

General Relativity and Cosmology

Leonardo Cerasi

General Relativity and Cosmology

© 2025 Leonardo Cerasi. No rights reserved.

This document has been typeset by L^AT_EX with the `book` class.
Source code available on GitHub at [LeonardoCerasi/notes](https://github.com/LeonardoCerasi/notes).

Author's email: leonardo@cerasi.net

Notation

Conventions

In these notes, the Lorentz–Minkowski metric $\eta_{\mu\nu}$ has signature $(-, +, +, +)$, except for the Cosmology part where a $(+, -, -, -)$ signature is adopted, and Greek indices generally run over spacetime coordinates, while Latin indices are general \mathbb{N}_0 -indices defined in each context. Repeated indices are generally summed over, unless otherwise specified. The n -dimensional Levi–Civita symbol $\epsilon^{i_1 \dots i_n}$ is defined with the convention $\epsilon^{01\dots n} = +1$.

Given $\alpha \in \mathbb{C}^{n \times n}$, with $n \in \mathbb{N}_0$, its complex conjugate is denoted as α^* , its transpose as α^\top and its Hermitian conjugate as $\alpha^\dagger := (\alpha^*)^\top$. Given a Dirac spinor $\Psi \in \mathbb{C}^4$, its Dirac dual (or Dirac adjoint) is defined as $\bar{\Psi} := \Psi^\dagger \gamma^0$.

The Landau symbol for a function $f : D \subseteq \mathbb{R} \rightarrow \mathbb{C}$ is defined by the condition $\exists M \in \mathbb{R} : |o(f(x))| \leq M |f(x)| \ \forall x \in D$.

Mathematical notation

The empty set is denoted by \emptyset and the power set of a set A by $\mathcal{P}(A) := \{B : B \subseteq A\}$. The counting numbers are $\mathbb{N} \equiv \{1, 2, 3, \dots\}$, and the natural numbers are defined by $\mathbb{N}_0 \equiv \{0\} \cup \mathbb{N}$. The imaginary unit is denoted by i and the unit quaternions by i, j, k , so that $\mathbb{C}(\mathbb{R}) = \text{span}(1, i)$ and $\mathbb{H}(\mathbb{R}) = \text{span}(1, i, j, k)$.

The n -dimensional sphere is denoted by \mathbb{S}^n , while the n -dimensional disk by \mathbb{D}^n , so that $\partial\mathbb{D}^n = \mathbb{S}^{n-1}$, where in general $\partial\Sigma$ denotes the boundary of Σ . The measure on \mathbb{S}^n is denoted by $d\Omega_n$, while its line element as $d\Omega_n^2$.

The permutation group of n objects, i.e. the n^{th} symmetric group, is denoted by S_n .

Given two \mathbb{K} -vector spaces V and W , with \mathbb{K} a generic field, the space of all \mathbb{K} -linear applications $f : V \rightarrow W$ is denoted by $\text{Hom}_{\mathbb{K}}(V, W)$: in particular, $\text{Hom}_{\mathbb{K}}(V) \equiv \text{End}(V)$. The subset of $\text{End}(V)$ of all automorphisms of V is the automorphism group $\text{Aut}(V)$, which is a group under composition of morphisms.

Given a manifold \mathcal{M} , the space of all smooth scalar functions on \mathcal{M} is denoted by $\mathcal{C}^\infty(\mathcal{M})$, the space of all vector fields on \mathcal{M} by $\mathfrak{X}(\mathcal{M})$, the space of all p -forms on \mathcal{M} by $\Lambda^p(\mathcal{M})$ and the Grassmann algebra of \mathcal{M} by $\Lambda(\mathcal{M}) := \bigoplus_{k=0}^n \Lambda^k(\mathcal{M})$.

The exterior derivative is denoted by d , the partial derivative by ∂ , the nabla operator by $\nabla \equiv (\partial_1, \partial_2, \partial_3)$, the Laplacian operator by $\Delta \equiv \nabla^2$ and the D'Alembert operator by $\square \equiv \partial_0^2 - \Delta$.

A list of “important” Lie groups:

$\text{GL}(n, \mathbb{K}) := \{A \in \mathbb{K}^{n \times n} : \det A \neq 0\}$ general linear group (Lie group for $\mathbb{K} = \mathbb{R}, \mathbb{C}$)

$\text{SL}(n, \mathbb{K}) := \{A \in \text{GL}(n, \mathbb{K}) : \det A = 1\}$ special linear group (Lie group for $\mathbb{K} = \mathbb{R}, \mathbb{C}$)

$\text{O}(n) := \{A \in \mathbb{R}^{n \times n} : AA^\top = A^\top A = I_n\}$ orthogonal group

$\text{SO}(n) := \{A \in \text{O}(n) : \det A = 1\}$ special orthogonal group

$\text{U}(n) := \{A \in \mathbb{C}^{n \times n} : AA^\dagger = A^\dagger A = I_n\}$ unitary group

$\text{SU}(n) := \{A \in \text{U}(n) : \det A = 1\}$ special unitary group

Given a Lie group G , its associated Lie algebra is denoted by \mathfrak{g} .

Physical notation

Natural units are used, with $c = \hbar = k_B \equiv 1$, so that mass, energy, temperature, inverse length and inverse time all have the same units.

In the Cosmology part, the metric signature is inverted to $(+, -, -, -)$, and quantities with a subscript ‘0’ are evaluated at the present coordinate time t_0 .

Contents

I Differential Geometry	1
1 Manifolds	3
1.1 Differentiable manifolds	3
1.1.1 Definitions	3
1.1.2 Maps between manifolds	4
1.2 Tangent spaces	5
1.2.1 Tangent vectors	5
1.2.1.1 Changing coordinates	6
1.2.1.2 Curves	6
1.2.2 Vector fields	6
1.2.2.1 Lie brackets	7
1.2.2.2 Integral curves	7
1.2.3 Lie derivative	8
1.3 Tensors	10
1.3.1 Dual Spaces	10
1.3.2 Cotangent vectors	10
1.3.3 Tensor fields	11
1.3.4 Operations on tensors	12
1.4 Differential forms	14
1.4.1 de Rham cohomology	16
1.4.2 Integration	17
1.4.2.1 Stokes' theorem	18
2 Riemannian Geometry	20
2.1 Metric manifolds	20
2.1.1 Riemannian manifolds	20
2.1.2 Lorentzian manifolds	21
2.1.3 Metric properties	22
2.1.3.1 Hodge theory	23
2.2 Connections	27
2.2.1 Covariant derivative	27
2.2.2 Covariant derivative of tensors	28
2.2.2.1 Torsion and curvature	29
2.2.2.2 Levi–Civita connection	30
2.2.2.3 Gauss' theorem	31
2.2.2.4 Maxwell action	32
2.3 Parallel transport	36
2.3.1 Normal coordinates	37
2.3.1.1 Exponential map	37

	2.3.1.2 Equivalence principle	38
2.3.2	Curvature and torsion	39
	2.3.2.1 Torsion	40
2.3.3	Geodesic deviation	41
2.4	Riemann tensor	43
2.4.1	Ricci and Einstein tensors	43
2.4.2	Connection and curvature forms	44
	2.4.2.1 Vielbeins	44
	2.4.2.2 Connection 1-form	45
	2.4.2.3 Curvature 2-form	46
II	General Relativity	48
3	External Gravitational Fields	50
3.1	Geodetic motion	50
	3.1.1 Non-relativistic particles	50
	3.1.2 Relativistic particles	51
	3.1.3 Interactions	52
	3.1.3.1 Electromagnetism	52
	3.1.3.2 Gravity	53
3.2	Equivalence principle	53
	3.2.1 Kottler–Möller metric	54
	3.2.2 Einstein’s equivalence principle	54
	3.2.3 Gravitational time dilation	55
4	Geometrodynamics	57
4.1	Einstein–Hilbert action	57
	4.1.1 Equations of motion	57
	4.1.1.1 Dimensional analysis	59
	4.1.1.2 Cosmological constant	59
	4.1.2 Diffeomorphisms	59
4.2	Simple solutions	61
	4.2.1 Minkowski space	61
	4.2.2 de Sitter space	61
	4.2.2.1 Riemann tensor	61
	4.2.2.2 Ricci tensor	62
	4.2.2.3 de Sitter geodesics	62
	4.2.2.4 de Sitter embeddings	63
	4.2.3 Anti-de Sitter space	65
	4.2.3.1 Anti-de Sitter geodesics	65
	4.2.3.2 Anti-de Sitter embeddings	66
4.3	Symmetries	68
	4.3.1 Isometries	68
	4.3.1.1 Minkowski spacetime	68
	4.3.1.2 de Sitter and anti-de Sitter space	69
	4.3.2 Conserved quantities	70
	4.3.3 Komar integrals	70
4.4	Asymptotics of spacetime	72

4.4.1	Conformal transformations	72
4.4.2	Penrose diagrams	73
 4.4.2.1	Minkowski spacetime	73
 4.4.2.2	de Sitter space	75
 4.4.2.3	Anti-de Sitter space	78
4.5	Matter coupling	80
 4.5.1	Field theories in curved spacetime	80
 4.5.2	Einstein equations with matter	80
 4.5.3	Energy-momentum tensor	81
 4.5.3.1	Field theories	82
 4.5.3.2	Perfect fluids	83
 4.5.4	Energy conservation	83
 4.5.4.1	Conserved energy from Killing vectors	84
 4.5.5	Energy conditions	84
 4.5.5.1	Weak energy condition	85
 4.5.5.2	Strong energy condition	85
 4.5.5.3	Null energy condition	86
 4.5.5.4	Dominant energy condition	86
4.6	Schwarzschild solution	87
 4.6.1	Birkhoff theorem	87
 4.6.2	Motion in a Schwarzschild field	90
 4.6.2.1	Freely-falling emitter	91
 4.6.3	Classical tests of General Relativity	92
 4.6.3.1	Gravitational redshift	92
 4.6.3.2	Periastron precession	92
 4.6.3.3	Deflection of light	94
 4.6.3.4	Gravitational time delay	95
 4.6.3.5	Gravitational microlensing	96
5	Weak Gravity	99
5.1	Linerarized gravity	99
 5.1.1	Gauge symmetry	100
 5.1.2	Newtonian limit	101
5.2	Gravitational waves	102
 5.2.1	Polarizations	103
 5.2.1.1	+ polarization	103
 5.2.1.2	× polarization	104
 5.2.1.3	Gravitational wave detection	104
 5.2.2	Exact solutions	104
5.3	Perturbing spacetime	106
 5.3.1	Green's function	106
 5.3.1.1	Multipole expansion	108
 5.3.2	Radiated power	109
 5.3.2.1	Fierz–Pauli action	109
 5.3.2.2	Conceptual issues	110
 5.3.2.3	Gauge invariance	111
 5.3.2.4	Orders of magnitude	111

6 Black Holes	113
6.1 Schwarzschild black holes	113
6.1.1 Horizon	113
6.1.1.1 Near horizon limit	114
6.1.2 Eddington–Finkelstein coordinates	115
6.1.2.1 Ingoing coordinates	115
6.1.2.2 Outgoing coordinates	117
6.1.3 Kruskal spacetime	117
6.1.3.1 Kruskal diagram	118
6.1.3.2 Penrose diagram	121
6.1.4 Weak cosmic censorship	122
6.1.4.1 Naked singularities	122
6.1.5 de Sitter black holes	123
III Cosmology	125
7 Cosmological Geometrodynamics	127
7.1 Geometry	127
7.1.1 FLRW metric	128
7.2 Kinematics	130
7.2.1 Particle motion	130
7.2.2 Redshift	131
7.2.3 Distances	131
7.3 Dynamics	134
7.3.1 Matter sources	134
7.3.1.1 Number density	134
7.3.1.2 Energy-momentum tensor	134
7.3.1.3 Matter content	135
7.3.2 Curvature tensors	136
7.3.3 Friedmann equations	137
7.3.3.1 Λ CDM	138
7.3.3.2 Single-component universe	138
8 Inflation	140
8.1 Horizon problem	140
8.1.1 Particle horizon	140
8.1.2 CMB uniformity	141
8.1.3 SEC violation and Inflation	142
8.1.3.1 Estimates	143
8.2 Inflationary model	144
8.2.1 Scalar field dynamics	144
8.2.1.1 Slow-roll inflation	145
8.2.2 Reheating	145
9 Thermal History	147
9.1 Hot Big Bang	147
9.2 Equilibrium	150
9.2.1 Equilibrium thermodynamics	150

9.2.2	Density and pressure	151
9.2.3	Effective number of relativistic species	153
9.2.3.1	Entropy	154
9.2.3.2	Neutrino decoupling	156
9.2.3.3	Electron-positron annihilation	157
9.2.4	Thermal history	158
9.3	Beyond equilibrium	160
9.3.1	Boltzmann equation	160
9.3.2	Dark matter relics	161
9.3.2.1	Dark matter freeze-out	161
9.3.2.2	WIMP miracle	162
9.3.3	Big Bang nucleosynthesis	163
9.3.3.1	Equilibrium abundances	163
9.3.3.2	Neutron freeze-out and decay	164
9.3.3.3	Helium fusion	164
9.3.3.4	Constraints on BSM physics	165
9.3.4	Recombination	167
9.3.4.1	Saha equilibrium	168
9.3.4.2	Hydrogen recombination	168
9.3.4.3	Photon decoupling	169
10	Structure Formation	170
10.1	Newtonian perturbation theory	170
10.1.1	Flat background	170
10.1.1.1	Matter perturbations	171
10.1.1.2	Gravitational perturbations	171
10.1.2	FLRW background	172
10.1.2.1	Matter era	174
10.1.2.2	Radiation era	174
10.1.2.3	Vacuum era	175
10.2	Relativistic perturbation theory	177
10.2.1	Perturbations	177
10.2.1.1	Metric perturbations	177
10.2.1.2	Matter perturbations	178
10.2.2	Evolution	179
10.2.2.1	Metric evolution	180
10.2.2.2	Radiation evolution	181
10.2.2.3	CMB power spectrum	184
Bibliography		185

Part I

Differential Geometry

Chapter 1

Manifolds

§1.1 Differentiable manifolds

Definition 1.1.1 (Topological space)

The **topology** \mathcal{T} of a set X is a family of subsets of X , i.e. $\mathcal{T} \subseteq \mathcal{P}(X)$, defined as **open sets**, with the following properties:

1. $\emptyset, X \in \mathcal{T}$;
2. $O_\alpha, O_\beta \in \mathcal{T} \implies O_\alpha \cap O_\beta \in \mathcal{T}$;
3. $\{O_\alpha\}_{\alpha \in \mathcal{I}} \subset \mathcal{T} \implies \bigcup_{\alpha \in \mathcal{I}} O_\alpha \in \mathcal{T}$.

A **topological space** M is a set of points, endowed with a topology \mathcal{T} .

Given a topological space (M, \mathcal{T}) , $O \in \mathcal{T}$ is a **neighbourhood** of a point $p \in M$ if $p \in O$: then, (M, \mathcal{T}) is **Hausdorff** if $\forall p, q \in M \exists O_1, O_2 \in \mathcal{T}$ neighbourhoods of p and q respectively such that $O_1 \cap O_2 = \emptyset$.

Topological spaces allow to introduce the concept of continuity: given two topological spaces (M_1, \mathcal{T}_1) and (M_2, \mathcal{T}_2) , a map $f : M_1 \rightarrow M_2$ is **continuous** if $O \in \mathcal{T}_2 \Rightarrow f^{-1}(O) \in \mathcal{T}_1$.

Definition 1.1.2 (Homeomorphism)

Given two topological spaces (M_1, \mathcal{T}_1) and (M_2, \mathcal{T}_2) , a map $f : M_1 \rightarrow M_2$ is a **homeomorphism** if it is bijective and bicontinuous, i.e. both f and f^{-1} are continuous.

§1.1.1 Definitions

Definition 1.1.3 (Differentiable manifold)

An n -dimensional **differentiable manifold** \mathcal{M} is a Hausdorff topological space such that:

1. \mathcal{M} is locally homeomorphic to \mathbb{R}^n , i.e. $\forall p \in \mathcal{M} \exists O \in \mathcal{T}(\mathcal{M}) : p \in O \wedge \exists \varphi : O \rightarrow U \in \mathcal{T}(\mathbb{R}^n)$ homeomorphism;
2. given $O_\alpha, O_\beta \in \mathcal{T}(\mathcal{M}) : O_\alpha \cap O_\beta \neq \emptyset$, the corresponding maps $\varphi_\alpha : O_\alpha \rightarrow U_\alpha$ and $\varphi_\beta : O_\beta \rightarrow U_\beta$ must be *compatible*, i.e. $\varphi_\beta \circ \varphi_\alpha^{-1} : \varphi_\alpha(O_\alpha \cap O_\beta) \rightarrow \varphi_\beta(O_\alpha \cap O_\beta)$ and its inverse must be smooth (of \mathcal{C}^∞ class).

The maps φ_α are called **charts** and a collection of compatible charts is called an **atlas**: a maximal atlas \mathcal{A} is an atlas such that $\bigcup_{\alpha \in \mathcal{I}} O_\alpha = \mathcal{M}$. Two atlases are compatible if each chart of one atlas is compatible with every chart of the other: they define the same differentiable structure on the manifold.

