ik}\partial_j\beta^k + K_{jk}\partial_i\beta^k - 8\pi G\alpha\left(S_{ij} - \frac{1}{2}\gamma_{ij}(S-\rho)\right)\end{aligned} \tag{10.39}$$

- **Canonical momentum**:

$$\begin{aligned}\dot{\pi}^{ij} =& -N\sqrt{\gamma}\left(R^{ij} - \frac{1}{2}\gamma^{ij}R\right) + \frac{N}{2\sqrt{\gamma}}\left(\pi_{cd}\pi^{cd} - \frac{\pi^2}{2}\right)\gamma^{ij} \\ &- \frac{2N}{\sqrt{\gamma}}\left(\pi^{ic}\pi_c{}^j - \frac{1}{2}\pi\pi^{ij}\right) + \sqrt{\gamma}\left(D^iD^jN - \gamma^{ij}D_cD^cN\right) \\ &+ D_c\left(\pi^{ij}N^c\right) - \pi^{ic}D_cN^j - \pi^{jc}D_cN^i\end{aligned} \tag{10.40}$$

10.5 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* $\Psi[\gamma_{ab}]$:

$$\hat{\gamma}_{ij}(t, x^k)|\Psi\rangle \to \gamma_{ij}(t, x^k)\Psi[\gamma_{ab}] \tag{10.41}$$

$$\hat{\pi}^{ij}(t, x^k)|\Psi\rangle \to -i\frac{\delta}{\delta\gamma_{ij}(t, x^k)}\Psi[\gamma_{ab}] \tag{10.42}$$

The first and third evolution equations then become

$$\hat{R}^0|\Psi\rangle \doteq -\left[\sqrt{\gamma}R + \frac{1}{\sqrt{\gamma}}\left(\frac{\pi^2}{2} - \pi^{ij}\pi_{ij}\right)\right]\Psi\left[\gamma_{kl}\right] = 0 \tag{10.43}$$

$$\hat{R}^i|\Psi\rangle \doteq -2D_j\pi^{ij}\Psi\left[\gamma_{kl}\right] = 0 \tag{10.44}$$

where first one can be rewritten as

[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 $p = mv$.

Theorem 10.8 (Wheeler-Dewitt equation)

$$\left[\sqrt{\gamma}R - \frac{\hbar^2}{\sqrt{\gamma}}\left(\frac{1}{2}\gamma_{ab}\gamma_{cd} - \gamma_{ac}\gamma_{bd}\right)\frac{\delta}{\delta\gamma_{ab}}\frac{\delta}{\delta\gamma_{cd}}\right]\Psi\left[\gamma_{kl}\right] = 0 \tag{10.45}$$

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

Trilobites [DRAFT]

Chapter 11

Cosmology before GR

Quote 11.1 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.

11.1 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 11.1 (Comoving coordinates) The *comoving coordinates cosmological reference frame* is a set of coordinates in which physical quantities are homogeneous and isotropic.

Remark 11.1 They are called 'comoving' coordinates because they move along with the Hubble flow.

Note 11.1 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 11.2 (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:

$$\vec{r}(t) = \frac{a(t)}{a_0}\vec{r}_0 \quad \text{where} \quad a_0 = a(t_0) \tag{11.1}$$

Remark 11.2 The scale factor is dimensionless and characterises the expansion of the universe.

Definition 11.3 (Comoving observer) The *comoving observer* is an observer at rest in comoving coordinates.

Definition 11.4 (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 11.3 We observe a large dipole in the CMB that results from the proper motion of the Milky Way with respect to comoving coordinates.

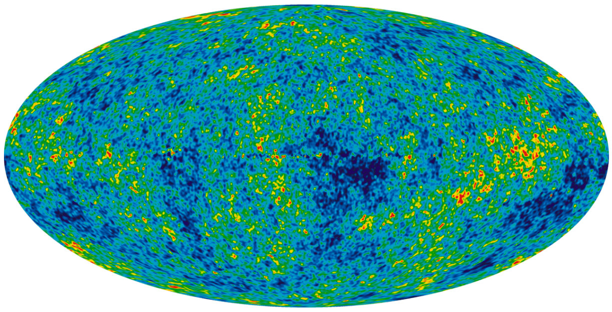


Figure 11.1: 'These are the axes of evil... you must conquer each!'

11.2 Historical overview

Now with all the basic concepts by our side, we can proceed with a recap of the history of cosmology.

> **Quote 11.2** 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 11.5 (Homogeneity) A system is *homogeneous* if it is invariant under translations

$$X_a \to X_a + X_c \tag{11.2}$$

Definition 11.6 (Isotropy) A system is *isotropic* if it is invariant under rotations. i.e. it looks the same in all directions.

Theorem 11.1 (Cosmological principle) There exist coordinate systems such that the universe appears spatially isotropic and homogeneous on sufficiently large scales for every observer.

Remark 11.4 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 11.2 (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 11.3** 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 11.7 (Staticity) Static solutions are models in which the universe remains constant in size.

Derivation 11.1 (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

$$n(r) = n_* = \text{constant} \tag{11.3}$$

Then the number of stars in a spherical shell located at distance r with thickness dr is

$$N_* = 4\pi n_* r^2 dr \tag{11.4}$$

Assume a star located at r has luminosity L. Then, the apparent luminosity (flux density) is

$$F' = \frac{L}{(4\pi r^2)} \tag{11.5}$$

We can then calculate the total luminosity (flux density) that reaches us from all stars in an infinite, static universe:

$$\int_0^\infty F'(r) 4\pi r^2 dr = \int_0^\infty \frac{L}{4\pi r^2} 4\pi n_* r^2 dr = n_* \int_0^\infty dr = \infty \tag{11.6}$$

This means that the night sky would be infinitely bright - which it is not.

Remark 11.5 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:

$$d \propto z := \frac{\Delta\lambda}{\lambda} \tag{11.7}$$

We thus have *Hubble's law*:

Theorem 11.3 (Hubble's law)

$$\vec{v} = H(t)\vec{r} \tag{11.8}$$

where significantly, H is independent of $\vec{r}$ and is only a function of t.

Note 11.2 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 11.4 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 11.6 A new development is the *Hubble tension*.

11.3 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 11.7 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 11.4 (Non-relativistic fluid equations) The equations of non-relativistic fluid dynamics are:

- **Continuity equation**: This describes the conservation of mass in the fluid.

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \vec{v}) = 0 \tag{11.9}$$

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.

$$\frac{D\vec{v}}{Dt} = \frac{\partial \vec{v}}{\partial t} + \vec{v} \cdot \underbrace{\nabla \vec{v}}_{\text{covariant derivative}} = -\frac{\nabla p}{\rho} + \vec{f} \tag{11.10}$$

where D is the *material derivative* we saw in (3.60), 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 $\vec{f} = \vec{g}$:

$$\nabla^2 \Phi = 4\pi G \rho \tag{11.11}$$

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 11.8 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 11.3 Remember that $\vec{r} = t\vec{v}$ and all that follows.

Remark 11.9 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 11.8 (Hubble parameter) We can represent the *Hubble parameter* H by the scale factor:

$$H(t) = \frac{\dot{a}(t)}{a(t)} \tag{11.12}$$

Noting (11.8) and that:

- The divergence of ϕ observes, by the definition of $\vec{g}$:

$$\nabla \Phi = \vec{g} \tag{11.13}$$

- The universe, being isotropic and homogeneous, has no pressure gradients[2]:

$$\nabla p = 0 \tag{11.14}$$

[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 11.5 (Newtonian cosmology fluid equations) Continuity equation:

$$\frac{\partial \rho}{\partial t} + 3\rho H(t) = 0 \tag{11.15}$$

Euler's equation:

$$\frac{\partial H}{\partial t} + H(t)^2 = \vec{f}(t) \tag{11.16}$$

Gravity equation:

$$g(t) = -\frac{4\pi G\rho}{3} \tag{11.17}$$

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 (13.7), (13.8) and (13.9). They reduce to

Theorem 11.6 (Friedmann equations for non-relativistic matter)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{\Lambda}{3} \tag{11.18}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\rho \tag{11.19}$$

$$\frac{d}{dt}(\rho a^3) = 0 \tag{11.20}$$

Einstein introduced the cosmological constant to maintain the static universe[3]:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\left(\rho - \frac{\Lambda c^2}{4\pi G}\right) \tag{11.21}$$

Remark 11.10 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 12

FLRW model

Quote 12.1 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.

12.1 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 12.1 (Lie derivative of the metric)

$$\mathcal{L}_\xi g_{\mu\nu} = \nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu \tag{12.1}$$

For local isometries this then leads to the following:

Theorem 12.2 (Killing equation)

$$\mathcal{L}_\xi g_{\mu\nu} = 0 \quad \text{or} \quad \nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu = 0 \tag{12.2}$$

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 12.1 (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 12.1 Linear combinations of Killing vectors are also Killing vectors.

Quote 12.2 Killing is not killing the metric. It is named after Wilhelm Killing.

Betti Hartmann, 19 January 2024

[1] Oh, no, you don't!

Figure 12.1: Wilhelm Killing, c. 1889-1891.

Note 12.1 Remember that the Killing vector is tensorial. Switching its indices would require applying the metric.

Quote 12.3 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 12.2 (Maximally symmetric space) A *maximally symmetric space* observes

$$R_{\mu\nu} = \frac{R}{d} g_{\mu\nu} \tag{12.3}$$

where the Ricci scalar R is constant for $d \geq 2$. i.e.

$$R_{\mu\nu} \propto g_{\mu\nu} \tag{12.4}$$

Theorem 12.3 (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 12.2 An example would be Minkowski space, which, in 4D, has $\frac{4 \cdot 5}{2} = 10$ and is hence maximally symmetric and homogeneous.

Theorem 12.4 (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 12.1 Give an example of metric such that $R_{\mu\nu} = cg_{\mu\nu}$ for some constant c, but which is not maximally symmetric.

12.2 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 12.5 (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 12.3 (Conformally-adjusted hypersurface metric)

$$ds^2 = \frac{1}{\left(1 + \frac{kr^2}{4}\right)^2} \left(dr^2 + \tilde{r}^2 d\Omega_2^2\right) \tag{12.5}$$

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

$$\rho = r\left(1 + \frac{kr^2}{4}\right)^{-1} \tag{12.6}$$

which gives

$$ds^2 = \frac{d\rho^2}{1 - k\rho^2} + \rho^2 d\Omega^2 \tag{12.7}$$

But what is this k? This can be investigated by solving for the spatial part of the Ricci tensor, which is

$$R_{\text{spatial}} = \frac{6k}{a^2(t)} \tag{12.8}$$

Already, the Ricci scalar is understood to be the scalar curvature. As such, we conclude the physical significance of k:

Definition 12.4 (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 12.1 (3-sphere) Considering the 3-sphere and for simplicity, setting $k = 1$, we impose another coordinate transformation

$$\rho = \sin\chi \tag{12.9}$$

[2] If you have read Part II, you will be able to recall this.

[3] We do this in GR exercises all the time.

Definition 12.5 (Metric for a constant time 3-sphere)

$$ds^2 = d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta d\phi^2) \tag{12.10}$$

Exercise 12.2 Use the following coordinate transformations

$$x = \sin\chi\sin\theta\cos\phi \quad y = \sin\chi\sin\theta\sin\phi \quad z = \sin\chi\cos\theta \quad w = \cos\chi \tag{12.11}$$

to find that this metric is indeed that of a 3-sphere.

Derivation 12.2 (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.

$$ds^2 = \frac{d\rho^2}{1 - k\rho^2} + \rho^2 d\Omega^2 \tag{12.12}$$

Derivation 12.3 (Hyperboloid) Here we have the so-called *hyperboloid* or *hyperbolic space* $\mathbb{H}^3$. We can impose a hyperbolic coordinate transformation

$$\rho = \sinh\chi \tag{12.13}$$

We can insert this into the metric for a 4D constant time hypersurface, which is identical to (12.10) except for the addition of the time coordinate. From here, we can write down the hyperbolid metric:

$$ds^2 = -dt^2 + a(t)^2\left[d\chi^2 + S_k(\chi)^2(d\theta^2 + \sin^2\theta d\phi^2)\right] \tag{12.14}$$

where $S_k(\chi)$ represents the spatial curvature term and observes

$$S_k(\chi) = \begin{cases} \sin\chi & \text{for } k = +1 \\ \chi & \text{for } k = 0 \\ \sinh\chi & \text{for } k = -1 \end{cases} \tag{12.15}$$

12.3 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 12.6 (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 (12.14) in terms of k, we have

Definition 12.6 (FLRW metric) In isotropic coordinates, the FLRW metric is

$$ds^2 = -dt^2 + a^2(t)\frac{dx^2 + dy^2 + dz^2}{\left(1 + \frac{k}{4}(x^2 + y^2 + z^2)\right)^2} \tag{12.16}$$

In polar coordinates, it is

$$ds^2 = -dt^2 + a(t)^2 \left[\frac{dr^2}{1 - kr^2} + r^2 d\Omega^2 \right] \tag{12.17}$$

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 13

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.

13.1 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 13.1 (Friedmann equations) By substitution, one can find the Einstein tensor components corresponding to the FLRW metric as:

$$G^t_t = -9\frac{\dot{a}^2}{a^2} - 9\frac{k}{a^2} \quad G^x_x = G^y_y = G^z_z = -2\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} - \frac{k}{a^2} \tag{13.1}$$

Inserting these results to the field equations with the cosmological term, as seen in (7.28) and the stress-energy tensor for a perfect fluid in (5.14) gives the first two Friedmann equations:

$$-3\frac{\dot{a}^2}{a^2} - 3\frac{k}{a^2} + \Lambda = -8\pi\rho \tag{13.2}$$

$$2\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2} - \frac{k}{a^2} + \Lambda = 8\pi p \tag{13.3}$$

The third equation is slightly tricky. We first differentiate (13.2):

$$2\left(\frac{\dot{a}}{a}\right)\left(\frac{\ddot{a}}{a} - \frac{\dot{a}^2}{a^2}\right) = \frac{8\pi}{3}\dot{\rho} + 2\frac{k}{a^2}\frac{\dot{a}}{a} \tag{13.4}$$

Now we want to remove the term $\frac{\ddot{a}}{a}$. To do so, we can add (13.2) and (13.3), which yields

$$-\frac{\ddot{a}}{a} = \frac{4\pi}{3}(\rho + 3p) - \frac{\Lambda}{3} \tag{13.5}$$

Substitute $\frac{\ddot{a}}{a}$ with this result, and we find the third equation:

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0 \tag{13.6}$$

To sum up:

Theorem 13.1 (Friedmann equations) Field equations:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi\rho}{3} + \frac{\Lambda}{3} - \frac{k}{a^2} \tag{13.7}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi}{3}(\rho + 3p) + \frac{\Lambda}{3} \tag{13.8}$$

Conservation equation:

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + p) = 0 \tag{13.9}$$

13.2 Equation of state and beginnings of domination

Theorem 13.2 (Equation of state) Perfect fluids observe the so-called *equation of state*:

$$p = w\rho \tag{13.10}$$

where w is a constant called the *equation of state parameter*.

Derivation 13.2 (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 (13.9) 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

$$\rho(t) \propto 1/a^3(t) \tag{13.11}$$

- $w = 1/3$: radiation. This leads to $p = \rho/3$ and

$$\rho(t) \propto 1/a^4(t) \tag{13.12}$$

- $w = -1/3$: curvature. This leads to $p = -\rho/3$ and

$$\rho(t) \propto 1/a^2(t) \tag{13.13}$$

- $w = 1$: stiff matter[a]. This leads to $p = \rho$ and

$$\rho(t) \propto 1/a^6(t) \tag{13.14}$$

  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

$$\rho = \text{constant} \tag{13.15}$$

[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.

13.3 Energy-momentum conservation

The definition of perfect fluids can be adapted for cosmology.

Definition 13.1 (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 (5.14). In comoving coordinates, one has $u = (1, 0, 0, 0)$. As such, the stress-energy tensor of a perfect fluid in comoving coordinstes is

$$T_{\mu\nu} = \rho u_\mu u_\nu = \mathrm{diag}(\rho, p, p, p) \quad \text{or} \quad T^\mu_\nu = \mathrm{diag}(-\rho, p, p, p) \tag{13.16}$$

Remark 13.1 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 (7.25). If $\Lambda = 0$, the same applies to the stress-energy tensor:

$$\nabla_\mu T^{\mu\nu} = 0 \tag{13.17}$$

The implication? We then recall our previous derivation:

$$\dot{\rho} - 3H(\rho + p) = 0 \tag{13.18}$$

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 13.2 (Proper volume)

$$V(t) = a^3 v \tag{13.19}$$

which is the 'scaled' version of the volume.

Differentiation by t yields

$$\frac{dV}{dt} = 3a^2\dot{a} = 3HV \tag{13.20}$$

Note 13.1 If we insert this into the definition of entropy, we will find that instead of energy, the property conserved is *entropy*.

Remark 13.2 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 13.3 (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.

$$\rho T_{\mu\nu} V^\mu V^\nu \geq 0 \tag{13.21}$$

where V is a time-like vector.

> **Quote 13.1** 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.

13.4 History and ultimate fate of the universe

Quote 13.2 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

$$H^2 = \frac{8\pi\rho}{3} + \frac{\Lambda}{3} - \frac{k}{a^2} \to 1 = \frac{8\pi\rho}{3H^2} + \frac{\Lambda}{3H^2} - \frac{k}{a^2H^2} \tag{13.22}$$

We can thus define a series of parameters.

Definition 13.3 (Density parameter) The *density parameters* are

$$\underbrace{\Omega_m = \frac{\rho_m}{\rho_c} = \frac{8\pi\rho}{3H^2}}_{\text{matter}} \quad \underbrace{\Omega_r = \frac{\rho_r}{\rho_c}}_{\text{radiation}} \quad \underbrace{\Omega_k = \frac{\rho_k}{\rho_c} = \frac{k}{a^2H^2}}_{\text{curvature}} \quad \underbrace{\Omega_\Lambda = \frac{\rho_\Lambda}{\rho_c} = \frac{\Lambda}{3H^2}}_{\text{dark energy}} \tag{13.23}$$

where we have the *critical density*

$$\rho_c = \frac{3H^2}{8\pi G} - \frac{3H^2}{8\pi} \tag{13.24}$$

which is the density that leads to a spherically flat universe. G is omitted as usual.