Each chart φ_α provides a coordinate system on O_α : $\varphi_\alpha(p) = (x^1(p), \dots, x^\mu(p), \dots, x^n(p))$. The transition functions $\varphi_\beta \circ \varphi_\alpha^{-1}$ are therefore coordinate transformations on overlapping regions.

Example 1.1.1 (Spheres)

S^n is a differentiable manifold for $n \in \mathbb{N}$. In particular, to define a differentiable structure on S^1 , an atlas of two charts is needed: the standard parametrization $\vartheta \in [0, 2\pi)$ is not a well-defined chart because $[0, 2\pi)$ is not an open set in the Euclidean topology of \mathbb{R} , therefore the elimination of a point is necessary; usually, the two charts of the atlas are defined by $\vartheta_1 \in (0, 2\pi)$, excluding $(1, 0)$ (in the embedding space \mathbb{R}^2), and $\vartheta_2 \in (-\pi, \pi)$, excluding $(-1, 0)$: they are evidently compatible, thus they form a maximal atlas.

In the remainder of these notes, \mathcal{M} is always taken to be an n -dimensional differentiable manifold.

§1.1.2 Maps between manifolds

Locally mapping \mathcal{M} to \mathbb{R}^n allows for the extension of the concepts of Analysis from \mathbb{R}^n to \mathcal{M} .

Definition 1.1.4 (Smooth maps)

function $f : \mathcal{M} \rightarrow \mathbb{R}$ on a differentiable manifold $(\mathcal{M}, \mathcal{A})$ is **smooth** if $f \circ \varphi_\alpha^{-1} : U_\alpha \rightarrow \mathbb{R}$ is smooth for all charts $(U_\alpha, \varphi_\alpha) \in \mathcal{A}$.

A map $f : \mathcal{M} \rightarrow \mathcal{N}$ between two differentiable manifolds $(\mathcal{M}, \mathcal{A}_1), (\mathcal{N}, \mathcal{A}_2)$ is **smooth** if $\psi_{\alpha_2} \circ f \circ \varphi_{\alpha_1}^{-1} : U_{\alpha_1} \rightarrow V_{\alpha_2}$ is smooth for all charts $(U_{\alpha_1}, \varphi_{\alpha_1}) \in \mathcal{A}_1, (V_{\alpha_2}, \varphi_{\alpha_2}) \in \mathcal{A}_2$.

A smooth homeomorphism $f : \mathcal{M} \rightarrow \mathcal{N}$ between two differentiable manifolds \mathcal{M} and \mathcal{N} is called a **diffeomorphism**.

Proposition 1.1.1 (Diffeomorphic manifolds)

If \mathcal{M} and \mathcal{N} are **diffeomorphic**, then $\dim_{\mathbb{R}} \mathcal{M} = \dim_{\mathbb{R}} \mathcal{N}$.

Example 1.1.2 (Differentiable structures)

S^7 can be covered by multiple incompatible atlases: the resulting manifolds are homeomorphic but not diffeomorphic.

\mathbb{R}^n has a unique differentiable structure for all $n \in \mathbb{N}$, except for $n = 4$: \mathbb{R}^4 can be covered by infinitely-many incompatible atlases.

§1.2 Tangent spaces

The notions of calculus can be defined on a differential manifold $(\mathcal{M}, \mathcal{A})$ with the notion of tangent spaces. Indeed, the derivative of a function $f : \mathcal{M} \rightarrow \mathbb{R}$ at a point $p \in \mathcal{M}$, covered by the chart (φ, U) , is defined as:

$$\frac{\partial f}{\partial x^\mu} \Big|_p := \frac{\partial(f \circ \varphi^{-1})}{\partial x^\mu} \Big|_{\varphi(p)} \quad (1.2.1)$$

Evidently, this definition depends on the choice of coordinates x^μ , thus it depends on the chart.

§1.2.1 Tangent vectors

Definition 1.2.1 (Space of smooth functions)

The set of all smooth functions on \mathcal{M} is denoted by $\mathcal{C}^\infty(\mathcal{M})$.

Definition 1.2.2 (Tangent vector)

A **tangent vector** to \mathcal{M} in $p \in \mathcal{M}$ is an operator $X_p : \mathcal{C}^\infty(\mathcal{M}) \rightarrow \mathbb{R}$ such that:

1. $X_p(f + g) = X_p(f) + X_p(g) \quad \forall f, g \in \mathcal{C}^\infty(\mathcal{M})$;
2. $X_p(f) = 0$ for all constant functions $f \in \mathcal{C}^\infty(\mathcal{M})$;
3. $X_p(fg) = X_p(f)g(p) + f(p)X_p(g) \quad \forall f, g \in \mathcal{C}^\infty(\mathcal{M})$.

Conditions 2. and 3. trivially imply that $X_p(\alpha f) = \alpha X_p(f) \quad \forall \alpha \in \mathbb{R}$, which means that X_p is a linear operator, i.e. $X_p \in \text{Hom}_{\mathbb{R}}(\mathcal{C}^\infty(\mathcal{M}), \mathbb{R})$. Moreover, it is simple to check that $\partial_\mu|_p$ satisfies the conditions of Def. 1.2.2.

Theorem 1.2.1 (Tangent space)

The set $T_p\mathcal{M}$ of all tangent vectors at a point $p \in \mathcal{M}$ forms an n -dimensional space, called **tangent space**, and $\{\partial_\mu|_p\}_{\mu=1,\dots,n}$ is a basis of such space.

Proof. Defining $f \circ \varphi^{-1} \equiv F : U \subset \mathcal{M} \rightarrow \mathbb{R}$, with $f : \mathcal{M} \rightarrow \mathbb{R}$ and $(\varphi, U) \in \mathcal{A}$, it can be shown that, in some neighbourhood of p (not necessarily U), F can always be written as:

$$F(x) = F(x^\mu(p)) + (x^\mu - x^\mu(p)) F_\mu(x)$$

for some functions $\{F_\mu\}_{\mu=1,\dots,n}$ (e.g. $F(x) = F(0) + x \int_0^1 dt F(xt)$). Applying $\partial_\mu|_{x(p)}$:

$$\frac{\partial F}{\partial x^\mu} \Big|_{x(p)} = F_\mu(x(p))$$

Defining $f_\mu \equiv F_\mu \circ \varphi$, for any $q \in \mathcal{M}$ in an appropriate neighbourhood of p :

$$f(q) = f(p) + (x^\mu(q) - x^\mu(p)) f_\mu(q)$$

Moreover, remembering Eq. Eq. 1.2.1:

$$f_\mu(p) = F_\mu \circ \varphi(p) = F_\mu(x(p)) = \frac{\partial F}{\partial x^\mu} \Big|_{x(p)} = \frac{\partial f}{\partial x^\mu} \Big|_p$$

Using these facts, the action of a tangent vector can be written explicitly:

$$\begin{aligned} X_p(f) &= X_p(f(p) + (x^\mu - x^\mu(p)) f_\mu) \\ &= X_p(f(p)) + X_p((x^\mu - x^\mu(p))) f_\mu(p) + (x^\mu - x^\mu(p))(p) X_p(f_\mu) \\ &= X_p(x^\mu) f_\mu(p) \end{aligned}$$

because $f(p)$ is a constant and $(x^\mu - x^\mu(p))(p) = x^\mu(p) - x^\mu(p) = 0$. Therefore, remembering the expression for $f_\mu(p)$:

$$X_p = X_p(x^\mu) \frac{\partial}{\partial x^\mu} \Big|_p \equiv X^\mu \frac{\partial}{\partial x^\mu} \Big|_p$$

Thus, $T_p \mathcal{M} = \langle \{\partial_\mu|_p\}_{\mu=1,\dots,n} \rangle$. To check for linear independence, suppose $\alpha = \alpha^\mu \partial_\mu|_p \equiv 0$: acting on $f = x^\nu$, it gives $\alpha(f) = \alpha_\mu \partial_\mu(x^\nu)|_p = \alpha_\nu = 0$. This concludes the proof. \square

§1.2.1.1 Changing coordinates

Although $\partial_\mu|_p$ depends on the choice of coordinates (it is a **coordinate basis**), the existence of X_p is independent of that choice. If two different charts (φ, U) and $(\tilde{\varphi}, V)$ intersect in a neighbourhood of $p \in U \cap V$, the transition from x^μ to y^μ can be expressed as:

$$X_p(f) = X^\mu \frac{\partial f}{\partial x^\mu} \Big|_p = X^\mu \frac{\partial y^\nu}{\partial x^\mu} \Big|_{\varphi(p)} \frac{\partial f}{\partial y^\nu} \Big|_p \quad (1.2.2)$$

This equation can have two interpretations, namely the alibi and the alias interpretation:

$$\frac{\partial}{\partial x^\mu} \Big|_p = \frac{\partial y^\nu}{\partial x^\mu} \Big|_{\varphi(p)} \frac{\partial}{\partial y^\nu} \Big|_p \quad \tilde{X}^\nu = X^\mu \frac{\partial y^\nu}{\partial x^\mu} \Big|_{\varphi(p)} \quad (1.2.3)$$

Components of vectors which transform this way are called **contravariant**.

§1.2.1.2 Curves

Consider a smooth curve on \mathcal{M} , i.e. a smooth map $\sigma : I \in \mathcal{T}(\mathbb{R}) \rightarrow \mathcal{M}$, WLOG parametrized as $\sigma(t) : \sigma(0) = p \in \mathcal{M}$; with a given chart (φ, U) , this curve becomes $\varphi \circ \sigma : I \rightarrow \mathbb{R}^n$, parametrized by $x^\mu(t)$. The tangent vector to the curve in p is:

$$X_p = \frac{dx^\mu(t)}{dt} \Big|_{t=0} \frac{\partial}{\partial x^\mu} \Big|_p \quad (1.2.4)$$

This operator, applied to a function $f \in \mathcal{C}^\infty(\mathcal{M})$, computes the directional derivative of f along the curve. Every tangent vector can be written as in Eq. Eq. 1.2.4, therefore the tangent space can be seen as the space of all possible tangent vectors to curves passing through p .

It must be noted that tangent spaces at different points are entirely different spaces: there is no way to directly compare vectors between them.

§1.2.2 Vector fields

Definition 1.2.3 (Vector fields)

A **vector field** X is a smooth map $X : p \in \mathcal{M} \mapsto X_p \in T_p \mathcal{M}$. The space of all vector fields

on \mathcal{M} is denoted by $\mathfrak{X}(\mathcal{M})$.

Note that a vector field can also be viewed as a smooth map $X : \mathcal{C}^\infty(\mathcal{M}) \rightarrow \mathcal{C}^\infty(\mathcal{M})$, since $(X(f))(p) = X_p(f) \in \mathbb{R}$. Given a chart (φ, U) , a vector field X can be expressed as:

$$X = X^\mu \frac{\partial}{\partial x^\mu} \quad (1.2.5)$$

with $X^\mu \in \mathcal{C}^\infty(\mathcal{M})$. This expression is only defined on U .

§1.2.2.1 Lie brackets

Given two vector fields $X, Y \in \mathfrak{X}(\mathcal{M})$, their product is clearly not a vector field, as it does not satisfy Leibniz' rule:

$$XY(fg) = XY(f)g + Y(f)X(g) + X(f)Y(g) + fXY(g) \neq XY(f)g + fXY(g)$$

where $XY(f) \equiv X(Y(f))$.

Definition 1.2.4 (Lie brackets)

Given two vector fields $X, Y \in \mathfrak{X}(\mathcal{M})$, their commutator (or **Lie bracket**) is defined as:

$$[X, Y](f) = XY(f) - YX(f) \quad (1.2.6)$$

With a given chart:

$$\begin{aligned} [X, Y](f) &= X^\mu \frac{\partial}{\partial x^\mu} \left(Y^\nu \frac{\partial f}{\partial x^\nu} \right) - Y^\mu \frac{\partial}{\partial x^\mu} \left(X^\nu \frac{\partial f}{\partial x^\nu} \right) \\ &= \left(X^\mu \frac{\partial Y^\nu}{\partial x^\mu} - Y^\mu \frac{\partial X^\nu}{\partial x^\mu} \right) \frac{\partial f}{\partial x^\nu} \end{aligned}$$

therefore:

$$[X, Y] = \left(X^\mu \frac{\partial Y^\nu}{\partial x^\mu} - Y^\mu \frac{\partial X^\nu}{\partial x^\mu} \right) \frac{\partial}{\partial x^\nu} \quad (1.2.7)$$

Theorem 1.2.2 (Jacobi identity)

Given $X, Y, Z \in \mathfrak{X}(\mathcal{M})$, then:

$$[X, [Y, Z]] + [Y, [Z, X]] + [Z, [X, Y]] = 0 \quad (1.2.8)$$

With Lie brackets, $\mathfrak{X}(\mathcal{M})$ can be given the structure of a Lie algebra.

§1.2.2.2 Integral curves

Definition 1.2.5 (Flow)

A **flow** on \mathcal{M} is a one-parameter family of diffeomorphisms $\sigma_t : \mathcal{M} \rightarrow \mathcal{M}$, labelled by $t \in \mathbb{R}$, with group structure: $\sigma_0 = \mathbb{1}_{\mathcal{M}}$ and $\sigma_s \circ \sigma_t = \sigma_{s+t}$, thus $\sigma_{-t} = \sigma_t^{-1}$.

Such flows give rise to streamlines on the manifold: these streamlines are required to be smooth. Defining $x^\mu(\sigma_t) \equiv x^\mu(t)$, a vector field can be defined by the tangent to the streamlines at each point on the manifold:

$$X^\mu(x^\mu(t)) = \frac{dx^\mu(t)}{dt} \quad (1.2.9)$$

The inverse reasoning is also possible: given $X \in \mathfrak{X}(\mathcal{M})$, the streamlines defined by Eq. 1.2.9 are the **integral curves** of X .

Proposition 1.2.1 (Infinitesimal flow)

The **infinitesimal flow** generated by $X \in \mathfrak{X}(\mathcal{M})$ is:

$$x^\mu(t) = x^\mu(0) + tX^\mu(x(0)) + o(t^2) \quad (1.2.10)$$

A vector field which generates a flow defined for all $t \in \mathbb{R}$ is called **complete**.

Theorem 1.2.3

If \mathcal{M} is compact, then all $X \in \mathfrak{X}(\mathcal{M})$ are complete.

Example 1.2.1 (Integral curves on the 2-sphere)

On \mathbb{S}^2 , the flow generated by $X = \partial_\phi$ is described by $\dot{\phi} = 1, \dot{\theta} = 0$, thus $\theta(t) = \theta_0$ and $\phi(t) = \phi_0 + t$: the flow lines are lines of constant latitude.

§1.2.3 Lie derivative

Defining calculus for vector fields requires a way to compare vectors of different tangent spaces.

Definition 1.2.6 (Pull-back of functions)

Given a diffeomorphism between two manifolds $\varphi : \mathcal{M} \rightarrow \mathcal{N}$ and a function $f : \mathcal{N} \rightarrow \mathbb{R}$, the **pull-back** of f is the function $\varphi^*f : \mathcal{M} \rightarrow \mathbb{R}$ such that $\varphi^*f(p) = f(\varphi(p))$.

Definition 1.2.7 (Push-forward of vector fields)

Given a diffeomorphism between two manifolds $\varphi : \mathcal{M} \rightarrow \mathcal{N}$ and a vector field $X \in \mathfrak{X}(\mathcal{M})$, the **push-forward** of X is the vector field $\varphi_*X \in \mathfrak{X}(\mathcal{N})$ such that $\varphi_*X(f) = X(\varphi^*f)$.

This last equality must be evaluated at the appropriate points: $[\varphi_*X(f)](\varphi(p)) = [X(\varphi^*f)](p)$. With the appropriate charts on \mathcal{M} and \mathcal{N} , the definitions above can be rewritten with coordinates:

$$\varphi^*f(x) = f(y(x)) \quad (1.2.11)$$

$$\varphi_*X(f) = X^\mu \frac{\partial f(y(x))}{\partial x^\mu} = X^\mu \frac{\partial y^\alpha}{\partial x^\mu} \frac{\partial f(y)}{\partial y^\alpha} \quad (1.2.12)$$

The notions of pull-back and push-forward allow to compare tangent vectors at neighbouring points and, in particular, to define the derivative along a vector field.

Definition 1.2.8 (Lie derivative of functions)

Given a function $f : \mathcal{M} \rightarrow \mathbb{R}$ and a vector field $X \in \mathfrak{X}(\mathcal{M})$, the derivative of f along X ,

called **Lie derivative**, is defined as:

$$\mathcal{L}_X f(x) := \lim_{t \rightarrow 0} \frac{f(\sigma_t(x)) - f(x)}{t} = \frac{df(\sigma_t(x))}{dt} \Big|_{t=0} \quad (1.2.13)$$

where σ_t is the flow generated by X .

Lemma 1.2.1

$$\mathcal{L}_X f = X(f) \quad (1.2.14)$$

$$\text{Proof. } \mathcal{L}_X f = \frac{df(\sigma_t)}{dt} = \frac{\partial f}{\partial x^\mu} \frac{dx^\mu(t)}{dt} = X^\mu \frac{\partial f}{\partial x^\mu} = X(f).$$

□

The Lie derivative can be extended to vector fields.