Remark 13.3 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:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}(\rho_m + \rho_r - \rho_k + \rho_\Lambda) \tag{13.25}$$

and find rather miraculously that

Theorem 13.4 (Density parameters)

$$\Omega_m + \Omega_r - \Omega_k + \Omega_\Lambda = 1 \tag{13.26}$$

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 13.5 (Domination in various eras) Solving for the proportionality constant in each era, we have

$$\dot{a}^2 = \frac{C_0}{a} + \frac{C_{1/3}}{a^2} - k + \frac{\Lambda}{3}a^2 \tag{13.27}$$

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 13.4 (Various scale factors)

$$a(t) = a_0 \left(\frac{t}{t_0}\right)^{1/2} \quad a(t) = a_0 \left(\frac{t}{t_0}\right)^{2/3} \quad a(t) = a_0 \left(\frac{t}{t_0}\right) \quad a(t) = a_0 \exp\left[\pm\sqrt{\frac{\Lambda}{3}}(t - t_0)\right] \tag{13.28}$$

Note 13.2 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 13.4 When dark energy dominates, we can infer that

$$\frac{\dot{a}}{a} = H = \pm\sqrt{\frac{\Lambda}{3}} \tag{13.29}$$

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:

$$\frac{k}{a^2 H^2} = \Omega - 1 \tag{13.30}$$

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 13.6 (Time evolution of the density parameter)

$$\Omega(t) = 1 + \frac{k}{\dot{a}^2(t)} \tag{13.31}$$

> **Quote 13.3** 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.

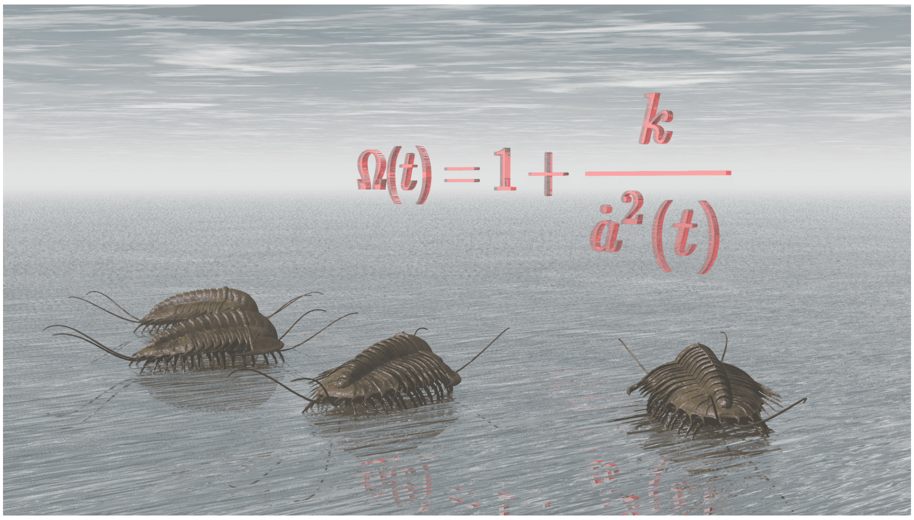


Figure 13.1: Trilobites discuss the density parameter (3D trilobite by Thorsten Brand).

We can also investigate if the expansion of the universe is accelerating or otherwise from the so-called *deceleration parameter*:

Definition 13.5 (Deceleration parameter)

$$q(t) = \frac{a(t)\ddot{a}(t)}{\dot{a}^2(t)} \tag{13.32}$$

If we insert this into the last Friedmann equation, we get

$$q = \frac{4\pi}{3H^2}(\rho + 3p) \tag{13.33}$$

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 13.5 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 $\ddot{a} < 0$.

13.5 Age of the universe

Naively, we might be tempted to insert $H_0 = 2.3 \times 10^{-18}$ into

$$t_0 = \frac{a_0 t}{a(t)} \tag{13.34}$$

which gives $t_0 = 4.4 \times 10^{18}$.

Quote 13.4 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 13.3 (Age of the universe) Recalling (13.26), one can write down the expression

$$\frac{H^2}{H_0^2} = \frac{H^2}{H_0^2}\left(\Omega_m + \Omega_r - \Omega_k + \Omega_\Lambda\right) \tag{13.35}$$

Meanwhile, we note that

$$\frac{H^2}{H_0^2} = \left(\frac{a_0}{a}\right)^{3(1+w)} \tag{13.36}$$

Inserting this and taking the total derivative, we find

$$dt = \frac{da}{aH_0}\left[\left(\frac{a_0}{a}\right)^3 \Omega_m + \left(\frac{a_0}{a}\right)^4 \Omega_r - \left(\frac{a_0}{a}\right)^2 \Omega_k + \Omega_\Lambda\right]^{-1/2} \tag{13.37}$$

We consider both cases of the flat universe.

- **Matter only**

$$\Omega_r = \Omega_k = \Omega_\Lambda = 0 \quad \Omega_m = 1 \tag{13.38}$$

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.

- **Matter and cosmological constant**

$$\Omega_r = \Omega_k = 0 \quad \Omega_m = 1 - \Omega_\Lambda \quad \Omega_\Lambda \neq 0 \tag{13.39}$$

In this case, we find that the age of the universe is about 13 billion years old, a much better value.

Exercise 13.1 Calculate the age of the universe for the matter and radiation-dominated cases.

Quote 13.5 The radiation in the current era is negligible.

Betti Hartmann, 20 February 2024

13.6 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 13.6 (Redshift in cosmology)

$$z = \frac{a(t_{\text{obs}})}{a(t)} - 1 \tag{13.40}$$

Taking the total derivative of this expression gives

$$dz = -\frac{a(t_{\text{obs}})}{a(t)^2}\dot{a}dt = -\frac{a(t_{\text{obs}})}{a(t)}Hdt = (1+z)Hdt \tag{13.41}$$

Integrating with respect to t:

Theorem 13.7 (Signals)

$$t_{\text{obs}} - t_{\text{e}} = \int_0^z \frac{1}{(1+z')H}dz' \tag{13.42}$$

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

$$dr = \frac{da}{a^2H} \tag{13.43}$$

This can then be integrated to find the *distance-redshift relation.*

Theorem 13.8 (Distance-redshift relation)

$$r(z) = \frac{1}{a_0 H_0} \int_0^z dz \left[(1+z)^3\Omega_m + (1+z)^4\Omega_r + \Omega_\Lambda\right]^{-1/2} \tag{13.44}$$

13.7 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 13.7 (Conformal time) As the name suggests, the *conformal time* is time with the scale factor taken into consideration

$$d\eta = dt/a(t) \tag{13.45}$$

The FLRW model contains two horizons:

Derivation 13.4 (Particle horizon) We begin with the FLRW metric in isotropic coordinates:

$$ds^2 = -dt^2 + a(t)^2 \left[\frac{dr^2}{1-kr^2} + r^2 d\Omega^2\right] \tag{13.46}$$

For a photon, the metric reduces to

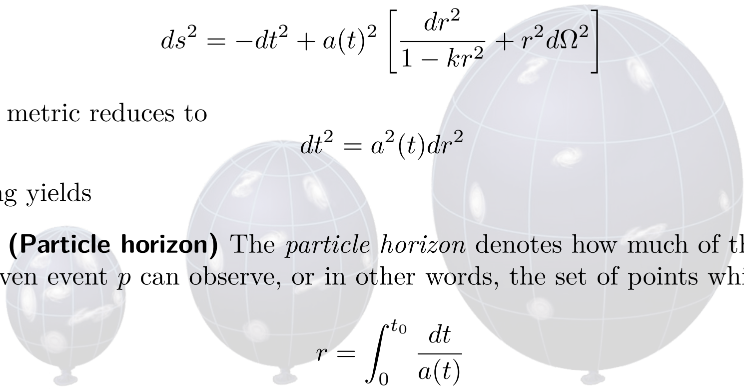
$$dt^2 = a^2(t)dr^2 \tag{13.47}$$

Further simplifying yields

Definition 13.8 (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.

$$r = \int_0^{t_0} \frac{dt}{a(t)} \tag{13.48}$$

where r is the particle horizon.



Remark 13.6 In terms of the comoving distance, the particle horizon is equal to the conformal time η that has passed since the Big Bang.

Definition 13.9 (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:

$$r_\gamma(t) = \eta(\infty) = \eta(t) \tag{13.49}$$

[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 13.2 Prove these horizon results.

Definition 13.10 (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

$$D = \frac{H^{-1}}{a(t)} = -\frac{\ddot{a}(t)}{\dot{a}^2(t)} \tag{13.50}$$

Two regions greater than this length cannot communicate with each other.

[5]Although you can make as convincing a case that it has created more problems than it has solved.

Remark 13.7 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 13.11 (Apparent horizon) The *apparent* is the distance at which galaxies move away from us with the speed of light c.

$$R_H(t) = \frac{c}{H(t)} \tag{13.51}$$

Now we take a few pages off as an intermission, and discuss some exact solutions of Friedmann equations.

13.8 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

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}\rho - \frac{k}{a^2} \quad \frac{\ddot{a}}{a} = -\frac{4\pi}{3}\rho \tag{13.52}$$

while the conservation equation stays the same. Integrating it yields

$$\rho = \frac{\rho_0}{a^3} \tag{13.53}$$

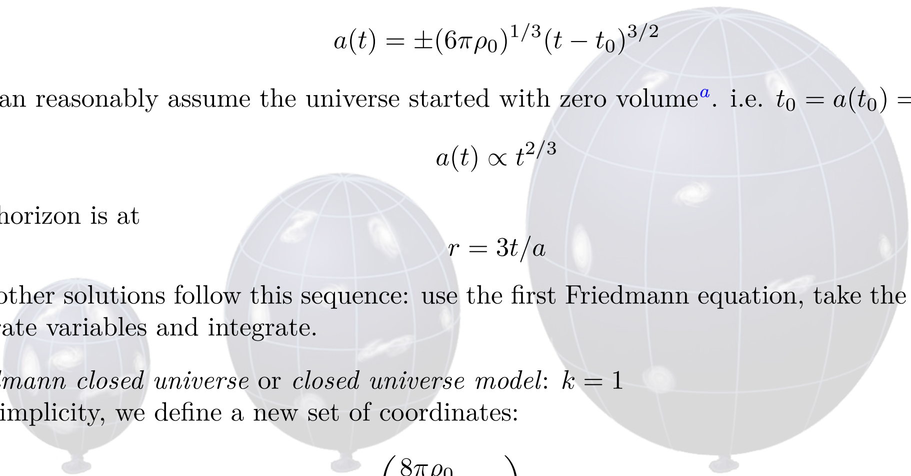
Derivation 13.5 (Matter-only Friedmann solutions) We first consider the matter-only case:

- *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

$$dt = \pm\frac{1}{\sqrt{\frac{8\pi}{3}\rho_0}}\sqrt{a}da \tag{13.54}$$

  Integrate and we find the scale factor:

$$a(t) = \pm(6\pi\rho_0)^{1/3}(t - t_0)^{3/2} \tag{13.55}$$



  We can reasonably assume the universe started with zero volume[a]. i.e. $t_0 = a(t_0) = 0$

$$a(t) \propto t^{2/3} \tag{13.56}$$

  The horizon is at

$$r = 3t/a \tag{13.57}$$

  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:

$$\left(\frac{8\pi\rho_0}{3a} - 1\right) = \tan u \tag{13.58}$$

  Subbing this in, we can find

$$a(u) = \frac{4\pi}{3}\rho_0(1 - \cos u) \tag{13.59}$$

  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

$$a(u) = \frac{4\pi\rho_0}{3}(\cosh 2u - 1) \tag{13.60}$$

  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 13.6 (Radiation-only Friedmann solutions) Now we consider the radiation-only case:

- *Lemaître-Tolman metric*: $k = 0$

$$a(t) \propto t^{1/2} \quad a(t) = \sqrt{2}(8\pi\rho_0/3)^{1/4}t^{1/2} \tag{13.61}$$

  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

$$r = 2t/a \tag{13.62}$$

Derivation 13.7 (Curvature-only Friedmann solutions) Now we consider the curvature-only case:

- *Milne model*: $k = -1$

$$a(t) \propto t \quad a(t) = H_0\sqrt{\Omega_k}t \tag{13.63}$$

  As mentioned previously, there is no curvature-dominated era, and the Milne model is inconsistent with cosmological observations and thus unphysical.

13.9 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 13.8 (Friedmann-Lemaître solutions) We likewise consider possible curvature parameters:

- *Einstein's static universe*: $k = 1$
  Remark 13.8 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*:

$$a = \frac{1}{\sqrt{4\pi G\left(\rho_m + \rho_r + p_\lambda\right)}} \tag{13.64}$$

  As a is constant, its derivative of any order is zero. Using this to our advantage in the Friedmann equations, we can derive that

$$\Lambda = 4\pi(\rho + 3p) \tag{13.65}$$

  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

$$a(t) = e^{\sqrt{\Lambda/3}t} \tag{13.66}$$

  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

$$ds^2 = -dv^2 + dw + 2 + dx^2 + dy^2 + dz^2 \tag{13.67}$$

This can be accomplished by a series of coordinate transformations or *slices* which we call the *de Sitter foliations*:

Definition 13.12 (De Sitter foliations)

$$\begin{aligned} v &= \sqrt{\frac{3}{\Lambda}} \sinh \sqrt{\frac{\Lambda}{3}} t \\ w &= \sqrt{\frac{3}{\Lambda}} \cosh \sqrt{\frac{\Lambda}{3}} t \cos \chi \\ x &= \sqrt{\frac{3}{\Lambda}} \cosh \sqrt{\frac{\Lambda}{3}} t \sin \chi \cos \theta \\ y &= \sqrt{\frac{3}{\Lambda}} \cosh \sqrt{\frac{\Lambda}{3}} t \sin \chi \sin \theta \cos \phi \\ z &= \sqrt{\frac{3}{\Lambda}} \cosh \sqrt{\frac{\Lambda}{3}} t \sin \chi \sin \theta \sin \phi \end{aligned} \tag{13.68}$$

Through the coordinate transformations

$$T = \sqrt{\frac{3}{\Lambda}} \log \left[\frac{v+w}{\sqrt{3/\Lambda}}\right] \quad X = \sqrt{\frac{3}{\Lambda}} \frac{x}{v+w} \quad Y = \sqrt{\frac{3}{\Lambda}} \frac{y}{v+w} \quad Z = \sqrt{\frac{3}{\Lambda}} \frac{z}{v+w} \tag{13.69}$$

the metric in so-called static coordinates can be recovered. For the sake of familiarity, we can rewrite capitals to small letters:

$$ds^2 = -dt^2 + e^{2\sqrt{\Lambda/3}t}(dx^2 + dy^2 + dz^2) \tag{13.70}$$

Remark 13.9 We can do the same thing in spaces with zero and negative curvature, both of which yield 5 slices each as well.

Derivation 13.9 (Staticity of the de Sitter solution) Consider a time translation to the de Sitter solution $t \to t' + T$. The metric then becomes

$$ds^2 = -dt'^2 + e^{2\sqrt{\Lambda/3}(t'+T)}(dz^2 + dy^2 + dz^2) \tag{13.71}$$

Only the spatial part is rescaled. We use a rather dirty trick and impose a second rescaling

$$X'^\mu = e^{2\sqrt{\Lambda/3}T} X^\mu \tag{13.72}$$

and everything returns to

$$ds^2 = -dt'^2 + e^{2\sqrt{\Lambda/3}t'}(dz'^2 + dy'^2 + dz'^2) \tag{13.73}$$

13.10 Solutions without matter & radiation

These solutions are also known as the *vacuum solutions*.

Derivation 13.10 (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*: $\Lambda > 0$

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{\Lambda}{3} - \frac{k}{a^2} \to \dot{a}^2 = \underbrace{\frac{\Lambda}{3}a^2 - k}_{\geq 0} \tag{13.74}$$

for which we have

- $k = 1$

$$a^2 = \frac{\Lambda}{3} \pm 1 \to a(t) = \sqrt{\frac{3}{\Lambda}} \cosh\left(\pm\sqrt{\frac{\Lambda}{3}}t\right) \tag{13.75}$$

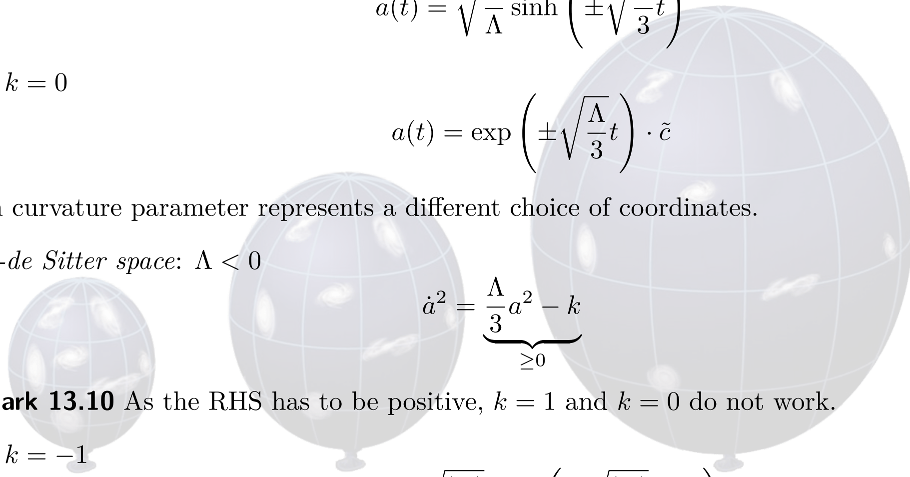
- $k = -1$

$$a(t) = \sqrt{\frac{3}{\Lambda}} \sinh\left(\pm\sqrt{\frac{\Lambda}{3}}t\right) \tag{13.76}$$

- $k = 0$

$$a(t) = \exp\left(\pm\sqrt{\frac{\Lambda}{3}}t\right) \cdot \tilde{c} \tag{13.77}$$

Each curvature parameter represents a different choice of coordinates.