Definition 1.2.9 (Lie derivative of vector fields)

Given two vector fields $X, Y \in \mathfrak{X}(\mathcal{M})$, the **Lie derivative** of Y along X is defined as:

$$\mathcal{L}_X Y_p := \lim_{t \rightarrow 0} \frac{((\sigma_{-t})_* Y)_p - Y_p}{t} \quad (1.2.15)$$

where σ_t is the flow generated by X .

The use of the inverse flow σ_{-t} is necessary because to evaluate the vector field $\mathcal{L}_X Y$ at the point $p \in \mathcal{M}$, the tangent vector $Y_{\sigma_t(p)} \in T_{\sigma_t(p)}\mathcal{M}$ must be “pushed-back” to $T_p\mathcal{M} = T_{\sigma_0(p)}\mathcal{M}$. With $t \rightarrow 0$, the infinitesimal flow σ_{-t} is, according to Eq. 1.2.10, $x^\mu(t) = x^\mu(0) - tX^\mu + o(t)$, therefore the Lie derivative of basis tangent vectors can be expressed as:

$$(\sigma_{-t})_* \partial_\mu = \frac{\partial x^\nu(t)}{\partial x^\mu} \frac{\partial}{\partial x^\nu(t)} = \left(\delta_\mu^\nu - t \frac{\partial X^\nu}{\partial x^\mu} + o(t) \right) \partial_\nu(t) \implies \mathcal{L}_X \partial_\mu = - \frac{\partial X^\nu}{\partial x^\mu} \partial_\nu \quad (1.2.16)$$

Moreover, by the Jacobi identity it follows that:

$$\mathcal{L}_X \mathcal{L}_Y Z - \mathcal{L}_Y \mathcal{L}_X Z = \mathcal{L}_{[X,Y]} Z \quad (1.2.17)$$

Lemma 1.2.2

$$\mathcal{L}_X Y = [X, Y] \quad (1.2.18)$$

$$\text{Proof. } \mathcal{L}_X Y = \mathcal{L}_X(Y^\mu \partial_\mu) = (\mathcal{L}_X Y^\mu) \partial_\mu + Y^\mu (\mathcal{L}_X \partial_\mu) = X^\nu \frac{\partial Y^\mu}{\partial x^\nu} \partial_\mu - Y^\mu \frac{\partial X^\nu}{\partial x^\mu} \partial_\nu = [X, Y]. \quad \square$$

§1.3 Tensors

§1.3.1 Dual Spaces

Definition 1.3.1 (Dual space)

Given a vector space V , its **dual space** V^* is the space of all linear maps $f : V \rightarrow \mathbb{R}$.

Given a basis $\{\mathbf{e}_\mu\}_{\mu=1,\dots,n}$ of V , its **dual basis** $\{\mathbf{f}^\mu\}_{\mu=1,\dots,n}$ of V^* can be defined by:

$$\mathbf{f}^\nu(\mathbf{e}_\mu) = \delta_\mu^\nu \quad (1.3.1)$$

A general vector in V can be written as $X = X^\mu \mathbf{e}_\mu$, thus, according to Eq. 1.3.1, $X^\mu = \mathbf{f}^\mu(X)$. Clearly, the map $f : \mathbf{e}_\mu \mapsto \mathbf{f}^\mu$ is a non-canonical isomorphism between V and V^* , hence $\dim_{\mathbb{R}} V = \dim_{\mathbb{R}} V^*$.

Proposition 1.3.1 (Dual of the dual)

$$(V^*)^* \cong V \quad (1.3.2)$$

Proof. The natural isomorphism between $(V^*)^*$ and V is basis-independent: suppose $X \in V$ and $\omega \in V^*$, so that $\omega(X) \in \mathbb{R}$; X can be viewed as $X \in (V^*)^*$ by setting $X(\omega) \equiv \omega(X)$. \square

§1.3.2 Cotangent vectors

Definition 1.3.2 (Cotangent space)

Given a differentiable manifold $(\mathcal{M}, \mathcal{A})$ and a point $p \in \mathcal{M}$, the **cotangent space** to \mathcal{M} at p is defined as $T_p^*\mathcal{M} := (T_p\mathcal{M})^*$.

Elements of $T_p^*\mathcal{M}$ are called cotangent vectors (or **covectors**).

Definition 1.3.3 (Covector field)

A covector field (or **1-form**) is a smooth map $\omega : p \in \mathcal{M} \mapsto \omega_p \in T_p^*\mathcal{M}$. The space of all 1-forms on \mathcal{M} is denoted by $\Lambda^1(\mathcal{M})$.

Note that a 1-form can also be viewed as a smooth map $\omega : \mathfrak{X}(\mathcal{M}) \rightarrow \mathcal{C}^\infty(\mathcal{M})$, since $(\omega(X))(p) = \omega_p(X_p) \in \mathbb{R}$.

Proposition 1.3.2

$\{dx^\mu\}_{\mu=1,\dots,n}$ is a basis of $\Lambda^1(\mathcal{M})$, dual to the basis $\{\partial_\mu\}_{\mu=1,\dots,n}$ of $\mathfrak{X}(\mathcal{M})$.

Proof. Consider $f \in \mathcal{C}^\infty(\mathcal{M})$ and define $df \in \Lambda^1(\mathcal{M})$ by $df(X) = X(f)$: taking $f = x^\mu$ and $X = \partial_\mu$, $df(X) = \partial_\nu(x^\mu) = \delta_\nu^\mu$, therefore $\{dx^\mu\}_{\mu=1,\dots,n}$ is the dual basis of $\Lambda^1(\mathcal{M})$. \square

This is also confirmed by $\mathrm{d}f = \frac{\partial f}{\partial x^\mu} \mathrm{d}x^\mu$. These are coordinate bases: in fact, given two different charts $(\varphi, U), (\tilde{\varphi}, V)$:

$$\mathrm{d}y^\mu = \frac{\mathrm{d}y^\mu}{\mathrm{d}x^\nu} \mathrm{d}x^\nu \quad (1.3.3)$$

which is the inverse of Eq. 1.2.3 (not evaluated at a specific point). This ensures that:

$$\mathrm{d}y^\mu \left(\frac{\partial}{\partial y^\nu} \right) = \frac{\partial y^\mu}{\partial x^\alpha} \frac{\partial x^\beta}{\partial y^\nu} \mathrm{d}x^\alpha \left(\frac{\partial}{\partial x^\beta} \right) = \frac{\partial y^\mu}{\partial x^\alpha} \frac{\partial x^\alpha}{\partial y^\nu} = \delta_\nu^\mu$$

A 1-form $\omega \in \Lambda^1(\mathcal{M})$ can thus be expressed both as $\omega = \omega_\mu \mathrm{d}x^\mu = \tilde{\omega}_\mu \mathrm{d}y^\mu$, with:

$$\tilde{\omega}_\mu = \frac{\partial x^\nu}{\partial y^\mu} \omega_\nu \quad (1.3.4)$$

Components of 1-forms which transform this way are called **covariant**.

Definition 1.3.4 (Pull-back of 1-forms)

Given a diffeomorphism between two manifolds $\varphi : \mathcal{M} \rightarrow \mathcal{N}$ and a 1-form $\omega \in \Lambda^1(\mathcal{N})$, the **pull-back** of ω is the 1-form $\varphi^*\omega \in \Lambda^1(\mathcal{M})$ such that $\varphi^*\omega(X) = \omega(\varphi_*X)$.

With the appropriate charts on \mathcal{M} and \mathcal{N} , the definition above can be rewritten with coordinates:

$$\varphi^*\omega = \omega_\alpha \frac{\partial y^\alpha}{\partial x^\mu} \mathrm{d}x^\mu \quad (1.3.5)$$

Definition 1.3.5 (Lie derivative of 1-forms)

Given a vector field $X \in \mathfrak{X}(\mathcal{M})$ and a 1-form $\omega \in \Lambda^1(\mathcal{M})$, the **Lie derivative** of ω along X is defined as:

$$\mathcal{L}_X \omega_p := \lim_{t \rightarrow 0} \frac{(\sigma_t^* \omega)_p - \omega_p}{t} \quad (1.3.6)$$

where σ_t is the flow generated by X .

In contrast with the Lie derivative of a vector field, which pushes forward with σ_{-t} (i.e. pushes back), the Lie derivative of a 1-form pulls back with σ_t : this results in the difference of a minus sign with respect to Eq. 1.2.16, giving:

$$\mathcal{L}_X \mathrm{d}x^\mu = \frac{\partial X^\mu}{\partial x^\nu} \mathrm{d}x^\nu \quad (1.3.7)$$

Therefore, on a general 1-form $\omega = \omega_\mu \mathrm{d}x^\mu$:

$$\mathcal{L}_X \omega = (X^\nu \partial_\nu \omega_\mu + \omega_\nu \partial_\mu X^\nu) \mathrm{d}x^\mu \quad (1.3.8)$$

§1.3.3 Tensor fields

Definition 1.3.6 (Tensor)

A **tensor** of rank (r, s) at a $p \in \mathcal{M}$ of a differentiable manifold $(\mathcal{M}, \mathcal{A})$ is a multilinear map defined as:

$$T_p : \overbrace{T_p^* \mathcal{M} \times \cdots \times T_p^* \mathcal{M}}^r \times \overbrace{T_p \mathcal{M} \times \cdots \times T_p \mathcal{M}}^s \rightarrow \mathbb{R} \quad (1.3.9)$$

For example, a cotangent vector $\omega_p \in T_p^*\mathcal{M}$ is a tensor of rank $(1, 0)$, while a tangent vector $X_p \in T_p\mathcal{M}$ is a tensor of rank $(0, 1)$.

Definition 1.3.7 (Tensor field)

A **tensor field** of rank (r, s) is a smooth map $T : p \in \mathcal{M} \mapsto T_p$ tensor of rank (r, s) at p . It can also be viewed as a smooth map $T : [\Lambda^1(\mathcal{M})]^r \times [\mathfrak{X}(\mathcal{M})]^s \rightarrow \mathcal{C}^\infty(\mathcal{M})$.

Given appropriate bases for vector fields $\{\mathbf{e}_\mu\}_{\mu=1,\dots,n}$ and 1-forms $\{\mathbf{f}^\mu\}_{\mu=1,\dots,n}$, the components of a tensor field are defined as:

$$T^{\mu_1 \dots \mu_r}_{\nu_1 \dots \nu_s} := T(\mathbf{f}^{\mu_1}, \dots, \mathbf{f}^{\mu_r}, \mathbf{e}_{\nu_1}, \dots, \mathbf{e}_{\nu_s}) \quad (1.3.10)$$

Proposition 1.3.3 (Components of tensor fields)

On an n -dimensional manifold, a (r, s) tensor field has n^{r+s} components, each being an element of $\mathcal{C}^\infty(\mathcal{M})$.

Consider two general basis transformations, for vector fields and 1-forms, described by invertible matrices A and B such that $\tilde{\mathbf{e}}_\mu = A^\nu_\mu \mathbf{e}_\nu$ and $\tilde{\mathbf{f}}^\mu = B^\mu_\nu \mathbf{f}^\nu$, with necessary condition $A^\mu_\nu B^\rho_\mu = \delta_\nu^\rho$ to ensure duality: this implies $B = A^{-1}$, i.e. covectors transform inversely with respect to vectors. Thus:

$$\tilde{T}^{\mu_1 \dots \mu_r}_{\nu_1 \dots \nu_s} = B^{\mu_1}_{\rho_1} \dots B^{\mu_r}_{\rho_r} A^{k_1}_{\nu_1} \dots A^{k_s}_{\nu_s} T^{\rho_1 \dots \rho_r}_{k_1 \dots k_s} \quad (1.3.11)$$

If the considered basis are coordinate basis, then $A^\mu_\nu = \frac{\partial x^\mu}{\partial y^\nu}$ and $B^\mu_\nu = \frac{\partial y^\mu}{\partial x^\nu}$.

§1.3.4 Operations on tensors

Algebraic addition and multiplication by functions are trivially defined on tensors of the same rank, and in fact the space of all (r, s) tensors at a point $p \in \mathcal{M}$ is a vector space.

Definition 1.3.8 (Tensor product)

Given two tensor fields S of rank (p, q) and T of rank (r, s) , their **tensor product** is defined as:

$$\begin{aligned} S \otimes T(\omega_1, \dots, \omega_p, \eta_1, \dots, \eta_r, X_1, \dots, X_q, Y_1, \dots, Y_s) \\ = S(\omega_1, \dots, \omega_p, X_1, \dots, X_q) T(\eta_1, \dots, \eta_r, Y_1, \dots, Y_s) \end{aligned} \quad (1.3.12)$$

or, in components:

$$(S \otimes T)^{\mu_1 \dots \mu_p \nu_1 \dots \nu_r}_{\rho_1 \dots \rho_q \dots \sigma_1 \dots \sigma_s} = S^{\mu_1 \dots \mu_p}_{\rho_1 \dots \rho_q} T^{\nu_1 \dots \nu_r}_{\sigma_1 \dots \sigma_s} \quad (1.3.13)$$

It is also possible to contract tensor $((r, s) \mapsto (r-1, s-1))$: for example, given a rank $(2, 1)$ tensor, a rank $(1, 0)$ tensor can be defined as $S(\omega) = T(\omega, \mathbf{f}^\mu, \mathbf{e}_\mu)$, with components $S^\mu = T^{\nu\mu}_\mu$; obviously, in general, $T^{\nu\mu}_\mu \neq T^{\mu\nu}_\mu$.

It is convenient to introduce some notation: given an object $T_{\mu_1 \dots \mu_n}$ dependent on some indices, its symmetric and antisymmetric parts are respectively defined as:

$$T_{(\mu_1 \dots \mu_n)} := \frac{1}{n!} \sum_{\sigma \in S^n} T_{\sigma(\mu_1) \dots \sigma(\mu_n)} \quad (1.3.14)$$

$$T_{[\mu_1 \dots \mu_n]} := \frac{1}{n!} \sum_{\sigma \in S^n} \operatorname{sgn}(\sigma) T_{\sigma(\mu_1) \dots \sigma(\mu_n)} \quad (1.3.15)$$

Conventionally, indices surrounded by $|\cdot|$ are ignored, e.g. $T_{[\mu|\nu|\rho]} = \frac{1}{2} (T_{\mu\nu\rho} - T_{\rho\nu\mu})$. As previously seen, vector fields are pushed forward and 1-form are pulled back: tensors will thus behave in a mixed way.

Definition 1.3.9 (Push-forward of tensors)

Given a diffeomorphism between two manifolds $\varphi : \mathcal{M} \rightarrow \mathcal{N}$ and a (r, s) tensor field T on \mathcal{M} , the **push-forward** of T is the (r, s) tensor field $\varphi_* T$ on \mathcal{N} such that, for $\{\omega_j\}_{j=1,\dots,r} \subseteq \Lambda^1(\mathcal{N})$ and $\{X_j\}_{j=1,\dots,s} \subseteq \mathfrak{X}(\mathcal{N})$:

$$\varphi_* T(\omega_1, \dots, \omega_r, X_1, \dots, X_s) = T(\varphi^* \omega_1, \dots, \varphi^* \omega_r, \varphi_*^{-1} X_1, \dots, \varphi_*^{-1} X_s) \quad (1.3.16)$$

Definition 1.3.10 (Lie derivative of tensors)

Given a vector field $X \in \mathfrak{X}(\mathcal{M})$ and a (r, s) tensor field T on \mathcal{M} , the **Lie derivative** of T along X is defined as:

$$\mathcal{L}_X T_p := \lim_{t \rightarrow 0} \frac{((\sigma_{-t})_* T)_p - T_p}{t} \quad (1.3.17)$$

where σ_t is the flow generated by X .

It can be shown that the components of the Lie derivative of a general (r, s) tensor are:

$$\begin{aligned} (\mathcal{L}_X T)^{\mu_1 \dots \mu_r}_{\nu_1 \dots \nu_s} &= T^{\mu_1 \dots \mu_r}_{\nu_1 \dots \nu_s, \rho} X^\rho - T^{\rho \mu_2 \dots \mu_r}_{\nu_1 \dots \nu_s} X^{\mu_1}_{,\rho} - \dots - T^{\mu_1 \dots \mu_{r-1} \rho}_{\nu_1 \dots \nu_s} X^{\mu_r}_{,\rho} + \\ &\quad + T^{\mu_1 \dots \mu_r}_{\rho \nu_2 \dots \nu_s} X^\rho_{,\nu_1} + \dots + T^{\mu_1 \dots \mu_r}_{\nu_1 \dots \nu_{s-1} \rho} X^\rho_{,\nu_s} \end{aligned} \quad (1.3.18)$$

§1.4 Differential forms

Definition 1.4.1 (p -forms)

A **p -form** totally anti-symmetric $(0, p)$ tensor. The set of all p -forms over a manifold \mathcal{M} is denoted as $\bigwedge^p(\mathcal{M})$.

Lemma 1.4.1 (Independent components of p -forms)

$$\dim_{\mathbb{R}} \bigwedge^p(\mathcal{M}) = \binom{n}{p} \quad (1.4.1)$$

The maximum degree of differential forms thus is $p = n \equiv \dim_{\mathbb{R}} \mathcal{M}$: forms in $\bigwedge^n(\mathcal{M})$ are called **top forms**.

Definition 1.4.2 (Wedge product)

Given $\omega \in \bigwedge^p(\mathcal{M}), \eta \in \bigwedge^q(\mathcal{M})$, their **wedge product** is a $(p + q)$ -form defined as:

$$(\omega \wedge \eta)_{\mu_1 \dots \mu_p \nu_1 \dots \nu_q} = \frac{(p+q)!}{p!q!} \omega_{[\mu_1 \dots \mu_p} \eta_{\nu_1 \dots \nu_q]} \quad (1.4.2)$$

Example 1.4.1 (1-forms to 2-form)

Given $\omega, \eta \in \bigwedge^1(\mathcal{M})$, their wedge product is $(\omega \wedge \eta)_{\mu\nu} = \omega_\mu \eta_\nu - \omega_\nu \eta_\mu$.

The wedge product is essentially a linear map $\bigwedge^p(\mathcal{M}) \times \bigwedge^q(\mathcal{M}) \rightarrow \bigwedge^{p+q}(\mathcal{M})$. Moreover, it can be shown to be associative.

Proposition 1.4.1 (Skew-symmetry)

Given $\omega \in \bigwedge^p(\mathcal{M}), \eta \in \bigwedge^q(\mathcal{M})$, then:

$$\omega \wedge \eta = (-1)^{pq} \eta \wedge \omega \quad (1.4.3)$$

Then, clearly, $\omega \wedge \omega = 0 \ \forall \omega \in \bigwedge^p(\mathcal{M}) : p \text{ is odd.}$

Proposition 1.4.2 (Bases for p -forms)

If $\{\mathbf{f}^\mu\}_{\mu=1,\dots,n}$ is a basis of $\bigwedge^1(\mathcal{M})$, then $\{\mathbf{f}^{\mu_1} \wedge \dots \wedge \mathbf{f}^{\mu_p}\}_{\mu_1,\dots,\mu_p=1,\dots,n}$ is a basis of $\bigwedge^p(\mathcal{M})$.

Locally $\{dx^\mu\}_{\mu=1,\dots,n}$ is a basis of $T_p^*\mathcal{M}$, thus a general p -form can be locally written as:

$$\omega = \frac{1}{p!} \omega_{\mu_1 \dots \mu_p} dx^{\mu_1} \wedge \dots \wedge dx^{\mu_p} \quad (1.4.4)$$

Definition 1.4.3 (Exterior derivative)

The **exterior derivative** is a map $d : \bigwedge^p(\mathcal{M}) \rightarrow \bigwedge^{p+1}(\mathcal{M})$ defined as:

$$(d\omega)_{\mu_1 \dots \mu_{p+1}} = (p+1) \partial_{[\mu_1} \omega_{\mu_2 \dots \mu_{p+1}]} \quad (1.4.5)$$

In local coordinates:

$$d\omega = \frac{1}{p!} \frac{\partial \omega_{\mu_1 \dots \mu_p}}{\partial x^\nu} dx^\nu \wedge dx^{\mu_1} \wedge \dots \wedge dx^{\mu_p} \quad (1.4.6)$$

Theorem 1.4.1 (Poincaré's theorem)

$$d^2 = 0 \quad (1.4.7)$$

Proof. Consequence of Schwarz lemma. \square

Proposition 1.4.3 (Product rule)

Given $\omega \in \Lambda^p(\mathcal{M}), \eta \in \Lambda^q(\mathcal{M})$, then:

$$d(\omega \wedge \eta) = d\omega \wedge \eta + (-1)^p \omega \wedge d\eta \quad (1.4.8)$$

Proposition 1.4.4 (Pull-back and exterior derivative)

Given a diffeomorphism between two manifolds $\varphi : \mathcal{M} \rightarrow \mathcal{N}$ and $\omega \in \Lambda^p(\mathcal{M})$, then $d(\varphi^*\omega) = \varphi^*(d\omega)$.

Corollary 1.4.4.1 (Lie derivative and exterior derivative)

Given $X \in \mathfrak{X}(\mathcal{M}), \omega \in \Lambda^p(\mathcal{M})$, then $d(\mathcal{L}_X \omega) = \mathcal{L}_X(d\omega)$.

Definition 1.4.4 (Closed and exact forms)

Given $\omega \in \Lambda^p(\mathcal{M})$, it is:

- **closed** if $d\omega = 0$;
- **exact** if $\exists \eta \in \Lambda^{p-1}(\mathcal{M}) : \omega = d\eta$.

By Poincaré's theorem, clearly $\omega \in \Lambda^p(\mathcal{M})$ exact $\implies \omega$ closed, while the converse is not true in general, but depends on the topology of the manifold: a full treatment of this matter led to the development of cohomology, as outlined in §1.4.1.

Definition 1.4.5 (Interior product)

Given a vector field $X \in \mathfrak{X}(\mathcal{M})$, the **interior product** determined by X is the linear map $\iota_X : \Lambda^p(\mathcal{M}) \rightarrow \Lambda^{p-1}(\mathcal{M})$ defined as:

$$\iota_X \omega(Y_1, \dots, Y_{p-1}) := \omega(X, Y_1, \dots, Y_{p-1}) \quad (1.4.9)$$

On 0-forms (i.e. scalar functions), the interior product is defined as $\iota_X f \equiv 0 \ \forall X \in \mathfrak{X}(\mathcal{M})$.

Proposition 1.4.5 (Anti-commutation)

Given $X, Y \in \mathfrak{X}(\mathcal{M})$, then $\iota_X \iota_Y = -\iota_Y \iota_X$.

Proof. Consequence of the total anti-symmetry of p -forms. \square

Proposition 1.4.6 (Distributivity)

Given $X \in \mathfrak{X}(\mathcal{M})$, $\omega \in \Lambda^p(\mathcal{M})$, $\eta \in \Lambda^q(\mathcal{M})$, then:

$$\iota_X(\omega \wedge \eta) = \iota_X \omega \wedge \eta + (-1)^p \omega \wedge \iota_X \eta \quad (1.4.10)$$

Theorem 1.4.2 (Cartan's theorem)

Given a vector field $X \in \mathfrak{X}(\mathcal{M})$, then:

$$\mathcal{L}_X = d \circ \iota_X + \iota_X \circ d \quad (1.4.11)$$

Proof. Consider $\omega \in \Lambda^1(\mathcal{M})$:

$$\begin{aligned} \iota_X(d\omega) &= \iota_X \frac{1}{2} (\partial_\mu \omega_\nu - \partial_\nu \omega_\mu) dx^\mu \wedge dx^\nu = X^\mu \partial_\mu \omega_\nu dx^\nu - X^\nu \partial_\mu \omega_\nu dx^\mu \\ d(\iota_X \omega) &= d(\omega_\mu X^\mu) = X^\mu \partial_\nu \omega_\mu dx^\nu + \omega_\mu \partial_\nu X^\mu dx^\nu \end{aligned}$$

Thus, adding these expressions and recalling Eq. 1.3.8:

$$(d\iota_X + \iota_X d)\omega = (X^\mu \partial_\mu \omega_\nu + \omega_\mu \partial_\nu X^\mu) dx^\nu = \mathcal{L}_X \omega$$

The $p > 1$ case is more complex. □

§1.4.1 de Rham cohomology

While exact \implies closed, the converse is not true, in general: it depends on the topological properties of the manifold.

Lemma 1.4.2 (Poincaré's lemma)

If \mathcal{M} is simply connected, then $\omega \in \Lambda^p(\mathcal{M})$ closed \implies ω exact.

In general, it is always possible to choose a simply connected neighbourhood of a point $p \in \mathcal{M}$, in which every closed form is exact, but that may not always be possible globally.

It is convenient to set the notation $d_p \equiv d : \Lambda^p(\mathcal{M}) \rightarrow \Lambda^{p+1}(\mathcal{M})$, so that the set of all closed p -forms on \mathcal{M} is denoted by $Z^p(\mathcal{M}) := \ker d_p$, while the set of all exact p -forms by $B^p(\mathcal{M}) := \text{ran } d_{p-1}$.

Definition 1.4.6 (Equivalent forms)

Two closed p -forms $\omega, \omega' \in Z^p(\mathcal{M})$ are said to be **equivalent** if $\omega = \omega' + \eta$ for some $\eta \in B^p(\mathcal{M})$.

Definition 1.4.7 (de Rahm cohomology)

The p^{th} **de Rham cohomology group** of a manifold \mathcal{M} is defined to be:

$$H^p(\mathcal{M}) := Z^p(\mathcal{M}) / B^p(\mathcal{M}) \quad (1.4.12)$$

and its dimension is the p^{th} **Betti number** of \mathcal{M} :

$$B_p := \dim_{\mathbb{R}} H^p(\mathcal{M}) \quad (1.4.13)$$

Theorem 1.4.3 (Betti numbers)

Given a differentiable manifold, its Betti numbers are always finite.

$B_0 = 1$ for any connected manifold: there exist constant functions, which are manifestly closed and not exact, due to the non-existence of “ (-1) -forms”. Higher Betti numbers are non-zero only if the manifold has some non-trivial topology.

Definition 1.4.8 (Euler's character)

The **Euler's character** of a manifold \mathcal{M} is defined as:

$$\chi(\mathcal{M}) := \sum_{p \in \mathbb{N}_0} (-1)^p B_p \quad (1.4.14)$$

Example 1.4.2 (Spheres and tori)

The n -sphere \mathbb{S}^n has only non-vanishing $B_0 = B_n = 1$, thus $\chi(\mathbb{S}^n) = 1 + (-1)^n$, while the n -torus \mathbb{T}^n has $B_p = \binom{n}{p}$, hence $\chi(\mathbb{T}^n) = 0$.

§1.4.2 Integration

Definition 1.4.9 (Volume form)

A **volume form** on an n -dimensional differentiable manifold \mathcal{M} is a nowhere-vanishing top form dV , i.e. locally $dV = v(x)dx^1 \wedge \cdots \wedge dx^n : v(x) \neq 0$.

If such a form exists, the manifold is said to be **orientable**, and its orientation is right/left-handed if respectively $v(x) > 0$ or $v(x) < 0$ locally on every subset of \mathcal{M} .

To ensure that the handedness of the manifold doesn't change on overlapping charts:

$$dV = v(x) \frac{\partial x^1}{\partial y^{\mu_1}} dy^{\mu_1} \wedge \cdots \wedge \frac{\partial x^n}{\partial y^{\mu_n}} dy^{\mu_n} = v(x) \det \left[\frac{\partial x^\mu}{\partial y^\nu} \right] dy^1 \wedge \cdots \wedge dy^n$$

It is therefore necessary that the two sets of coordinates on the overlapping region satisfy:

$$\det \left[\frac{\partial x^\mu}{\partial y^\nu} \right] > 0 \quad (1.4.15)$$

Non-orientable manifolds cannot be covered by overlapping charts satisfying this condition.

Example 1.4.3 (Projective spaces)

The real projective space \mathbb{RP}^n is orientable for odd n and non-orientable for even n , while the complex projective space \mathbb{CP}^n is orientable for all $n \in \mathbb{N}$.

Definition 1.4.10 (Integration over manifolds)

Given a function $f : \mathcal{M} \rightarrow \mathbb{R}$ on an orientable manifold \mathcal{M} with volume form dV and a chart (φ, U) on \mathcal{M} with coordinates $\{x^\mu\}_{\mu=1,\dots,n}$, the **integral** of f on $O = \varphi^{-1}(U) \subset \mathcal{M}$ is defined as:

$$\int_O f dV := \int_U dx_1 \dots dx_n f(x) v(x) \quad (1.4.16)$$

It is clear that the volume form acts like a measure on the manifold. To integrate over the whole manifold, it must be divided up into different regions, each covered by a single chart.

Definition 1.4.11 (Submanifold)

A k -dimensional manifold Σ is a **submanifold** of an n -dimensional manifold \mathcal{M} , with $n > k$, if there exists an injective map $\varphi : \Sigma \rightarrow \mathcal{M}$ such that $\varphi_* : T_p(\Sigma) \rightarrow T_{\varphi(p)}(\mathcal{M})$ is injective.

Definition 1.4.12 (Integration over submanifolds)

Given a k -form $\omega \in \Lambda^k(\mathcal{M})$, its integral over a k -dimensional submanifold Σ of \mathcal{M} is defined as:

$$\int_{\varphi(\Sigma)} \omega := \int_{\Sigma} \varphi^* \omega \quad (1.4.17)$$

Example 1.4.4 (Integration over a curve)

Consider a 1-form $\omega \in \Lambda^1(\mathcal{M})$ and a 1-dimensional submanifold γ of \mathcal{M} described by a curve $\sigma : \gamma \rightarrow \mathcal{M} : x^\mu = \sigma^\mu(t)$: locally $\omega = \omega_\mu(x) dx^\mu$, thus the integral of ω on γ can be calculated as $\int_{\sigma(\gamma)} \omega = \int_{\gamma} \sigma^* \omega = \int_{\gamma} d\tau \omega_\mu(x) \frac{dx^\mu}{d\tau}$.

§1.4.2.1 Stokes' theorem

Integration can be generalized beyond smooth (i.e. differentiable) manifolds.

Definition 1.4.13 (Manifold with boundary)

An n -dimensional **manifold with boundary** is a Hausdorff topological space, equipped with a compatible maximal atlas, which is locally homeomorphic to $\mathbb{R}^{n-1} \times [a, \infty) : a \in \mathbb{R}$.

The **boundary** $\partial\mathcal{M}$ is the 1-dimensional submanifold determined parametrically by $x^n = a$.

Theorem 1.4.4 (Stokes' theorem)

Given an n -dimensional manifold \mathcal{M} with boundary $\partial\mathcal{M}$, then for any $\omega \in \Lambda^{n-1}(\mathcal{M})$:

$$\int_{\mathcal{M}} d\omega = \int_{\partial\mathcal{M}} \omega \quad (1.4.18)$$

This important theorem unifies many different results.

Fundamental theorem of calculus Given the 1-dimensional manifold $I = [a, b] \subset \mathbb{R}$, then for any 0-form (i.e. scalar function) $\omega = \omega(x)$:

$$\int_I d\omega = \int_a^b \frac{d\omega}{dx} dx = \int_{\partial I} \omega = \omega(b) - \omega(a)$$

Green's theorem Given a 2-dimensional manifold with boundary $S \subset \mathbb{R}^2$ and a 1-form $\omega = \omega_1 dx^1 + \omega_2 dx^2$, then $d\omega = (\partial_1 \omega_2 - \partial_2 \omega_1) dx^1 \wedge dx^2$ and:

$$\int_S d\omega = \int_S \left(\frac{\partial \omega_2}{\partial x^1} - \frac{\partial \omega_1}{\partial x^2} \right) dx^1 \wedge dx^2 = \int_{\partial S} \omega = \int_{\partial S} \omega_1 dx^1 + \omega_2 dx^2$$

Gauss' theorem Given a 3-dimensional manifold with boundary $V \subset \mathbb{R}^3$ and a 2-form $\omega = \omega_1 dx^2 \wedge dx^3 + \omega_2 dx^3 \wedge dx^1 + \omega_3 dx^1 \wedge dx^2$, then $d\omega = (\partial_1 \omega_1 + \partial_2 \omega_2 + \partial_3 \omega_3) dx^1 \wedge dx^2 \wedge dx^3$ and:

$$\int_V d\omega = \int_V \left(\frac{\partial \omega_1}{\partial x^1} + \frac{\partial \omega_2}{\partial x^2} + \frac{\partial \omega_3}{\partial x^3} \right) = \int_{\partial V} \omega = \int_{\partial V} \omega_1 dx^2 \wedge dx^3 + \omega_2 dx^3 \wedge dx^1 + \omega_3 dx^1 \wedge dx^2$$

Chapter 2

Riemannian Geometry

§2.1 Metric manifolds

Definition 2.1.1 (Metric)

A **metric** g is a $(0,2)$ tensor field on a manifold \mathcal{M} that is:

1. symmetric: $g(X, Y) = g(Y, X)$;
2. non-degenerate: $\exists p \in \mathcal{M} : g(X, Y)|_p = 0 \ \forall Y \in T_p \mathcal{M} \implies X_p = 0$.

A manifold equipped with a metric (\mathcal{M}, g) is called a **metric manifold**.

With a choice of coordinates, the metric can be written as:

$$g = g_{\mu\nu}(x) dx^\mu \otimes dx^\nu \quad (2.1.1)$$

where:

$$g_{\mu\nu} = g \left(\frac{\partial}{\partial x^\mu}, \frac{\partial}{\partial x^\nu} \right) \quad (2.1.2)$$

It is often written also as $ds^2 = g_{\mu\nu}(x) dx^\mu dx^\nu$. The matrix $g_{\mu\nu}(x) \in \mathbb{R}^{n \times n}$ is symmetric, and there's always a choice of basis on each tangent space such that this matrix is diagonal: the non-degeneracy condition implies that none of the diagonal elements vanish.