- *Anti-de Sitter space*: $\Lambda < 0$

$$\dot{a}^2 = \underbrace{\frac{\Lambda}{3}a^2 - k}_{\geq 0} \tag{13.78}$$

Remark 13.10 As the RHS has to be positive, $k = 1$ and $k = 0$ do not work.

- $k = -1$

$$a(t) = \sqrt{\frac{|\Lambda|}{3}} \sinh\left(\pm\sqrt{\frac{|\Lambda|}{3}}t + c\right) \tag{13.79}$$

Remark 13.11 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 13.12 True to their name, all three spaces are solutions to the Einstein equations $G_{\mu\nu} = \Lambda g_{\mu\nu} = 0$.

Figure 13.2: De Sitter standing

Chapter 14

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.

14.1 Flatness problem

This is also known as the *oldness problem*. Solutions using the FLRW metric tell us that the universe is spatially flat

$$k = 0 \rightarrow \rho = \rho_c \rightarrow \Omega_0 = 1 \tag{14.1}$$

Observations give, with generous bounds:

$$0.9 \leq \Omega_0 \leq 1.1 \tag{14.2}$$

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 14.1 (Matter-dominated universe) We consider the matter-dominated case, in which

$$a \propto t^{2/3} \quad \text{and} \quad H \propto 1/t \rightarrow aH \propto t^{-1/3} \tag{14.3}$$

and therefore

$$|\Omega - 1| \propto t^{2/3} \tag{14.4}$$

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 14.1** In physics, this is called fine-tuning, and not very natural.
>
> *Betti Hartmann, 23 February 2024*

Derivation 14.2 (Radiation-dominated universe) We consider the radiation-dominated case, in which

$$a \propto t^{1/2} \quad \text{and} \quad H \propto 1/t \rightarrow aH \propto t^{-1/2} \tag{14.5}$$

and therefore

$$|\Omega - 1| \propto t \tag{14.6}$$

We know that the age of the universe is $t_a \approx 10^{17}$s and that the Planck time is $t_{\text{Pl}} \approx 10^{-43}s$. This results in

$$t/t_0 \approx 10^{-60} \tag{14.7}$$

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 14.1 The universe being spatially flat is not a 'natural' choice. It requires extreme fine-tuning in a very hot, very dense early universe!

14.2 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 (13.7):

Definition 14.1 (Proper distance)

$$R_p(t_2) = a(t_2)\int_{t_1}^{t_2} \frac{dt}{a(t)} \tag{14.8}$$

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 $\beta = 1/2$.

$$R_p(t_2) \sim \frac{1}{1-\beta} t_2^\beta (t_2^{1-\beta} - t_1^{1-\beta}) \tag{14.9}$$

The time since the Big Bang ($t_1 = 0$) until the decoupling of matter (formation of CMB) $t_2 = t_{\text{ls}}$ (last scattering of photons). This was radiation-dominated ($\beta = 1/2$). i.e. regions separated by distances that are no larger than $2t_{\text{ls}}$ were never in causal contact.

$$R_p(t_{\text{ls}}) \sim 2t_{\text{ls}} \tag{14.10}$$

Remark 14.2 That is to say, the causal patch has expanded due to the expansion of the universe.
Its size is given by

$$\frac{a(t_0)}{a(t_{\text{ls}})} R_p(t_{\text{ls}}) = 1 + Z_{\text{ls}} R_p(t_{\text{ls}}) \tag{14.11}$$

The distance travelled by photons of the CMB since $t = t_{\text{ls}}$ until today is the integral from before $t_1 = t_{\text{ls}}$, $t_2 = t_0$:

$$R_p(t_0) = a(t_0)\int_{t_{\text{ls}}}^{t_0} \frac{dt}{a(t)} \tag{14.12}$$

We assume that the universe is matter-dominated ($\beta = 2/3$):

$$R_p(t_0) \sim 3t_0^{2/3}(t_0^{1/3} - t_{\text{ls}}^{1/3}) \sim 3t_0 \quad \text{for } t_0 \ll t_{\text{ls}} \tag{14.13}$$

The light cone angle θ is then

$$\theta \approx \frac{(1+z_{\text{ls}})R_p(t_{\text{ls}})}{R_p(t_{\text{ls}})} \approx \frac{(1+z_{\text{ls}})2t_{\text{ls}}}{3t_0} \approx 0.02 \approx 1.15° \tag{14.14}$$

This means that, looking at patches of the CMB, under angles $> 1.15°$, we see photons that have not been at causal contact at the moment of last scattering is formation of the CMB.

Note 14.1 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?

14.3 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 14.2 ...the universe would have recollapsed!

Remark 14.3 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 15

Inflation

Quote 15.1 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.

15.1 Cosmological inflation

Inflation is essentially accelerated expansion of space. It can be formally defined as

Theorem 15.1 (Inflation) Inflation occurs when

$$\ddot{a} > 0 \tag{15.1}$$

or if the derivative of the so-called *comoving Hubble length* $H^{-1}/a(t)$ is negative:

$$\frac{d}{dt}\frac{H^{-1}}{a(t)} = \underbrace{\frac{d}{dt}\frac{1}{\dot{a}(t)}}_{①} = -\frac{\ddot{a}(t)}{(\dot{a})^2(t)} < 0 \tag{15.2}$$

Remark 15.1 The comoving Hubble length is the size of the casual patch and the two are conceptually identical.

Quote 15.2 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 15.1, 15 April 2024

Remark 15.2 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 15.3 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 15.2, 15 April 2024

In other words:

Quote 15.4 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 15.2, 15 April 2024

15.2 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.

15.3 Scalar fields

As before, the starting point is the FLRW metric as seen in (12.17). We define a new real scalar field, or a so-called *inflation field*, which dominates over all other energy-momentum content during inflation.

Definition 15.1 (Lagrangian density of a real scalar field) The Lagrangian density for the real scalar field $\phi(t) \in \mathbb{R}$ is

$$\mathcal{L}_\phi = \underbrace{-\frac{1}{2}g^{\mu\nu}\partial_\mu\phi\partial_\nu\phi}_{\text{kinetic + gradient}} - \underbrace{V(\phi)}_{\text{potential}} \tag{15.3}$$

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 (**??**). Thus, for $\phi = \phi(t) \in \mathbb{R}$ abd the FLRW metric, we have

$$\mathcal{L}_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi) \quad T_{00} = \frac{1}{2}\dot{\phi}^2 + V(\phi) \tag{15.4}$$

The other diagonal components of the energy-momentum tensor are

$$T_{xx} = T_{yy} = T_{zz} = \left(\frac{1}{2}\dot{\phi}^2 - V(\phi)\right)a^2 \tag{15.5}$$

Noting that the components are simpny density and pressure, we can write

Definition 15.2 (Stress and press under a scalar field)

$$\rho_\phi = \frac{1}{2}\dot{\phi}^2 + V(\phi) \quad p_\phi = \frac{1}{2}\dot{\phi}^2 - V(\phi) \tag{15.6}$$

We then consider the dynamics of the scalar field. By inserting 15.4 into the Euler-Lagrange equations seen in (4.25), we find that

$$\ddot{\phi} + \underbrace{\frac{3\dot{a}}{a}\dot{\phi}}_{\text{friction term}} + \frac{dV}{d\phi} = 0 \tag{15.7}$$

The dynamics of an FLRW universe filled solely with a real scalar field $\phi(t) \in \mathbb{R}$ is thus as follows:

Theorem 15.2 (Inflationary Friedmann equations)

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi}{3}\mathcal{L}_\phi = \frac{8\pi}{3}\left(\frac{1}{2}\dot{\phi}^2 - V(\phi)\right) \tag{15.8}$$

$$-\frac{\ddot{a}}{a} = \frac{4\pi}{3}(\rho_\phi + 3p_\phi) = \frac{8\pi G}{3}((\dot{\phi})^2 - V(\phi)) \tag{15.9}$$

[1]See previous bullet point.

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} + \frac{dV}{d\phi} = 0 \tag{15.10}$$

Remark 15.3 We do not really use (15.10). Nonetheless, it is there for completeness.