Proposition 2.1.1 (Signature)

The **signature** of a metric, i.e. the number of negative entries when diagonalized, is independent on the choice of basis.

Proof. From Sylvester's theorem of inertia. □

§2.1.1 Riemannian manifolds

Definition 2.1.2 (Riemannian manifold)

A **Riemannian manifold** (\mathcal{M}, g) is a manifold equipped with a metric with totally-positive signature.

Example 2.1.1 (Euclidean space)

The Euclidean space \mathbb{R}^n , equipped with the metric $g_{\mu\nu} = \delta_{\mu\nu}$ (in Cartesian coordinates), is a Riemannian manifold.

In a Riemannian manifold (\mathcal{M}, g) and $X \in \mathfrak{X}(\mathcal{M})$, the **length** of X at $p \in \mathcal{M}$ is defined as:

$$|X_p| := \sqrt{g(X, X)|_p} \quad (2.1.3)$$

Given $Y \in \mathfrak{X}(\mathcal{M})$, the **angle** between X and Y at $p \in \mathcal{M}$ is defined as:

$$\cos \theta := \frac{g(X, Y)|_p}{|X_p| |Y_p|} \quad (2.1.4)$$

This can be generalized to distances between points on a curve $\sigma : \mathbb{R} \rightarrow \mathcal{M}$:

$$d(p, q) = \int_a^b dt \sqrt{g(X, X)|_{\sigma(t)}} \quad (2.1.5)$$

where $\sigma(a) = p$, $\sigma(b) = q$ and X is the tangent vector field of the curve. With parametrization $x^\mu(t)$, the tangent vector has components $X^\mu = \frac{dx^\mu}{dt}$, thus:

$$d(p, q) = \int_a^b dt \sqrt{g_{\mu\nu}(x) \frac{dx^\mu}{dt} \frac{dx^\nu}{dt}} \quad (2.1.6)$$

It is important to note that this distance is independent of the parametrization.

§2.1.2 Lorentzian manifolds**Definition 2.1.3** (Lorentzian manifold)

A **Lorentzian manifold** (\mathcal{M}, g) is a manifold equipped with a metric which has a signature with a single negative sign.

Example 2.1.2 (Lorentz–Minkowski metric)

The simplest Lorentzian manifold is \mathbb{R}^n with the **Lorentz–Minkowski metric**:

$$\eta = -dx^0 \otimes dx^0 + dx^1 \otimes dx^1 + \cdots + dx^{n-1} \otimes dx^{n-1} \quad (2.1.7)$$

Its components are $\eta_{\mu\nu} = \text{diag}(-1, +1, \dots, +1)$, thus this is a Lorentzian manifold.

On a general Lorentzian manifold, at any point $p \in \mathcal{M}$ it is always possible to choose an orthonormal basis $\{e_\mu\}_{\mu=0,\dots,n-1}$ of $T_p \mathcal{M}$ such that $g_{\mu\nu}|_p = \eta_{\mu\nu}$: this fact is closely related to the equivalence principle. Consider a different basis $\tilde{e}_\mu = \Lambda^\nu{}_\mu e_\nu$: the condition for it to leave the Lorentz–Minkowski metric unchanged is:

$$\eta_{\mu\nu} = \Lambda^\rho{}_\mu \Lambda^\sigma{}_\nu \eta_{\rho\sigma} \quad (2.1.8)$$

This is the defining equation of a Lorentz transformation: on a Lorentzian manifold, the basic features of special relativity are locally recovered. Thus, other ideas from special relativity can be imported.

Definition 2.1.4 (Curves in a Lorentzian manifold)

Given a Lorentzian manifold (\mathcal{M}, g) and $X \in \mathfrak{X}(\mathcal{M})$, at $p \in \mathcal{M}$ the vector field is said to be **timelike** if $g(X_p, X_p) < 0$, **null** if $g(X_p, X_p) = 0$ or **spacelike** if $g(X_p, X_p) > 0$.

At each point $p \in \mathcal{M}$ it is possible to draw **lightcones**, i.e. the null tangent vectors at that point, which are past-directed or future-directed: these lightcones vary smoothly as the point is varied smoothly on the manifold, elucidating the causal structure of spacetime.

The distance between two points on a curve depends on the nature of the tangent vector field of the curve: a *timelike curve* is a curve whose tangent vector field is everywhere timelike, and analogously for the other cases. The distance on a spacelike curve is defined as in Eq. 2.1.5, while that on a timelike curve gets a negative sign in the square root. With parametrization $x^\mu(t)$, it is possible to define the **proper time** on a timelike curve as:

$$\tau = \int_a^b dt \sqrt{-g_{\mu\nu}(x) \frac{dx^\mu}{dt} \frac{dx^\nu}{dt}} \quad (2.1.9)$$

This is precisely the action of a free particle moving in spacetime.

§2.1.3 Metric properties

The metric defines a natural isomorphism between vectors and covectors.

Proposition 2.1.2

Given a metric manifold (\mathcal{M}, g) , the metric defines for each $p \in \mathcal{M}$ a natural isomorphism $g : X_p \in T_p \mathcal{M} \rightarrow \omega_p \in T_p^* \mathcal{M} : \omega_p(Y_p) = g(X_p, Y_p) \forall Y_p \in T_p \mathcal{M}$.

Truly, this is an isomorphism between the tangent bundle $T\mathcal{M}$ and the cotangent bundle $T^*\mathcal{M}$ of \mathcal{M} , which are defined as:

$$T\mathcal{M} := \bigsqcup_{p \in \mathcal{M}} T_p \mathcal{M} \quad T^*\mathcal{M} := \bigsqcup_{p \in \mathcal{M}} T_p^* \mathcal{M} \quad (2.1.10)$$

Then, the natural isomorphism determined by the metric is divided into two musical isomorphism: the **flat isomorphism** $\flat : T\mathcal{M} \rightarrow T^*\mathcal{M}$ and the **sharp isomorphism** $\sharp : T^*\mathcal{M} \rightarrow T\mathcal{M}$. In a chosen coordinate basis, the vector $X = X^\mu \partial_\mu$ is mapped to the one-form $X^\flat = X_\mu dx^\mu$, thus the following identity holds:

$$X_\mu = g_{\mu\nu} X^\nu \quad (2.1.11)$$

Being g non-degenerate, the matrix $g_{\mu\nu}$ is invertible, with inverse $g^{\mu\nu}$ such that:

$$g^{\mu\nu} g_{\nu\rho} = \delta_\rho^\mu \quad (2.1.12)$$

Its elements are the components of a (2,0) symmetric tensor $\hat{g} := g^{\mu\nu} \partial_\mu \otimes \partial_\nu$, which defines the inverse of the natural isomorphism in Prop. 2.1.2:

$$X^\mu = g^{\mu\nu} X_\nu \quad (2.1.13)$$

The metric also defines a natural volume form on the manifold:

$$dV := \sqrt{g} dx^1 \wedge \cdots \wedge dx^n \quad (2.1.14)$$

where $g \equiv |\det g_{\mu\nu}|$.

Proposition 2.1.3

The volume form is basis-independent.

Proof. Consider a new set of coordinates y^μ such that $dx^\mu = A^\mu{}_\nu dy^\nu$, where $A^\mu{}_\nu = \frac{\partial x^\mu}{\partial y^\nu}$. In general:

$$dx^1 \wedge \cdots \wedge dx^n = A^1{}_{\mu_1} \cdots A^n{}_{\mu_n} dy^{\mu_1} \wedge \cdots \wedge dy^{\mu_n}$$

Recalling the anti-symmetry of the wedge product and the definition of determinant, this can be rewritten as:

$$dx^1 \wedge \cdots \wedge dx^n = \sum_{\pi \in S^n} \text{sgn}(\pi) A^1{}_{\pi(1)} \cdots A^n{}_{\pi(n)} dy^1 \wedge \cdots \wedge dy^n = \det A dy^1 \wedge \cdots \wedge dy^n$$

Note the Jacobian factor which arises when changing the measure. On the other hand:

$$g_{\mu\nu} = \frac{\partial y^\rho}{\partial x^\mu} \frac{\partial y^\sigma}{\partial x^\nu} \tilde{g}_{\rho\sigma} = (A^{-1})^\rho{}_\mu (A^{-1})^\sigma{}_\nu \tilde{g}_{\rho\sigma} \implies \det g_{\mu\nu} = \frac{\det \tilde{g}_{\mu\nu}}{(\det A)^2}$$

The factors $\det A$ and $(\det A)^{-1}$ cancel, thus yielding the thesis. \square

The volume form can be rewritten as:

$$dV = \frac{1}{n!} v_{\mu_1 \dots \mu_n} dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_n} \equiv \frac{1}{n!} \sqrt{g} \epsilon_{\mu_1 \dots \mu_n} dx^{\mu_1} \wedge \cdots \wedge dx^{\mu_n} \quad (2.1.15)$$

where $\epsilon_{\mu_1 \dots \mu_n}$ is the totally-antisymmetric n -dimensional Levi-Civita symbol. $\epsilon_{\mu_1 \dots \mu_n}$ cannot be considered a proper tensor, as its components are always $+1, -1, 0$ independently if the indices are covariant or contravariant: it is, in fact, a **tensor density**, i.e. a tensor divided by \sqrt{g} . It can be shown that:

$$v^{\mu_1 \dots \mu_n} = g^{\mu_1 \nu_1} \cdots g^{\mu_n \nu_n} v_{\mu_1 \dots \mu_n} = \sigma \frac{1}{\sqrt{g}} \epsilon^{\mu_1 \dots \mu_n} \quad (2.1.16)$$

where σ is the sign of the signature, e.g. $\sigma = +1$ for Riemannian manifolds and $\sigma = -1$ for Lorentzian manifolds. As notation, the integral of a generic function f on \mathcal{M} is denoted as:

$$\int_{\mathcal{M}} f dV \equiv \int_{\mathcal{M}} d^n x \sqrt{g} f \quad (2.1.17)$$

§2.1.3.1 Hodge theory**Definition 2.1.5 (Hodge dual)**

Given an n -dimensional oriented metric manifold (\mathcal{M}, g) , the **Hodge dual** is defined as the map $\star : \bigwedge^p(\mathcal{M}) \rightarrow \bigwedge^{n-p}(\mathcal{M}) : \omega \mapsto \star \omega$ such that:

$$\star \omega_{\mu_1 \dots \mu_{n-p}} := \frac{1}{(n-p)!} \sqrt{g} \epsilon_{\mu_1 \dots \mu_{n-p} \nu_1 \dots \nu_p} \omega^{\nu_1 \dots \nu_p} \quad (2.1.18)$$

In this section, the orientedness and n -dimensionality of the manifold \mathcal{M} are implied.

Proposition 2.1.4 (Basis-independence)

The Hodge dual is basis-independent.

It is useful to state a lemma for future calculations.

Lemma 2.1.1 (Generalized Kronecker delta)

$$v^{\mu_1 \dots \mu_p \rho_1 \dots \rho_{n-p}} v_{\nu_1 \dots \nu_p \rho_1 \dots \rho_{n-p}} = \sigma p! (n-p)! \delta_{[\nu_1}^{\mu_1} \dots \delta_{\nu_p]}^{\mu_p}.$$

Proposition 2.1.5 (Nilpotence)

$$\star(\star\omega) = \sigma (-1)^{p(n-p)} \omega.$$

The Hodge dual defines an inner product on each $\Lambda^p(\mathcal{M})$:

$$\langle \omega, \eta \rangle := \int_{\mathcal{M}} \omega \wedge \star \eta \quad (2.1.19)$$

This allows to define operators and their adjoints on form spaces.

Proposition 2.1.6 (Adjoint exterior derivative)

Given a metric manifold (\mathcal{M}, g) and two forms $\omega \in \Lambda^p(\mathcal{M}), \alpha \in \Lambda^{p-1}(\mathcal{M})$, then:

$$\langle d\alpha, \omega \rangle = \langle \alpha, d^\dagger \omega \rangle \quad (2.1.20)$$

where the adjoint of the exterior derivative $d^\dagger : \Lambda^p(\mathcal{M}) \rightarrow \Lambda^{p-1}(\mathcal{M})$ is defined as:

$$d^\dagger := \sigma (-1)^{np+n-1} \star d \star \quad (2.1.21)$$

Proof. To simplify the proof, consider a closed manifold; then, from Stokes' theorem and Eq. 1.4.8:

$$0 = \int_{\mathcal{M}} d(\alpha \wedge \star \omega) = \langle d\alpha, \omega \rangle + \int_{\mathcal{M}} (-1)^{p-1} \alpha \wedge d \star \omega$$

The second term is proportional to $\langle \alpha, \star d \star \omega \rangle$: to determine the relative sign, note that $d \star \omega \in \Lambda^{n-p+1}(\mathcal{M})$, thus, from Prop. 2.1.5, $\star \star d \star \omega = \sigma (-1)^{(n-p+1)(p-1)} d \star \omega$. In conclusion:

$$\begin{aligned} \langle \alpha, \star d \star \omega \rangle &= \sigma (-1)^{(n-p)(p-1)} \int_{\mathcal{M}} (-1)^{p-1} \alpha \wedge d \star \omega \\ &\implies \langle d\alpha, \omega \rangle = \sigma (-1)^{(n-p)(p-1)+1} \langle \alpha, \star d \star \omega \rangle \end{aligned}$$

Noting that $(-1)^{(n-p)(p-1)+1} = (-1)^{np+n-1}$, as in general $(-1)^{-n} = (-1)^n$ and $(-1)^{-p^2+p+1} = (-1)^{-1}$ due to $p(p-1)$ being always even, concludes the proof. \square

Definition 2.1.6 (Laplacian)

Given a metric manifold (\mathcal{M}, g) , the **Laplacian** $\Delta : \Lambda^p(\mathcal{M}) \rightarrow \Lambda^p(\mathcal{M})$ is defined as the operator:

$$\Delta := (d + d^\dagger)^2 \quad (2.1.22)$$

Clearly, since $d^2 = d^\dagger 2 = 0$.

$$\Delta = dd^\dagger + d^\dagger d \equiv \{d, d^\dagger\} \quad (2.1.23)$$

It is possible to compute an explicit expression for the Laplacian of functions.

Lemma 2.1.2

Given $f \in \mathcal{C}^\infty(\mathcal{M})$, then $d^\dagger f = 0$.

Proof. Trivial noting that $\star f$ is a top-form. \square

Proposition 2.1.7 (G)

Given $f \in \mathcal{C}^\infty(\mathcal{M})$, then:

$$\Delta f = -\frac{\sigma}{\sqrt{g}} \partial_\nu (\sqrt{g} g^{\mu\nu} \partial_\mu f) \quad (2.1.24)$$

Proof. Via direct calculation, using Lemma 2.1.2:

$$\begin{aligned} \Delta f &= \sigma (-1)^{n^2+n-1} \star d \star (\partial_\mu f dx^\mu) = -\sigma \star d (\partial_\mu f \star dx^\mu) \\ &= -\frac{\sigma}{(n-1)!} \star d (\partial_\mu f g^{\mu\nu} \sqrt{g} \epsilon_{\nu\rho_1 \dots \rho_{n-1}} dx^{\rho_1} \wedge \dots \wedge dx^{\rho_{n-1}}) \\ &= -\frac{\sigma}{(n-1)!} \star \partial_\alpha (\sqrt{g} g^{\mu\nu} \partial_\mu f) \epsilon_{\nu\rho_1 \dots \rho_{n-1}} dx^\alpha \wedge dx^{\rho_1} \wedge \dots \wedge dx^{\rho_{n-1}} \\ &= -\sigma \star \partial_\nu (\sqrt{g} g^{\mu\nu} \partial_\mu f) dx^1 \wedge \dots \wedge dx^n = -\frac{\sigma}{\sqrt{g}} \partial_\nu (\sqrt{g} g^{\mu\nu} \partial_\mu f) \end{aligned}$$

which is the thesis. \square

The Laplacian operator is linked to the de Rham cohomology.

Definition 2.1.7 (Harmonic p -forms)

Given $\omega \in \bigwedge^p(\mathcal{M})$, it is said to be **harmonic** if $\Delta\omega = 0$. The space of harmonic p -forms on (\mathcal{M}, g) is denoted as $\text{Harm}^p(\mathcal{M})$.

Proposition 2.1.8

A harmonic form is both **closed** and **co-closed**.

Proof. $0 = \langle \omega, \Delta\omega \rangle = \langle d\omega, d\omega \rangle + \langle d^\dagger\omega, d^\dagger\omega \rangle$, thus $d\omega = 0$ and $d^\dagger\omega = 0$, for the inner product is positive-defined. \square

Theorem 2.1.1 (Harmonic decomposition of forms)

Given a compact Riemannian manifold (\mathcal{M}, g) , any $\omega \in \bigwedge^p(\mathcal{M})$ can be uniquely decomposed as $\omega = d\alpha + d^\dagger\beta + \gamma$, with $\alpha \in \bigwedge^{p-1}(\mathcal{M})$, $\beta \in \bigwedge^{p+1}(\mathcal{M})$ and $\gamma \in \text{Harm}^p(\mathcal{M})$.