> **Quote 15.5** Because it's there.
>
> *George Mallory, on his motivation of climbing Mount Everest, 18 March 1923*

15.4 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 15.3 (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

$$\ddot{\phi} \ll 3H\dot{\phi} \quad \dot{\phi}^2 \ll V(\phi) \tag{15.11}$$

Moreover, we use the reduced Planck mass, which substitutes several constants:

Definition 15.4 (Reduced Planck mass) On the scale of the reduced Planck mass, quantum effects of gravity become important.

$$M_{\text{Pl}} = \sqrt{\frac{\hbar c}{8\pi}} = \frac{1}{\sqrt{8\pi}} \quad (\hbar = c = 1) \tag{15.12}$$

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 15.3 (Slow roll Friedmann equations)

$$3H\dot{\phi} \approx -\frac{dV}{d\phi} \quad H^2 \approx \frac{8\pi}{3}V(\phi) = \frac{1}{3M_{\text{Pl}}^2}V(\phi) \tag{15.13}$$

Remark 15.4 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 15.1 (Slow roll conditions) We want to write the slow roll conditions in terms of the potential only:

- 1[st] slow roll condition:

$$H^2 \approx \frac{V(\phi)}{3M_{\text{Pl}}} \tag{15.14}$$

  We take the time derivative:

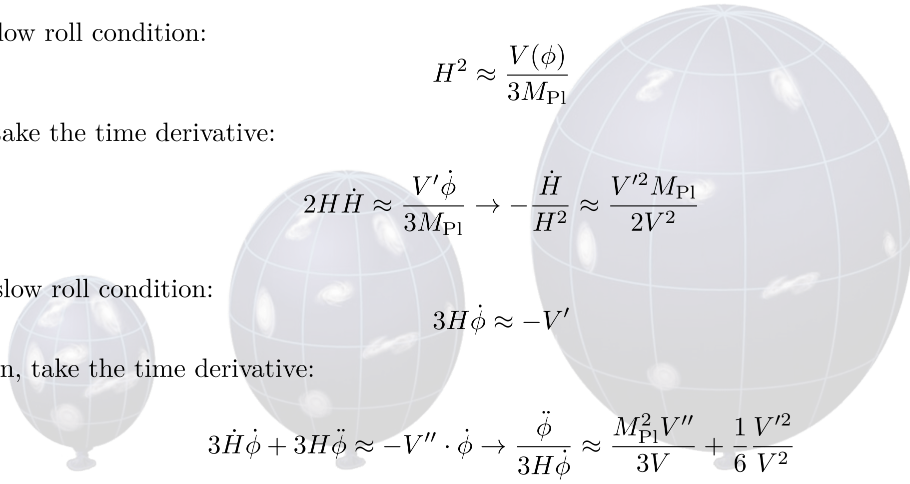
$$2H\dot{H} \approx \frac{V'\dot{\phi}}{3M_{\text{Pl}}} \rightarrow -\frac{\dot{H}}{H^2} \approx \frac{V'^2 M_{\text{Pl}}}{2V^2} \tag{15.15}$$

- 2[nd] slow roll condition:

$$3H\dot{\phi} \approx -V' \tag{15.16}$$

  Again, take the time derivative:

$$3\dot{H}\dot{\phi} + 3H\ddot{\phi} \approx -V'' \cdot \dot{\phi} \rightarrow \frac{\ddot{\phi}}{3H\dot{\phi}} \approx \frac{M_{\text{Pl}}^2 V''}{3V} + \frac{1}{6}\frac{V'^2}{V^2} \tag{15.17}$$



In summary:

[2]The reason behind whose naming will soon become apparent.

Definition 15.5 (Slow roll conditions, roll 2)

$$-\frac{\dot{H}}{H^2} \approx \frac{V'^2 M_{\text{Pl}}}{2V^2} \quad \frac{\ddot{\phi}}{3H\dot{\phi}} \approx \frac{M_{\text{Pl}}^2 V''}{3V} + \frac{1}{6}\frac{V'^2}{V^2} \tag{15.18}$$

This allows us to consider inflation. We can write the second derivative of a in terms of the Hubble constant:

$$H = \frac{\dot{a}}{a} \to \ddot{a} = (\dot{H} + H^2) \tag{15.19}$$

During the very rapidly accelerated expansion era, we want that $\ddot{a} \ll 0$. Thus we have

$$\dot{H} + H^2 \ll 0 \to -\frac{\dot{H}}{H^2} \ll 1 \to \frac{V'^2}{2V} M_{\text{Pl}}^2 \ll 1 \tag{15.20}$$

Definition 15.6 (Slow roll parameters)

$$\epsilon = \frac{1}{2} M_{\text{Pl}}^2 \frac{(V')^2}{V^2} \quad \eta = M_{\text{Pl}}^2 \frac{V''}{V} \tag{15.21}$$

Remark 15.5 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 15.6 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.

15.5 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 15.4 (End of inflation conditions)

$$\epsilon \approx 1 \quad \eta \approx 1 \tag{15.22}$$

Theorem 15.5 (E-folding) The number of e-foldings $\mathcal{N}$ is given by

$$\mathcal{N}(t) = \log\left(\frac{a(t_{\text{end}})}{a(t)}\right) \tag{15.23}$$

where t_{end} is the time that inflation ends.

We thus have

$$a(t_{\text{end}}) = e^{\mathcal{N}} a(t) \tag{15.24}$$

Remark 15.7 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 15.8 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 15.2 (End-of-inflation case) In the end-of-inflation case:

$$H = \frac{\dot{a}}{a} = \frac{d}{dt}(\log a(t)) \tag{15.25}$$

$$\mathcal{N} = \log \frac{a(t_{\text{end}})}{a(t)} = \int_t^{t_{\text{end}}} H(t)dt \tag{15.26}$$

[3]The slow roll parameters have risen!

Now recall the first slow roll Friedmann equations seen in (15.13). One can rewrite this as

$$3H\dot{\phi} \approx -V' \to 3H\frac{d\phi}{dt} \approx -\frac{dV}{d\phi} \to dt = \frac{3Hd\phi}{-V'} \tag{15.27}$$

By inserting this dt into 15.26

$$\mathcal{N}(t) = \frac{1}{M_{\text{Pl}}}\int_{\phi_{\text{end}}}^{\phi} \frac{V}{V'}d\phi \tag{15.28}$$

Derivation 15.3 (Scalar field potential) We introduce a massive scalar field. The potential, which then couples to mass, has the standard form

$$V = \frac{1}{2}m^2\phi^2 \tag{15.29}$$

> **Quote 15.6** 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

$$V' = m^2\phi \quad V'' = m^2 \tag{15.30}$$

As such, the slow roll conditions are

$$\epsilon = \eta = \frac{2M_{\text{Pl}}^2}{\phi^2} \tag{15.31}$$

Inflation happens when

$$\phi^2 \ll 2M_{\text{Pl}} \to \phi \ll \sqrt{2}M_{\text{Pl}} \tag{15.32}$$

Inflation ends when

$$\phi = \sqrt{2}M_{\text{Pl}} \tag{15.33}$$

This can be inserted into (15.28). Setting the time interval as t_{init} and t_{end}, we find

$$\mathcal{N}(t_{\text{init}}) = \frac{1}{M_{\text{Pl}}^2}\int_{\phi_{\text{init}}}^{\phi_{\text{end}}} \frac{V}{V'}d\phi = \frac{1}{M_{\text{Pl}}^2}\int_{\phi_{\text{init}}}^{\phi_{\text{end}}} \frac{1}{2}\phi d\phi = \left(\frac{\phi_{\text{init}}}{2M_{\text{Pl}}}\right)^2 - \frac{1}{2} \tag{15.34}$$

When we have around 70 e-foldings:

$$\phi_{\text{init}} = 2M_{\text{Pl}}\sqrt{\frac{141}{2}} \approx 17M_{\text{Pl}} \tag{15.35}$$

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].

Remark 15.9 (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.

[4] Here we see the nominal significance of the e-foldings.

Chapter 16

Emergence of cosmological perturbation theory

> **Quote 16.1** 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.

16.1 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 16.1 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 16.1 (Metric perturbations)

$$g_{\mu\nu} = \underbrace{{}^{(0)}g_{\mu\nu}}_{\text{background FLRW metric}} + \underbrace{{}^{(1)}g_{\mu\nu}}_{\text{linear } (\delta g_{\mu\nu})} + \underbrace{{}^{(2)}g_{\mu\nu} + \ldots}_{\text{non-linear}} \tag{16.1}$$

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.

16.2 Conformal coordinates

We recall *conformal time* defined in (13.45), 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 16.2 (FLRW metric in conformal coordinates)

$$ds^2 = a^2(\eta)[-d\eta^2 + \gamma_{\mu\nu}dX^\mu dX^\nu] \quad \text{where} \quad \gamma_{\mu\nu}dX^\mu dX^\nu = \frac{dx^2+dy^2+dz^2}{\left(1+\frac{k}{4}(z^2+y^2+z^2)\right)^2} \tag{16.2}$$

16.3 Scalar, vector and tensor perturbations

We decompose $\delta g_{\mu\nu}$ into scalar, vector and tensor perturbations:

$$\delta g_{\mu\nu} = \begin{pmatrix} \underbrace{\delta g_{00}}_{1\times 1} & \underbrace{\delta g_{0\nu}}_{1\times 3} \\ \underbrace{\delta g_{\mu 0}}_{3\times 1} & \underbrace{\delta g_{\mu\nu}}_{3\times 3} \end{pmatrix} \tag{16.3}$$

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 16.1 (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:

$$v = \underbrace{-\nabla\phi}_{\text{curl-free}} + \underbrace{\nabla \times A}_{\text{divergence-free}} \tag{16.4}$$

where ∇ is *not* the covariant derivative that produces the cute little Christoffel symbols.