Theorem 2.1.2 (Hodge's theorem)

Given a compact Riemannian manifold (\mathcal{M}, g) , then:

$$\text{Harm}^p(\mathcal{M}) \cong H^p(\mathcal{M}) \quad (2.1.25)$$

Proof. From Prop. 2.1.8 $\text{Harm}^p(\mathcal{M}) \subset Z^p(\mathcal{M})$, but the uniqueness of decomposition in Th. 2.1.1 implies $\forall \gamma \in \text{Harm}^p(\mathcal{M}) \exists \eta_\gamma \in \Lambda^{p-1}(\mathcal{M}) : \gamma \neq d\eta_\gamma$, thus $\text{Harm}^p(\mathcal{M}) \subset H^p(\mathcal{M})$. WTS that any equivalence class $[\omega] \in H^p(\mathcal{M})$ can be represented by a harmonic form. By Th. 2.1.1 $\omega = d\alpha + d^\dagger\beta + \gamma$, but $\omega \in H^p(\mathcal{M})$ implies $d\omega = 0$ by definition, so:

$$0 = \langle d\omega, \beta \rangle = \langle \omega, d^\dagger\beta \rangle = \langle d\alpha + d^\dagger\beta + \gamma, d^\dagger\beta \rangle = \langle d^\dagger\beta, d^\dagger\beta \rangle$$

The inner product is positive-definite, thus $d^\dagger\beta = 0$, hence $\omega = \gamma + d\alpha$. By definition $H^p(\mathcal{M}) := Z^p(\mathcal{M})/B^p(\mathcal{M})$, so $[\omega] = \gamma$. \square

Betti numbers can then be computed as $B_p = \dim_{\mathbb{R}} \text{Harm}^p(\mathcal{M})$.

§2.2 Connections

There is a different way to differentiate tensor fields, distinct from the Lie derivative, associated to a different way to map different vector spaces at different points: the covariant derivative. For the rest of this chapter, \mathcal{M} is implied to be an n -dimensional metric manifold with metric tensor g .

§2.2.1 Covariant derivative

Definition 2.2.1 (Connection)

The **connection** is a map $\nabla : \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \rightarrow \mathfrak{X}(\mathcal{M})$, usually written as $\nabla(X, Y) \equiv \nabla_X Y$, where ∇_X is called the **covariant derivative**, satisfying the following properties for all $X, Y, Z \in \mathfrak{X}(\mathcal{M})$:

1. $\nabla_X(Y + Z) = \nabla_X Y + \nabla_X Z$;
2. $\nabla_{fX+gY}Z = f\nabla_X Z + g\nabla_Y Z \quad \forall f, g \in \mathcal{C}^\infty(\mathcal{M})$;
3. $\nabla_X(fY) = f\nabla_X Y + X(f)Y \quad \forall f \in \mathcal{C}^\infty(\mathcal{M})$.

Usually $X(f) \equiv \nabla_X f$. The covariant derivative endows the manifold with more structure: in particular, given a basis $\{e_\mu\}_{\mu=1,\dots,n}$ of $\mathfrak{X}(\mathcal{M})$, its covariant derivative is expressed as:

$$\nabla_{e_\rho} e_\nu \equiv \Gamma_{\rho\nu}^\mu e_\mu \tag{2.2.1}$$

The $\Gamma_{\rho\nu}^\mu$ are the components of the connection on that basis. Conventionally $\nabla_{e_\mu} \equiv \nabla_\mu$, thus resembling a partial derivative. To elucidate how the covariant derivative acts on vector fields:

$$\begin{aligned} \nabla_X Y &= \nabla_X(Y^\mu e_\mu) \\ &= X(Y^\mu) e_\mu + Y^\mu \nabla_X e_\mu \\ &= X^\nu e_\nu(Y^\mu) e_\mu + Y^\mu X^\nu \nabla_\nu e_\mu \\ &= X^\nu [e_\nu(Y^\mu) + \Gamma_{\nu\rho}^\mu Y^\rho] e_\mu \\ &= X^\nu \nabla_\nu Y = X^\nu (\nabla_\nu Y)^\mu e_\mu \end{aligned}$$

The dependency on X can therefore be eliminated, and in components:

$$(\nabla_\nu Y)^\mu = e_\nu(Y^\mu) + \Gamma_{\nu\rho}^\mu Y^\rho \tag{2.2.2}$$

A sloppy notation is often used: $(\nabla_\nu Y)^\mu \equiv \nabla_\nu Y^\mu$. This must not be confused as the covariant derivative of Y^μ . Moreover $\nabla_\nu Y^\mu \equiv Y_{;\nu}^\mu$, while $\partial_\mu f \equiv f_{,\mu}$. On the coordinate basis $e_\mu = \partial_\mu$, then:

$$Y_{;\nu}^\mu = Y_{,\nu}^\mu + \Gamma_{\nu\rho}^\mu Y^\rho \tag{2.2.3}$$

Note that $Y_{;\nu}^\mu$ is the μ^{th} component of $\nabla_\nu Y$, while $Y_{,\nu}^\mu$ is the partial derivative of Y^μ along ∂_ν .

The covariant derivative coincides with other derivatives on $\mathcal{C}^\infty(\mathcal{M})$: it can be shown that $\nabla_X f = \mathcal{L}_X f = X(f)$ and $\nabla_\mu f = \partial_\mu f$. On $\mathfrak{X}(\mathcal{M})$, however, ∇_X and \mathcal{L}_X are distinct: while $\nabla_X = X^\mu \nabla_\mu$, there is no way to write the same relation for \mathcal{L}_X , for it depends not only on X but on its first derivative too. The covariant derivative is thus the natural generalization of the partial derivative to curved manifolds.

Proposition 2.2.1

$\Gamma_{\rho\nu}^\mu$ are not components of a tensor.

Proof. Given the basis transformation $\tilde{e}_\nu = A^\mu_\nu e_\mu$, with A and invertible matrix (if they are both coordinate basis, then $A^\mu_\nu = \frac{\partial x^\mu}{\partial y^\nu}$), the components of a (1,2) tensor must transform as:

$$\tilde{T}_{\rho\nu}^\mu = (A^{-1})^\mu_\tau A^\sigma_\rho A^\lambda_\nu T_{\sigma\lambda}^\tau$$

In the new basis:

$$\begin{aligned}\tilde{\Gamma}_{\rho\nu}^\mu \tilde{e}_\mu &= \nabla_{\tilde{e}_\rho} \tilde{e}_\nu = \nabla_{A^\sigma_\rho e_\sigma} (A^\lambda_\nu e_\nu) = A^\sigma_\rho \nabla_{e_\sigma} (A^\lambda_\nu e_\nu) \\ &= A^\sigma_\rho A^\lambda_\nu \Gamma_{\sigma\lambda}^\tau e_\tau + A^\sigma_\rho e_\lambda \partial_\sigma A^\lambda_\nu = [A^\sigma_\rho A^\lambda_\nu \Gamma_{\sigma\lambda}^\tau + A^\sigma_\rho \partial_\sigma A^\lambda_\nu] e_\tau \\ &= [A^\sigma_\rho A^\lambda_\nu \Gamma_{\sigma\lambda}^\tau + A^\sigma_\rho \partial_\sigma A^\lambda_\nu] (A^{-1})^\mu_\tau \tilde{e}_\mu\end{aligned}$$

Thus, there is a second term proportional to ∂A which deviates from the transformation law:

$$\tilde{\Gamma}_{\rho\nu}^\mu = (A^{-1})^\mu_\tau A^\sigma_\rho [A^\lambda_\nu \Gamma_{\sigma\lambda}^\tau + \partial_\sigma A^\lambda_\nu]$$

which means that $\Gamma_{\rho\nu}^\mu$ is not a tensor. \square

§2.2.2 Covariant derivative of tensors

First of all, it is necessary to elucidate how the covariant derivative acts on 1-forms. Given a 1-form ω , the 1-form $\nabla_X \omega$ is defined by its action on vector fields. By Leibniz rule:

$$\nabla_X(\omega(Y)) = (\nabla_X \omega)(Y) + \omega(\nabla_X Y)$$

Recalling that $\omega(Y)$ is a function, $\nabla_X(\omega(Y)) = X(\omega(Y))$, therefore:

$$(\nabla_X \omega)(Y) = X(\omega(Y)) - \omega(\nabla_X Y) \quad (2.2.4)$$

Expressing it in coordinates:

$$\begin{aligned}X^\mu (\nabla_\mu \omega)_\nu Y^\nu &= X^\mu \partial_\mu (\omega_\nu Y^\nu) - \omega_\nu X^\mu [\partial_\mu Y^\nu + \Gamma_{\mu\rho}^\nu Y^\rho] \\ &= X^\mu [\partial_\mu \omega_\rho - \Gamma_{\mu\rho}^\nu \omega_\nu] Y^\rho\end{aligned}$$

Crucially, the ∂Y terms cancel out, allowing to define $\nabla_X \omega$ without referencing Y :

$$(\nabla_\mu \omega)_\rho = \partial_\mu \omega_\rho - \Gamma_{\mu\rho}^\nu \omega_\nu \quad (2.2.5)$$

Using the same notation as for vector fields $(\nabla_\mu \omega)_\rho \equiv \nabla_\mu \omega_\rho \equiv \omega_{\rho;\mu}$:

$$\omega_{\rho;\mu} = \omega_{\rho,\mu} - \Gamma_{\mu\rho}^\nu \omega_\nu \quad (2.2.6)$$

This kind of argument can be extended to a general (p, q) tensor field:

$$\begin{aligned}\nabla_\rho T^{\mu_1 \dots \mu_p}_{\nu_1 \dots \nu_q} &= \partial_\rho T^{\mu_1 \dots \mu_p}_{\nu_1 \dots \nu_q} + \Gamma_{\rho\sigma}^{\mu_1} T^{\sigma \mu_2 \dots \mu_p}_{\nu_1 \dots \nu_q} + \dots + \Gamma_{\rho\sigma}^{\mu_p} T^{\mu_1 \dots \mu_{p-1} \sigma}_{\nu_1 \dots \nu_q} \\ &\quad - \Gamma_{\rho\nu_1}^\sigma T^{\mu_1 \dots \mu_p}_{\sigma \nu_2 \dots \nu_q} - \dots - \Gamma_{\rho\nu_q}^\sigma T^{\mu_1 \dots \mu_p}_{\nu_1 \dots \nu_{q-1} \sigma}\end{aligned} \quad (2.2.7)$$

The pattern is clear: for each upper index μ there is a $+\Gamma_{\rho\sigma}^\mu T^\sigma$ term, while for each lower index ν there is $-\Gamma_{\rho\nu}^\sigma T_\sigma$ term. Furthermore, it is necessary to generalize the comma-notation: for example, $X^\mu_{;\nu\rho} \equiv \nabla_\rho \nabla_\nu X^\mu$, so the rightmost index is the one acting first.

§2.2.2.1 Torsion and curvature

Even though the connection is not a tensor, it is used to construct two important tensors.

Definition 2.2.2 (Torsion tensor)

The **torsion** is a $(1,2)$ tensor defined on $\Lambda^1(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M})$ as:

$$\mathsf{T}(\omega, X, Y) := \omega(\nabla_X Y - \nabla_Y X - [X, Y]) \quad (2.2.8)$$

Alternatively, the torsion can be viewed as a map $\mathsf{T} : \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \rightarrow \mathfrak{X}(\mathcal{M})$ such that:

$$\mathsf{T}(X, Y) = \nabla_X Y - \nabla_Y X - [X, Y] \quad (2.2.9)$$

Definition 2.2.3 (Curvature tensor)

The **curvature** is a $(1,3)$ tensor defined on $\Lambda^1(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M})$ as:

$$\mathsf{R}(\omega, X, Y, Z) := \omega(\nabla_X \nabla_Y Z - \nabla_Y \nabla_X Z - \nabla_{[X,Y]} Z) \quad (2.2.10)$$

Alternatively, the curvature can be viewed as a map from $\mathfrak{X}(\mathcal{M}) \times \mathfrak{X}(\mathcal{M})$ to the space of differential operators on $\mathfrak{X}(\mathcal{M})$ such that:

$$\mathsf{R}(X, Y) = \nabla_X \nabla_Y - \nabla_Y \nabla_X - \nabla_{[X,Y]} \quad (2.2.11)$$

The fact that these are indeed tensors, i.e. they are linear in each argument, can be shown by direct calculation, recalling that $[fX, Y] = f[X, Y] - Y(f)X$.

Proposition 2.2.2

On the coordinate basis $\{\partial_\mu\}$ and $\{dx^\mu\}$ the torsion components are:

$$T_{\mu\nu}^\rho = \Gamma_{\mu\nu}^\rho - \Gamma_{\nu\mu}^\rho \quad (2.2.12)$$

Proof. By direct calculation:

$$\begin{aligned} T_{\mu\nu}^\rho &= T(dx^\rho, \partial_\mu, \partial_\nu) = dx^\mu(\nabla_\mu \partial_\nu - \nabla_\nu \partial_\mu - [\partial_\mu, \partial_\nu]) \\ &= dx^\mu(\partial_\mu \partial_\nu - \Gamma_{\mu\nu}^\sigma \partial_\sigma - \partial_\nu \partial_\mu + \Gamma_{\nu\mu}^\sigma \partial_\sigma) \\ &= [\Gamma_{\mu\nu}^\sigma - \Gamma_{\nu\mu}^\sigma] \delta_\sigma^\mu = \Gamma_{\mu\nu}^\rho - \Gamma_{\nu\mu}^\rho \end{aligned}$$

which is the thesis. \square

Interestingly, even though $\Gamma_{\mu\nu}^\rho$ is not a tensor, its anti-symmetric part $\Gamma_{[\mu\nu]}^\rho = \frac{1}{2}T_{\mu\nu}^\sigma$ is. Clearly, the torsion tensor is anti-symmetric in its lower indices, thus for connections which are symmetric in their lower indices the torsion is null: such connections are said to be **torsion-free**.

Proposition 2.2.3

On the coordinate basis $\{\partial_\mu\}$ and $\{dx^\mu\}$ the curvature components are:

$$R_{\mu\nu\rho}^\sigma = \partial_\mu \Gamma_{\nu\rho}^\sigma - \partial_\nu \Gamma_{\mu\rho}^\sigma + \Gamma_{\nu\rho}^\lambda \Gamma_{\mu\lambda}^\sigma - \Gamma_{\mu\rho}^\lambda \Gamma_{\nu\lambda}^\sigma \quad (2.2.13)$$

Proof. By direct calculation:

$$\begin{aligned}
 R(dx^\sigma, \partial_\mu, \partial_\nu, \partial_\rho) &= dx^\sigma (\nabla_\mu \nabla_\nu \partial_\rho - \nabla_\nu \nabla_\mu \partial_\rho - \nabla_{[\partial_\mu, \partial_\nu]} \partial_\rho) \\
 &= dx^\sigma (\nabla_\mu \nabla_\nu \partial_\rho - \nabla_\nu \nabla_\mu \partial_\rho) \\
 &= dx^\sigma (\nabla_\mu (\Gamma_{\nu\rho}^\lambda \partial_\lambda) - \nabla_\nu (\Gamma_{\mu\rho}^\lambda \partial_\lambda)) \\
 &= dx^\sigma ((\partial_\mu \Gamma_{\nu\rho}^\lambda) \partial_\lambda + \Gamma_{\nu\rho}^\lambda \Gamma_{\mu\lambda}^\tau \partial_\tau - (\partial_\nu \Gamma_{\mu\rho}^\lambda) \partial_\lambda - \Gamma_{\mu\rho}^\lambda \Gamma_{\nu\lambda}^\tau \partial_\tau) \\
 &= \partial_\mu \Gamma_{\nu\rho}^\sigma - \partial_\nu \Gamma_{\mu\rho}^\sigma + \Gamma_{\nu\rho}^\lambda \Gamma_{\mu\lambda}^\sigma - \Gamma_{\mu\rho}^\lambda \Gamma_{\nu\lambda}^\sigma
 \end{aligned}$$

which completes the proof. \square

Clearly, the curvature tensor is anti-symmetric in its last two lower indices, i.e. $R^\sigma_{\rho\mu\nu} = R^\sigma_{\rho[\mu\nu]}$, and it can in fact be written as:

$$R^\sigma_{\rho\mu\nu} = 2\partial_{[\mu}\Gamma_{\nu]\rho}^\sigma + 2\Gamma_{[\mu|\lambda|}^\sigma \Gamma_{\nu]\rho}^\lambda \quad (2.2.14)$$

Theorem 2.2.1 (Ricci identity)

$$2\nabla_{[\mu} \nabla_{\nu]} Z^\sigma = R^\sigma_{\rho\mu\nu} Z^\rho - T^\rho_{\mu\nu} \nabla_\rho Z^\sigma \quad (2.2.15)$$

Proof. By direct calculation:

$$\begin{aligned}
 \nabla_{[\mu} \nabla_{\nu]} Z^\sigma &= \partial_{[\mu} (\nabla_{\nu]} Z^\sigma) + \Gamma_{[\mu|\lambda|}^\sigma \nabla_{\nu]} Z^\lambda - \Gamma_{[\mu\nu]}^\rho \nabla_\rho Z^\sigma \\
 &= \partial_{[\mu} \partial_{\nu]} Z^\sigma + (\partial_{[\mu} \Gamma_{\nu]\rho}^\sigma) Z^\rho + (\partial_{[\mu} Z^\rho) \Gamma_{\nu]\rho}^\sigma + \Gamma_{[\mu|\lambda|}^\sigma \partial_{\nu]} Z^\lambda + \Gamma_{[\mu|\lambda|}^\sigma \Gamma_{\nu]\rho}^\lambda Z^\rho - \frac{1}{2} T^\rho_{\mu\nu} \nabla_\rho Z^\sigma \\
 &= (\partial_{[\mu} \Gamma_{\nu]\rho}^\sigma + \Gamma_{[\mu|\lambda|}^\sigma \Gamma_{\nu]\rho}^\lambda) Z^\rho - \frac{1}{2} T^\rho_{\mu\nu} \nabla_\rho Z^\sigma \\
 &= \frac{1}{2} R^\sigma_{\rho\mu\nu} Z^\rho - \frac{1}{2} T^\sigma_{\mu\nu} \nabla_\rho Z^\sigma
 \end{aligned}$$

which is the thesis. \square

§2.2.2.2 Levi–Civita connection

The discussion on the connection has so far been independent of the metric. Starting to consider it, an important result is the fundamental theorem of Riemannian geometry.