We then arrive at the following parameterisation:

$$\delta g_{0\mu} = \nabla_\mu b + s_\mu \quad \mu = 1,2,3 \tag{16.5}$$

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:

$$\delta g_{\mu\nu} = \nabla_\mu \nabla_\nu e + \frac{1}{2}(\nabla_\mu f_\nu + \nabla_\nu f_\mu) + h_{\mu\nu} \tag{16.6}$$

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 16.2 (Scalar perturbations)

$$\delta^{(s)} g_{\mu\nu} = a^2(\eta) \begin{pmatrix} -2\phi & \nabla_\nu B \\ \nabla_\mu B & -2(\psi\gamma_{\mu\nu} - \nabla_\mu\nabla_\nu E) \end{pmatrix} \tag{16.7}$$

where ϕ, B, ψ and E are scalar functions of space and time.

...the vector perturbations...

Theorem 16.3 (Vector perturbations)

$$\delta^{(v)} g_{\mu\nu} = a^2(\eta) \begin{pmatrix} 0 & -S_\nu \\ -S_\mu & \nabla_\nu F_\mu + \nabla_\mu F_\nu \end{pmatrix} \tag{16.8}$$

[1] Of the *Landau & Lifshitz* fame.

where S_μ and F_μ are 3-vector functions that are divergence-free.

Remark 16.2 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 16.4 (Tensor perturbations)

$$\delta^{(t)} g_{\mu\nu} = a^2(\eta) \begin{pmatrix} 0 & 0 \\ 0 & h_{\mu\nu} \end{pmatrix} \tag{16.9}$$

where $h_{\mu\nu}$ is a symmetric 3-tensor function. It is divergence- and trace-free.

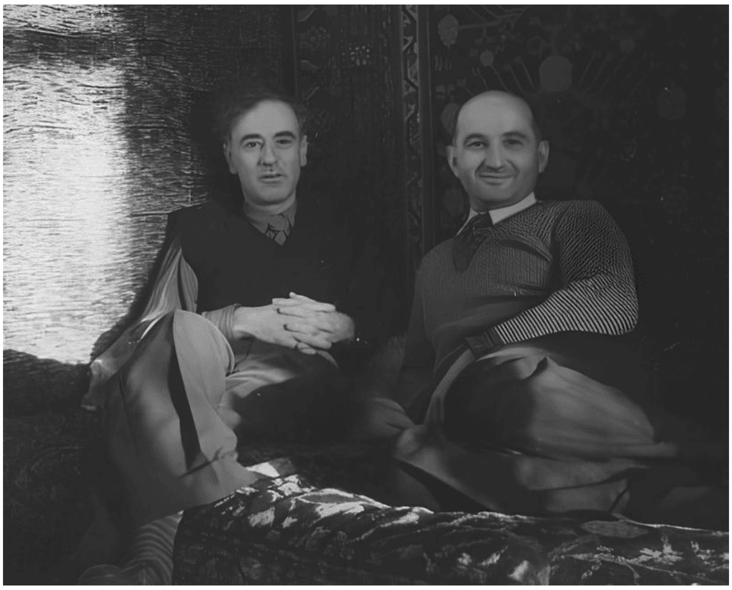


Figure 16.1: 'What are ya looking at, punk?'

Quote 16.2 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 16.3 In the scalar-vector-tensor decomposition the dynamical evaluation of the perturbations are such that the equation for scalar, vector and tensor perturbations decouple!

16.4 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*):

$$\eta \to \tilde{\eta} = \eta + \xi^0(\eta, x^\mu) \quad x^\mu \to \tilde{x}^\mu = x^\mu + \gamma^{\mu\nu} \nabla_\nu \xi(\eta, x^\mu) \quad \mu = 1, 2, 3 \tag{16.10}$$

We will now choose ξ^0 and ξ 'to our convenience' or, in field theory language, 'fix the gauge'.

Remark 16.4 Choosing the coordinate transformations to give in terms ξ^0 and ξ preserves the *scalar* nature of the perturbation.

Derivation 16.1 (Fixing the gauge) We recall, from *Spinors & Symmetries*, the definition of Lie derivatives:

$$(\mathcal{L}g)_{\mu\nu} = g_{\mu\nu}(x) -^{(')} g_{\mu\nu}(x) \tag{16.11}$$

The decomposed metric is given by

$$g_{\mu\nu} = \underbrace{g_{\mu\nu}(x)}_{\text{FLRW metric}} \quad -^{(0)}g_{\mu\nu} + \delta g_{\mu\nu} \tag{16.12}$$

which we insert into the Lie derivative. This yields

$$\mathcal{L}_\xi g_{\mu\nu} = \delta g_{\mu\nu}(x) - \delta^{(')} g_{\mu\nu}(x) \tag{16.13}$$

Using (16.7), it can be found that

$$\delta^{(')} g_{00} = -2a^2\phi + 2aa'\xi^0 + 2a^2(\xi^0) \tag{16.14}$$

From parametrisation, we have $\delta^{(')} g_{00} = 2a^2\tilde{\phi}$. Equating the two yields

$$\tilde{\phi} = \phi - \frac{a'}{a}\xi^0 - (\xi^0)' \tag{16.15}$$

In the same vein, we can derive the scalar functions seen in (16.7):

$$\tilde{\psi} = \psi + \frac{a'}{a}\xi^0 \quad \tilde{E} = E - \xi \quad \tilde{B} = B + \xi^0 - \xi' \tag{16.16}$$

And now, having no more degrees of freedom left, we have fixed the gauge. The metric that includes the scalar perturbations thus reads

Definition 16.3 (Scalar perturbation metric)

$$ds^2 = a^2(\eta)\left[-(1+2\Phi)d\eta^2 + (1-2\Psi)\gamma_{\mu\nu}dx^\mu dx^\nu\right] \tag{16.17}$$

where

$$\Phi = \phi + \frac{1}{a}\left[(B - E')\, a\right]' \quad \Psi = \psi - \frac{a'}{a}(B - E') \tag{16.18}$$

where $\phi(\eta, x, y, z)$ and $\psi(\eta, x, y, z)$ are functions that depend on all coordinates, representing 2 degrees of freedom.

16.5 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:

$$G_{\mu\nu} =^{(0)} G_{\mu\nu} + \delta G_{\mu\nu} \quad g_{\mu\nu} =^{(0)} g_{\mu\nu} + \delta g_{\mu\nu} \quad T_{\mu\nu} =^{(0)} T_{\mu\nu} + \delta T_{\mu\nu} \tag{16.19}$$

This gives, for the background space-time equation:

$$^{(0)}G_{\mu\nu} = 8\pi G^{(0)}T_{\mu\nu} \tag{16.20}$$

and for the the first-order perturbation:

$$\delta G_{\mu\nu} = 8\pi G \delta T_{\mu\nu} \tag{16.21}$$

Using the metric, we can shift an index upwards:

$$\begin{aligned}
\delta G_0^0 &= \frac{2}{a^2}\left[3\frac{a'}{a}\left(\Psi' + \frac{a'}{a}\Phi\right) - \Delta\Psi\right] \\
\delta G_\mu^0 &= -\frac{2}{a^2}\partial_\mu\left(\frac{a'}{a}\Phi + \Psi'\right) \\
\delta G_\mu^\nu &= \frac{2}{a^2}\left[\left(\Psi'' + \frac{a'}{a}(\Phi' + 2\Psi') + 2\frac{a''}{a}\Phi - \frac{a'^2}{a^2}\Phi\right)\delta_\mu^\nu + \frac{1}{2}\delta_\mu^\nu\Delta(\Phi-\Psi) - \frac{1}{2}\partial_\mu\partial^\nu(\Phi-\Psi)\right]
\end{aligned} \tag{16.22}$$

where we have the 3-dimensional Laplace operator

$$\Delta = \partial_{xx} + \partial_{yy} + \partial_{zz} \tag{16.23}$$

Remark 16.5 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:

$$\rho =^{(0)} \rho + \delta\rho \quad p =^{(0)} p + \delta p \tag{16.24}$$

Again, only considering scalar perturbations, the perturbed energy-momentum tensor is

$$\delta T_\mu^\nu = \begin{pmatrix} -\delta\rho & ({}^{(0)}+\rho^{(0)p})a^{-1}\partial^\nu V \\ ({}^{(0)}\rho +^{(0)} p)a\partial_\mu V & \delta p\delta_\mu^\nu - \nabla_\mu\nabla^\nu\sigma \end{pmatrix} \quad \mu,\nu = 1,2,3 \tag{16.25}$$

where V and σ are scalar functions.
Perturbations in the inflation field lead to scalar perturbations:

$$\phi(\eta,x,y,z) = \phi_0(\eta) + \delta\phi(\eta,x,y,z) \tag{16.26}$$

Assuming $\sigma = 0$, we find that $\Phi = \Psi$.

Note 16.1 (Primordial gravitational waves) Tensor perturbations of the energy-momentum tensor of the scalar (inflation) field ϕ give rise to the so-called *primordial gravitational waves*. Unlike 'normal' gravitational waves which emerge from astrophysical processes, primordial gravitational waves are a direct result of the big bang and are very weak in comparison. They are a current topic of research in early universe cosmology.