Theorem 2.2.2 (Riemann's theorem)

On a metric manifold (\mathcal{M}, g) , there exists a unique torsion-free connection that is compatible with the metric, i.e. for all $X \in \mathfrak{X}(\mathcal{M})$:

$$\nabla_X g = 0 \quad (2.2.16)$$

This is called the **Levi–Civita connection**.

Proof. WTS uniqueness: suppose such a connection exists. Then, by Leibniz rule:

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

Since $\nabla_X g = 0$, by cyclic permutations of X , Y and Z :

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

Since the connection is torsion-free, $\nabla_X Y - \nabla_Y X = [X, Y]$, thus these equations become:

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

Adding the first two and subtracting the third:

$$\begin{aligned} g(\nabla_Y X, Z) &= \frac{1}{2}[X(g(Y, Z)) + Y(g(Z, X)) + Z(g(X, Y)) \\ &\quad - g([X, Y], Z) - g([Y, Z], X) + g([Z, X], Y)] \end{aligned}$$

The metric is non-degenerate, thus this uniquely specifies the connection. By direct calculation it can be shown that it indeed satisfies all the properties of a connection. \square

It is possible to explicit the components of the Levi–Civita connection in terms of the metric.

Proposition 2.2.4 (Christoffel symbols)

On the coordinate basis $\{\partial_\mu\}$ and $\{dx^\mu\}$ the Levi–Civita connection's components, called **Christoffel symbols**, are:

$$\Gamma_{\mu\nu}^\lambda = \frac{1}{2}g^{\lambda\rho}(\partial_\mu g_{\nu\rho} + \partial_\nu g_{\mu\rho} - \partial_\rho g_{\mu\nu}) \quad (2.2.17)$$

Proof. Recalling that $[\partial_\mu, \partial_\nu] = 0$, $\Gamma_{\mu\nu}^\lambda g_{\lambda\rho} = g(\nabla_\mu \partial_\nu, \partial_\rho) = \frac{1}{2}(\partial_\nu g_{\mu\rho} + \partial_\mu g_{\nu\rho} - \partial_\rho g_{\mu\nu})$. \square

Example 2.2.1 (Euclidean space)

In flat space \mathbb{R}^n , endowed with either Euclidean or Minkowski metric, it is always possible to choose Cartesian coordinates, in which case the Christoffel symbols vanish. Being the Riemann tensor a genuine tensor, it therefore will vanish in all possible coordinate systems on \mathbb{R}^n , even in those with $\Gamma_{\mu\nu}^\rho \neq 0$: this expresses the flatness of \mathbb{R}^n .

§2.2.2.3 Gauss' theorem

The divergence theorem (or Gauss' theorem) states that the integral of a total derivative is a boundary term. It is possible to express this theorem on curved manifolds in a convenient way.

Lemma 2.2.1

$$\Gamma_{\mu\nu}^\mu = \frac{1}{\sqrt{g}}\partial_\nu \sqrt{g} \quad (2.2.18)$$

Proof. A useful identity for invertible matrices: $\text{tr} \log A = \log \det A$. Thus (WLOG $\det g > 0$):

$$\Gamma_{\mu\nu}^\mu = \frac{1}{2}g^{\mu\rho}\partial_\nu g_{\mu\rho} = \frac{1}{2}\text{tr}(g^{-1}\partial_\nu g) = \frac{1}{2}\text{tr}(\partial_\nu \log g) = \frac{1}{2}\partial_\nu \log \det g = \frac{1}{\sqrt{\det g}}\partial_\nu \sqrt{\det g}$$

which is the thesis. \square

Theorem 2.2.3 (Gauss' theorem)

Given a Riemannian manifold (\mathcal{M}, g) , consider a region $M \subseteq \mathcal{M}$ with boundary ∂M and let n^μ be an outward-pointing unit vector orthogonal to ∂M . Then, for any vector field X^μ on M :

$$\int_M d^n x \sqrt{g} \nabla_\mu X^\mu = \int_{\partial M} d^{n-1} x \sqrt{\gamma} n_\mu X^\mu \quad (2.2.19)$$

where γ_{ij} is the pull-back of the metric to ∂M and $\gamma \equiv \det \gamma_{ij}$.

Proof. From Lemma 2.2.1:

$$\sqrt{g} \nabla_\mu X^\mu = \sqrt{g} (\partial_\mu X^\mu + \Gamma_{\mu\nu}^\mu X^\nu) = \sqrt{g} \left(\partial_\mu X^\mu + X^\nu \frac{1}{\sqrt{g}} \partial_\nu \sqrt{g} \right) = \partial_\mu (\sqrt{g} X^\mu)$$

The integral becomes:

$$\int_M d^n x \sqrt{g} \nabla_\mu X^\mu = \int_M d^n x \partial_\mu (\sqrt{g} X^\mu)$$

This is the integral of an ordinary partial derivative, so the ordinary divergence theorem applies. To evaluate the integral on the boundary, it is convenient to pick coordinates so that ∂M is a surface at constant x^n . Moreover, to simplify the proof, the possible metrics will be restricted to $g_{\mu\nu} = \text{diag}(\gamma_{ij}, N^2)$. By usual integration rules:

$$\int_M d^n x \partial_\mu (\sqrt{g} X^\mu) = \int_{\partial M} d^{n-1} x \sqrt{\gamma N^2} X^n$$

The unit normal vector is $n^\mu = (0, \dots, 0, \frac{1}{N})$, so that $g_{\mu\nu} n^\mu n^\nu = 1$, therefore $n_\mu = g_{\mu\nu} n^\nu = (0, \dots, 0, N)$. The proof is then concluded because:

$$\int_{\partial M} d^{n-1} x \sqrt{\gamma N^2} X^n = \int_{\partial M} d^{n-1} x \sqrt{\gamma} n_\mu X^\mu$$

which is the thesis. \square

Note that this theorem holds on Lorentzian manifolds too, with the condition that ∂M must be purely timelike or purely spacelike, ensuring that $\gamma \neq 0$ at any point.

§2.2.2.4 Maxwell action

Consider spacetime as a manifold \mathcal{M} . The electromagnetic field can be described by a form on this manifold: indeed, the electromagnetic gauge field $A_\mu = (\phi, \mathbf{A})$ is to be thought as the

components of a 1-form $A = A_\mu(x)dx^\mu$. The exterior derivative of this form is a 2-form $F = dA$:

$$F = \frac{1}{2}F_{\mu\nu}dx^\mu \wedge dx^\nu = \frac{1}{2}(\partial_\mu A_\nu - \partial_\nu A_\mu)dx^\mu \wedge dx^\nu$$

The components $F_{\mu\nu}$ are in reality the components of a tensor, the Faraday tensor. By construction, a useful identity holds, sometimes called the **Bianchi identity**:

$$dF = 0 \quad (2.2.20)$$

From this identity derive two Maxwell equations: $\nabla \cdot \mathbf{B} = 0$ and $\nabla \times \mathbf{E} + \partial_t \mathbf{B} = 0$. Moreover, note that the gauge field is not unique: the gauge transformation $A \mapsto A + d\alpha$, which equals $A_\mu \mapsto A_\mu + \partial_\mu \alpha$, leaves F unchanged.

To study the dynamics of these fields, an action is needed: Differential Geometry allows very few actions to be written down.

For example, suppose that on the considered manifold no metric is defined. To integrate over \mathcal{M} a 4-form is needed, but F is a 2-form, thus the only possible action is:

$$\mathcal{S}_{\text{top}} = -\frac{1}{2} \int F \wedge F \quad (2.2.21)$$

The integrand become $dx^0 dx^1 dx^2 dx^3 \mathbf{E} \cdot \mathbf{B}$. Actions of this kind, independent of the metric, are called **topological actions** and are of no interest in classical physics: in fact, $F \wedge F = d(A \wedge F)$, so the action is a total derivative and doesn't affect the equations of motion.

To construct an action of classical interest, a metric is needed. This allows to introduce a second 2-form, $\star F$, so to construct the Maxwell action:

$$\mathcal{S}_M = -\frac{1}{2} \int F \wedge \star F \quad (2.2.22)$$

The integrand can then be expanded as:

$$\mathcal{S}_M = -\frac{1}{4} \int d^4x \sqrt{g} g^{\mu\nu} g^{\rho\sigma} F_{\mu\rho} F_{\nu\sigma} = -\frac{1}{4} \int d^4x \sqrt{g} F^{\mu\nu} F_{\mu\nu}$$

In flat spacetime $F^{\mu\nu} F_{\mu\nu} = 2(\mathbf{B}^2 - \mathbf{E}^2)$. In a general curved spacetime, the equation of motion resulting from the variation of the Maxwell action is $d \star F = 0$.

To complete the theory, consider a gauge field coupled to a current, described by a 1-form J . The Maxwell action then becomes:

$$\mathcal{S}_M = \int -\frac{1}{2}F \wedge \star F + A \wedge \star J \quad (2.2.23)$$

This action must retain its gauge invariance, but under $A \mapsto A + d\alpha$ it transforms as $\mathcal{S}_M \mapsto \mathcal{S}_M + \int d\alpha \wedge \star J$, therefore, after integrating by parts, the condition of gauge invariance translates to:

$$d \star J = 0 \quad (2.2.24)$$

This is current conservation in the language of forms. Varying the action in Eq. 2.2.22 now leads to the Maxwell equations with source terms:

$$d \star F = \star J \quad (2.2.25)$$

To define electric and magnetic charges, integrate over submanifolds. Consider a 3-dimensional spatial submanifold Σ : the electric charge in Σ is defined as:

$$Q_e(\Sigma) := \int_{\Sigma} \star J \quad (2.2.26)$$

This agrees with the usual definition in flat spacetime $Q_e = \int_{\Sigma} d^3x J^0$. Using the equations of motion and Stokes' theorem, a general form of Gauss' law is obtained:

$$Q_e(\Sigma) = \int_{\partial\Sigma} \star F \quad (2.2.27)$$

Similarly, the magnetic charge in Σ is defined as:

$$Q_m(\Sigma) := \int_{\partial\Sigma} F \quad (2.2.28)$$

The non-existence of magnetic charges, following from Bianchi identity, can be evaded in topologically interesting manifolds.

From charge conservation in Eq. 2.2.24, it follows that the electric charge in a region cannot change, unless current flows in or out of that region. Consider a cylindrical region of spacetime V , ending in two spatial hypersurfaces Σ_1 and Σ_2 : its boundary is $\partial V = \Sigma_1 \cup \Sigma_2 \cup B$, where B is a cylindrical timelike hypersurface. The statement that no current flows in or out of V means that $J|_B = 0$. Then:

$$Q_e(\Sigma_1) - Q_e(\Sigma_2) = \int_{\Sigma_1} \star J - \int_{\Sigma_2} \star J = \int_{\partial V} \star J - \int_B \star J = \int_{\partial V} \star J = \int_V d \star J = 0$$

Thus, electric charge in remains constant in time.

Maxwell equations from connections First note that, given the gauge field $A \in \Lambda^1(\mathcal{M})$, the field strength can be expressed via covariant derivatives:

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu = \nabla_\mu A_\nu - \nabla_\nu A_\mu$$

The Christoffel symbols cancel out due to anti-symmetry: this is what allows to define the exterior derivative without introducing connections first.

Proposition 2.2.5 (Current conservation)

Current conservation can be written as: $d \star J = 0 \iff \nabla_\mu J^\mu = 0$.

Proof. Recalling Lemma 2.2.1:

$$\nabla_\mu J^\mu = \partial_\mu J^\mu + \Gamma_{\mu\rho}^\mu J^\rho = \partial_\mu J^\mu + \partial_\rho (\log \sqrt{g}) J^\rho = \frac{1}{\sqrt{g}} \partial_\mu (\sqrt{g} J^\mu) \propto d \star J$$

Hence, $d \star J = 0 \iff \nabla_\mu J^\mu = 0$. □

As an aside, in general the divergence in different coordinate systems can be computed using the formula $\nabla_\mu J^\mu = \frac{1}{\sqrt{g}} \partial_\mu (\sqrt{g} J^\mu)$.

Proposition 2.2.6

$$d \star F = \star J \iff \nabla_\mu F^{\mu\nu} = J^\nu \quad (2.2.29)$$

Proof. Recalling Lemma 2.2.1:

$$\nabla_\mu F^{\mu\nu} = \partial_\mu F^{\mu\nu} + \Gamma_{\mu\rho}^\mu F^{\rho\nu} + \Gamma_{\mu\rho}^\nu F^{\mu\rho} = \frac{1}{\sqrt{g}} \partial_\mu (\sqrt{g} F^{\mu\nu})$$

where $\Gamma_{\mu\rho}^\nu F^{\mu\rho} = 0$ because $\Gamma_{\mu\rho}^\nu$ is symmetric in μ and ρ , while $F^{\mu\rho}$ is anti-symmetric. The proof follows recalling the definition of the Hodge dual in Eq. 2.1.18. \square

§2.3 Parallel transport

The connection connects tangent spaces, or more generally any tensor vector space, at different points of the manifold: this map is called parallel transport and it's necessary for the definition of differentiation.

Definition 2.3.1 (Parallel transport)

Consider a vector field X and some associated integral curve γ , with coordinates $x^\mu(\tau)$ such that:

$$X^\mu|_\gamma = \frac{dx^\mu(\tau)}{d\tau}$$

A tensor field T is said to be **parallelly transported** along γ if:

$$\nabla_X T = 0 \quad (2.3.1)$$

Suppose that γ connects two points $p, q \in \mathcal{M}$: Eq. 2.3.1 provides a map from the tensor vector space defined at p to that defined at q . To illustrate this, consider the parallel transport of a vector field Y :

$$X^\nu(\partial_\nu Y^\mu + \Gamma_{\nu\rho}^\mu Y^\rho) = 0$$

Evaluating this equation on γ , considering $Y^\mu = Y^\mu(x(\tau))$:

$$\frac{dY^\mu}{d\tau} + X^\nu \Gamma_{\nu\rho}^\mu Y^\rho = 0 \quad (2.3.2)$$

These are a set of coupled ODEs, thus, given an initial condition (e.g. at $\tau = 0$, i.e. at p), these equations can be solved to find a unique vector at each point along the curve.

Note that the parallel transport depends both on the path (characterized by the vector field X) and on the connection.

Definition 2.3.2 (Geodesic)

Given a vector field X , a **geodesic** is a curve tangent to X such that:

$$\nabla_X X = 0 \quad (2.3.3)$$

Proposition 2.3.1 (Geodesic equation)

A geodesic is described by:

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma_{\nu\rho}^\mu \frac{dx^\nu}{d\tau} \frac{dx^\rho}{d\tau} = 0 \quad (2.3.4)$$

Proof. From the above calculations, along γ :

$$0 = \frac{dX^\mu}{d\tau} + \Gamma_{\nu\rho}^\mu X^\nu X^\rho = \frac{d^2x^\mu}{d\tau^2} + \Gamma_{\nu\rho}^\mu \frac{dx^\nu}{d\tau} \frac{dx^\rho}{d\tau}$$

which is the geodesic equation. □

For the Levi-Civita connection $\nabla_X g = 0$. If $Y \in \mathfrak{X}(\mathcal{M})$ is parallelly transported along a geodesic associated to $X \in \mathfrak{X}(\mathcal{M})$, then $\nabla_X Y = \nabla_X X = 0$, therefore $\frac{d}{d\tau}g(X, Y) = 0$: this ensures that the two tangent vectors always determine the same angles along the geodesic.

§2.3.1 Normal coordinates

Theorem 2.3.1 (Equivalence principle)

Given a Riemannian manifold (\mathcal{M}, g) and $p \in \mathcal{M}$, in a neighbourhood of p it's always possible to find coordinates, called **normal coordinates**, such that $g_{\mu\nu}(p) = \delta_{\mu\nu}$ and $g_{\mu\nu,\rho}(p) = 0$.