The perturbed energy-momentum tensor reads

$$\begin{aligned}
\delta T_0^0 &= \frac{1}{a^2}\left(\phi_0'^2\Phi - \phi_0'(\delta\phi)' - \frac{\partial V}{\partial\phi}a^2\delta\phi\right) \\
\delta T_0^\mu &= -\frac{1}{2}\phi_0'\partial_\mu(\delta\phi) \\
\delta T_\nu^\mu &= \delta_\mu^\nu\left(-\phi_0'\Phi + \phi_0'(\delta\phi)' - \frac{\partial V}{\partial\phi}a^2\delta\phi\right)
\end{aligned} \tag{16.27}$$

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

$$(\delta\phi)'' + 2\frac{a'}{a}(\delta\phi)' - \Delta(\delta\phi) = 0 \tag{16.28}$$

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].

$$\mathcal{F}[f(\eta,x,y,z)] = \int f(\eta,K_x,K_y,K_z)e^{-i\vec{K}\cdot\vec{r}}d^3x \tag{16.29}$$

This yields the damped harmonical oscillator solution:

$$(\delta\phi)'' + 2\frac{a'}{a}(\delta\phi)' + K^2\delta\phi = 0 \tag{16.30}$$

Remark 16.6 Note that this solution is in Fourier space, not real space!

[2]This makes sense when you consider the definition of the wavevector.

16.6 Density perturbations

In physical cosmology, cosmological perturbations manifest in the form of very large *density perturbations* and collapsed structures such as stars and galaxies.

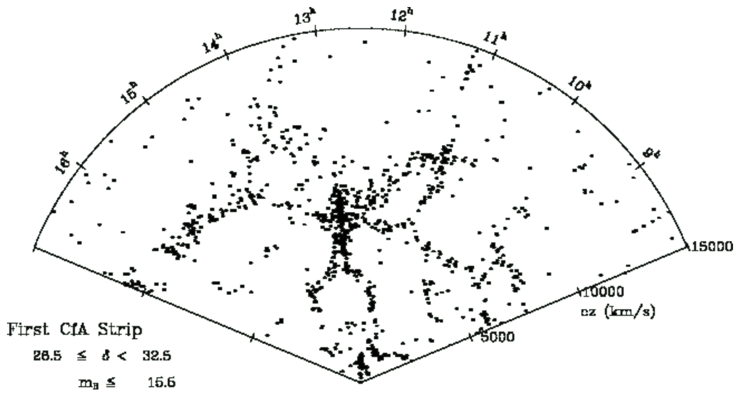


Figure 16.2: Infamous 'Stick Man' image from the CfA Redshift Survey.

Definition 16.4 (Overdensity and underdensity)

$$\delta(\vec{X}) = \frac{\rho(\vec{X}) - \bar{\rho}}{\bar{\rho}} \tag{16.31}$$

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 16.7 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*:

Definition 16.5 (Speed of sound)

$$c_s = \sqrt{\frac{\partial p}{\partial \rho}} \tag{16.32}$$

With (over)density perturbations taken into consideration, we introduce a modified Friedmann equation or the so-called *growth equation*:

Theorem 16.5 (Growth equation)

$$\ddot{\delta} + 2\frac{\dot{a}}{a}\dot{\delta} = \delta\left(4\pi G\bar{\rho} - \frac{\vec{K}^2 c_s^2}{a^2}\right) \tag{16.33}$$

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 16.6 (Generalised density purturbation)

$$\delta(t, \vec{X}) \propto \int e^{-i\vec{K}\cdot\vec{X}} \delta(t, \vec{K}) d^3\vec{K} \tag{16.34}$$

[3] Not the equation of state!

Exercise 16.1 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 16.2 (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

$$\delta_i = \frac{\rho_i - \bar{\rho}_i}{\bar{\rho}_i} \tag{16.35}$$

Considering the cold dark matter case:

$$\delta_m'' + 2\frac{\dot{a}}{a}\delta_m' - 4\pi G(\bar{\rho}_m \delta_m + \bar{\rho}_r \delta_r) = 0 \tag{16.36}$$

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 16.6 (Mészáros effect)

$$\delta'' + \frac{2+3y}{2y(1+y)}\delta' - \frac{3}{2y(1+y)}\delta = 0 \tag{16.37}$$

where $y = \bar{\rho}_m/\bar{\rho}_r$.

Remark 16.8 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 16.3 (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

$$H^2 = \frac{8\pi G}{3}\bar{\rho} \quad H^2 = \frac{8\pi G}{3}\rho - \frac{R}{a^2} \tag{16.38}$$

where R is the curvature. Combining the equations yield

$$\delta = \frac{3Rc^2}{8\pi G}\frac{1}{a^2\rho} \tag{16.39}$$

The solutions, which describe the growth of big density perturbations, then reads

Theorem 16.7 (Evolution of big perturbations)

$$\delta = \begin{cases} a^2 & \text{matter-dominated universe} \\ a & \text{radiation-dominated universe} \end{cases} \tag{16.40}$$

Likewise, for perturbations smaller than the particle horizon, we can derive

Theorem 16.8 (Evolution of *smol* perturbations)

$$\delta = \begin{cases} a^0 & \text{matter-dominated universe} \\ a & \text{radiation-dominated universe} \end{cases} \tag{16.41}$$

Exercise 16.2 Prove the *smol* perturbation evolutions above.

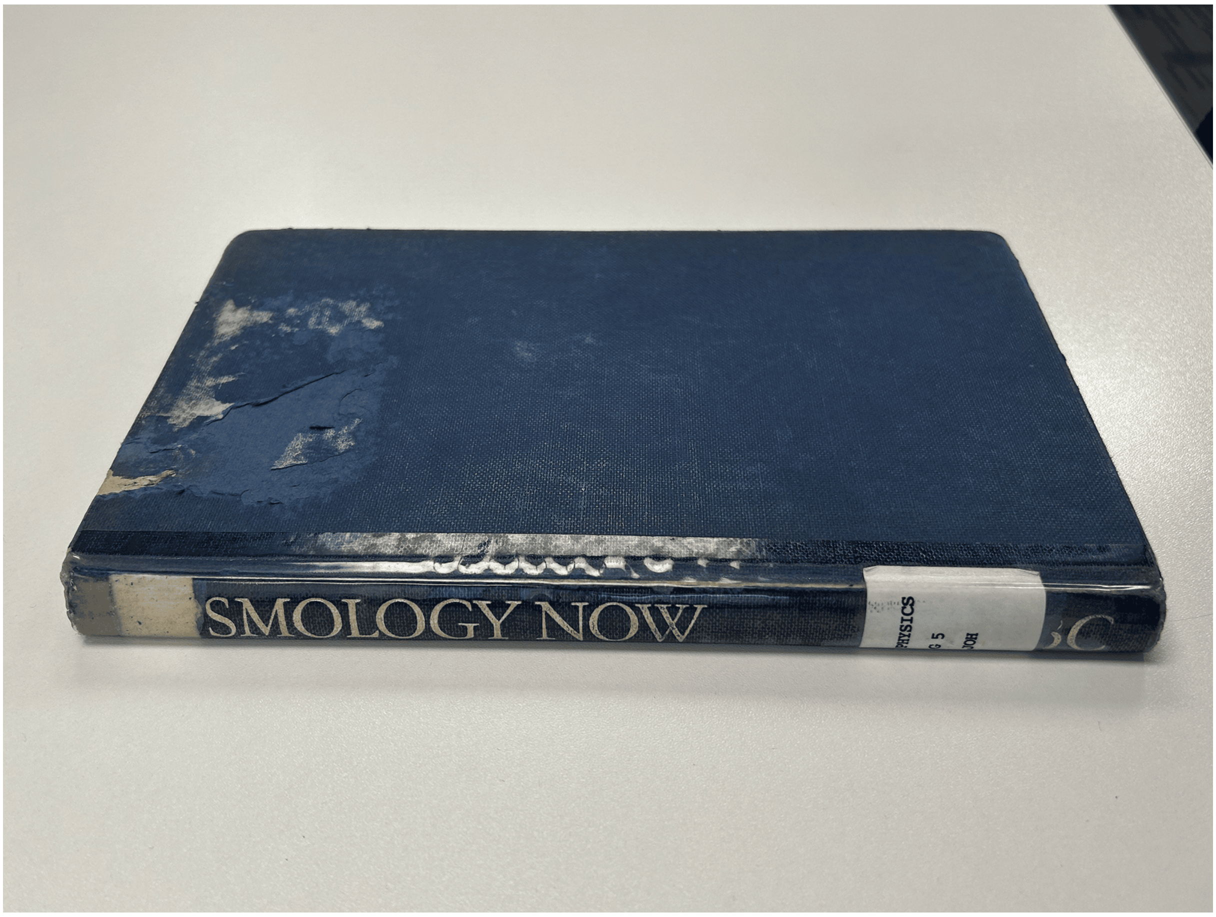


Figure 16.3: ‘Smology Now’

Chapter 17

Instead of a postscript [EMPTY]