Proof. By brute force, consider initial coordinates y^μ and find a change of coordinates x^μ which satisfy the requirements:

$$\frac{\partial y^\rho}{\partial x^\mu} \frac{\partial y^\sigma}{\partial x^\nu} \tilde{g}_{\rho\sigma} = g_{\mu\nu}$$

WLOG p is the origin of both sets of coordinates, so:

$$y^\rho = \frac{\partial y^\rho}{\partial x^\mu} \Big|_{x=0} x^\mu + \frac{1}{2} \frac{\partial^2 y^\rho}{\partial x^\mu \partial x^\nu} \Big|_{x=0} x^\mu x^\nu + \dots$$

This, together with the Taylor expansion of $\tilde{g}_{\rho\sigma}$, can be inserted in the transformation equation of the metric, thus finding a set of PDEs for each power of x , which can be solved to characterize the normal coordinates. For example, the first condition is:

$$\frac{\partial y^\rho}{\partial x^\mu} \Big|_{x=0} \frac{\partial y^\sigma}{\partial x^\nu} \Big|_{x=0} \tilde{g}_{\rho\sigma}(p) = \delta_{\mu\nu}$$

Given any $\tilde{g}_{\rho\sigma}(p)$, it's always possible to find $\frac{\partial y}{\partial x}$ that satisfies this condition. In fact, if $\dim_{\mathbb{R}} \mathcal{M} = n$, the Jacobian of the transformation has n^2 independent elements and the equation above puts $\frac{1}{2}n(n+1)$ constraints: the remaining free parameters are $\frac{1}{2}n(n-1)$, which is precisely the dimension of $\text{SO}(n)$, the symmetry group of the flat metric. A similar counting shows that $g_{\mu\nu,\rho}(p) = 0$ puts $\frac{1}{2}n^2(n+1)$ constraints, precisely the number of independent elements of the Hessian of the transformation. \square

This theorem holds for Lorentzian manifolds too, but the flat metric is now $\eta_{\mu\nu}$ and its symmetry group is $\text{SO}(1, n-1)$. The condition $g_{\mu\nu,\rho}(p) = 0$ implies that $\Gamma_{\nu\rho}^\mu(p) = 0$, but generally the Christoffel symbols won't vanish away from p . Note, however, that it's not generally possible to ensure the vanishing of second derivatives too: indeed, $g_{\mu\nu,\rho\sigma}(p) = 0$ would put $\frac{1}{4}n^2(n+1)^2$ constraints, but the independent $\frac{\partial^3 y}{\partial^3 x}$ terms are $\frac{1}{6}n^2(n+1)(n+2)$, thus leaving $\frac{1}{12}n^2(n^2-1)$ free terms: this is precisely the number of independent components of the Riemann tensor, therefore in general it's not possible to pick coordinates as to make the Riemann tensor vanish too.

§2.3.1.1 Exponential map

A simple way to construct normal coordinates is the following: given a tangent vector $X_p \in T_p \mathcal{M}$, there is a unique affinely parametrized geodesic through p with tangent vector X_p at p ; then, any point q in the neighbourhood of p is labelled by the coordinates of the geodesic that takes from p to q a fixed amount of time.

Analytically, introducing a coordinate system (not necessarily normal) \tilde{x}^μ in the neighbourhood of p , an affinely parametrized geodesic solves Eq. 2.3.4, with initial conditions $\frac{\partial \tilde{x}^\mu}{\partial \tau}|_{\tau=0} = \tilde{X}_p^\mu$ and $\tilde{x}^\mu(\tau=0) = 0$ that make the solution unique. The uniqueness of the solution allows to define a map $\text{Exp} : T_p \mathcal{M} \rightarrow \mathcal{M}$, called the **exponential map**, which acts as follows: given $X_p \in T_p \mathcal{M}$, construct the appropriate geodesic as above and follow it for a fixed affine distance,

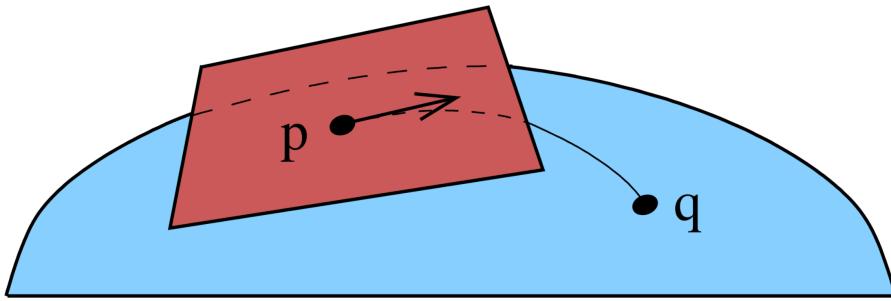


Figure 2.1: Visualization of the exponential map.

conventionally $\tau = 1$, to get a new point $q \in \mathcal{M}$. See Fig. Fig. 2.1 for visual aid. Obviously, there may be points which cannot be reached from p by geodesics, or there may be tangent vectors X_p for which Exp is ill-defined: in General Relativity, this occurs when spacetime has singularities, but these are not relevant issues.

Pick a basis $\{e_\mu\}$ of $T_p\mathcal{M}$. Then $\text{Exp} : T_p\mathcal{M} \ni X^\mu e_\mu \mapsto q \in \mathcal{M}$, thus it is possible to assign coordinates in the neighbourhood of p such that $x^\mu(q) = X^\mu$: these are the normal coordinates. To show this, note that if $\{e_\mu\}$ is orthonormal, then the geodesics will point in orthogonal directions, ensuring that $g_{\mu\nu}(p) = \delta_{\mu\nu}$. Now, fix a point q associated to a given tangent vector $X_p \in T_p\mathcal{M}$: this means that q is at distance $\tau = 1$ from p along the given geodesic. Note that the geodesic equation is homogeneous in τ , thus in general $\text{Exp} : \tau X_p \mapsto x^\mu(\tau) = \tau X^\mu$, which means that geodesics take a simple form in these coordinates:

$$x^\mu(\tau) = \tau X^\mu$$

Being these geodesics, they must solve Eq. 2.3.4, that is:

$$\Gamma_{\nu\rho}^\mu(x(\tau))X^\nu X^\rho = 0$$

which holds at any point along the geodesic, i.e. at any $\tau \in \mathbb{R}^+$. At most points $x(\tau)$, this equation only holds for those choices of X^μ tangent to the geodesics. However, at $x(0) = p$, i.e. at p , it must hold for any tangent vector: this means that $\Gamma_{(\nu\rho)}^\mu(p) = 0$ which, for a torsion-free connection, ensures that $\Gamma_{\nu\rho}^\mu(p) = 0$. But vanishing Christoffel symbols imply a vanishing first derivative of the metric: for the Levi–Civita connection $2g_{\mu\sigma}\Gamma_{\nu\rho}^\sigma = g_{\mu\nu,\rho} + g_{\mu\rho,\nu} - g_{\nu\rho,\mu}$, thus symmetrizing $(\mu\nu)$ cancels the last two terms, leaving an identity that, evaluated at p , gives $g_{\mu\nu,\rho}(p) = 0$. Hence, these are indeed normal coordinates.

§2.3.1.2 Equivalence principle

Normal coordinates are conceptually important in General Relativity: an observer at point p who parametrizes their immediate surroundings using coordinates constructed by geodesics will experience a locally flat metric. This is precisely Einstein’s equivalence principle: any free-falling observer, performing local experiments, will not experience a gravitational field. The formal definition of free-falling observer is an observer which follows geodesics, while the local lack of gravitational field means $g_{\mu\mu}(p) = \eta_{\mu\nu}$. In this context, normal coordinates are called **local inertial frame**.

To understand what “local” means, note that there is a way to distinguish whether a gravitational field is present at p : a non-vanishing Riemann tensor. This depends on the second derivatives of the metric, which in general will be non-vanishing. However, to measure the effects of the Riemann tensor, one needs to compare the results of experiments at p and at a nearby point q : this is a non-local observation.

§2.3.2 Curvature and torsion

With reference to Fig. 2.2, consider a tangent vector $Z_p \in T_p\mathcal{M}$ and two vector fields $X, Y \in \mathfrak{X}(\mathcal{M}) : [X, Y] = 0$, i.e. they are linearly independent. Construct two curves γ, γ' as in figure, both leading to a point $r \in \mathcal{M}$ which, for simplicity, is close to p . It is possible to impose normal coordinates centered at p such that $x^\mu = (\tau, \sigma, \dots)$, so that $X = \frac{\partial}{\partial \tau}$ and $Y = \frac{\partial}{\partial \sigma}$: then $x^\mu(p) = (0, 0, 0, \dots)$, $X^\mu(q) = (\delta\tau, 0, 0, \dots)$, $x^\mu(s) = (0, \delta\sigma, 0, \dots)$ and $x^\mu(r) = (\delta\tau, \delta\sigma, 0, \dots)$, with $\delta\tau, \delta\sigma$ small.

First, parallel transport Z_p along X to Z_q , so that Z^μ solves:

$$\frac{dZ^\mu}{d\tau} + X^\nu \Gamma_{\nu\rho}^\mu Z^\rho = 0$$

In normal coordinates $\Gamma_{\nu\rho}^\mu(p) = 0$, thus $\frac{dZ^\mu}{d\tau}|_{\tau=0} = 0$ and the Taylor expansion is:

$$\begin{aligned} Z_q^\mu &= Z_p^\mu + \frac{\delta\tau^2}{2} \frac{d^2 Z^\mu}{d\tau^2} \Big|_{\tau=0} + o(\delta\tau^3) \\ &= Z_p^\mu - \frac{\delta\tau^2}{s} \left[X^\nu Z^\rho \frac{d\Gamma_{\nu\rho}^\mu}{d\tau} + \frac{dX^\nu}{d\tau} Z^\rho \Gamma_{\nu\rho}^\mu + X^\nu \frac{dZ^\rho}{d\tau} \Gamma_{\nu\rho}^\mu \right]_{\tau=0} + o(\delta\tau^3) \\ &= Z_p^\mu - \frac{\delta\tau^2}{2} X^\nu Z^\rho \frac{d\Gamma_{\nu\rho}^\mu}{d\tau} \Big|_{\tau=0} + o(\delta\tau^3) = Z_p^\mu - \frac{\delta\tau^2}{2} [X^\nu X^\sigma Z^\rho \Gamma_{\nu\rho,\sigma}^\mu]_p + o(\delta\tau^3) \end{aligned}$$

where $\frac{d}{d\tau} = X^\sigma \partial_\sigma$. Now, Z_q needs to be parallelly transported along Y to Z_r , but this time $\frac{dZ^\mu}{d\sigma}|_{\sigma=0}$ doesn't vanish, in general. From the parallel transport equation:

$$\frac{dZ^\mu}{d\sigma} \Big|_{\sigma=0} = - [Y^\nu Z^\rho \Gamma_{\nu\rho}^\mu]_q = - [Y^\nu Z^\rho X^\sigma \Gamma_{\nu\rho,\sigma}^\mu]_p \delta\tau + o(\delta\tau^2)$$

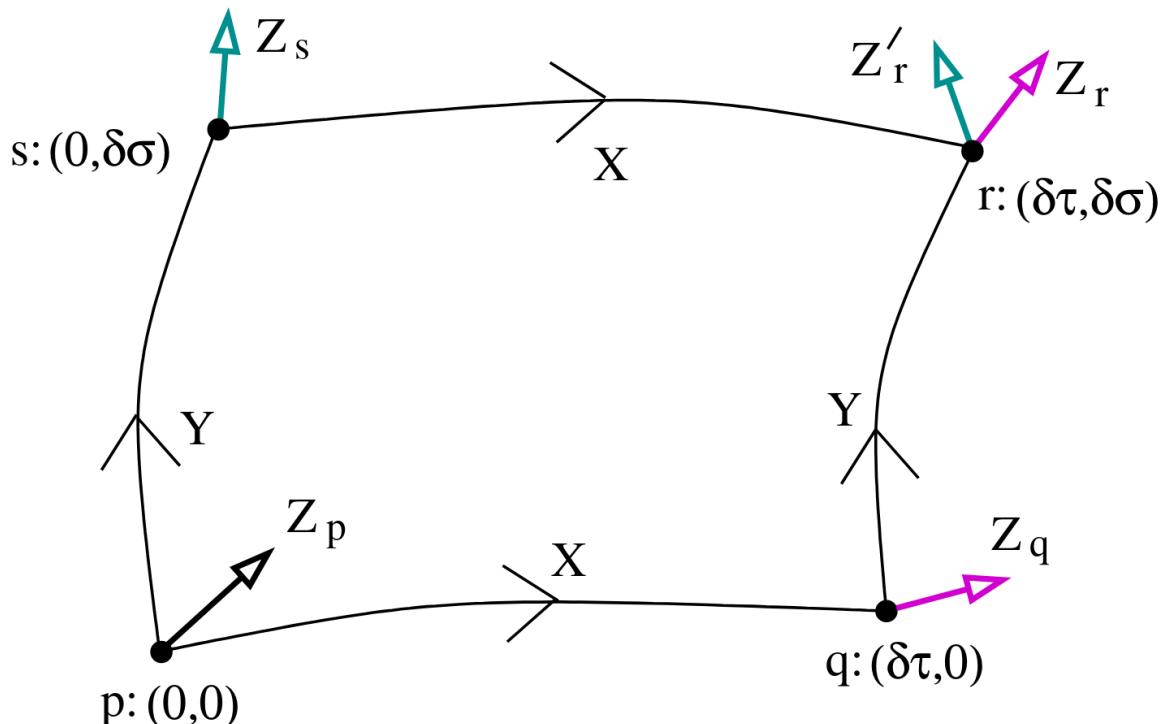


Figure 2.2: Parallel transport along different paths.

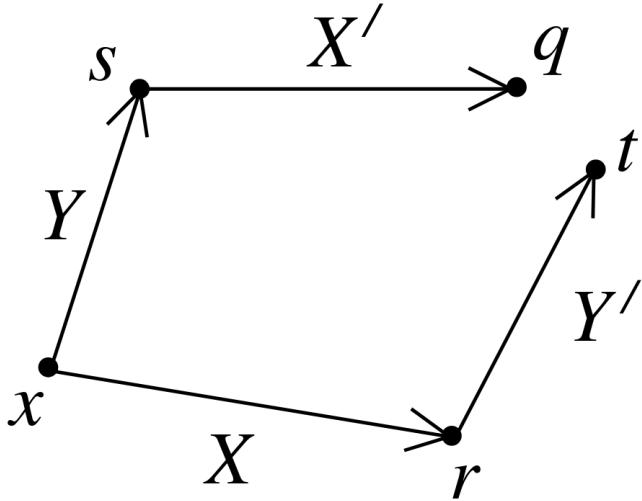


Figure 2.3: Visualization of torsion.

The expansions of Y^ν and Z^ρ at leading order multiply $\Gamma_{\nu\rho}^\mu(p) = 0$, thus only contribute to higher order terms. Next order in $\delta\sigma$:

$$\frac{d^2 Z^\mu}{d\sigma^2} \Big|_{\sigma=0} = - \left[\left(\frac{dY^\nu}{d\sigma} Z^\rho + Y^\nu \frac{dZ^\rho}{d\sigma} \right) \Gamma_{\nu\rho}^\mu + Y^\nu Z^\rho \frac{d\Gamma_{\nu\rho}^\mu}{d\sigma} \right]_q = - [Y^\nu Y^\sigma Z^\rho \Gamma_{\nu\rho,\sigma}^\mu]_p + o(\delta\tau)$$

The complete expansion thus is:

$$\begin{aligned} Z_r^\mu &= Z_q^\mu - [Y^\nu Z^\rho X^\sigma \Gamma_{\nu\rho,\sigma}^\mu]_p \delta\tau \delta\sigma - \frac{1}{2} [Y^\nu Y^\sigma Z^\rho \Gamma_{\nu\rho,\sigma}^\mu]_p \delta\sigma^2 + o(\delta^3) \\ &= Z_p^\mu - \frac{1}{2} \Gamma_{\nu\rho,\sigma}^\mu(p) [X^\nu X^\sigma Z^\rho \delta\tau^2 + 2Y^\nu Z^\rho X^\sigma \delta\tau \delta\sigma + Y^\nu Y^\sigma Z^\rho \delta\sigma^2] + o(\delta^3) \end{aligned}$$

Parallel transport along γ' leads to a similar expression (exchange of X and Y):

$$Z'_r^\mu = Z_p^\mu - \frac{1}{2} \Gamma_{\nu\rho,\sigma}^\mu(p) [Y^\nu Y^\sigma Z^\rho \delta\sigma^2 + 2X^\nu Z^\rho Y^\sigma \delta\sigma \delta\tau + X^\nu X^\sigma Z^\rho \delta\tau^2] + o(\delta^3)$$

The difference between the parallelly transported tangent vectors to leading order is:

$$\Delta Z_r^\mu = Z_r^\mu - Z'_r^\mu = - [\Gamma_{\nu\rho,\sigma}^\mu - \Gamma_{\sigma\rho,\nu}^\mu]_p [Y^\nu Z^\rho X^\sigma]_p \delta\tau \delta\sigma + o(\delta^3)$$

Recalling that $\Gamma_{\nu\rho}^\mu(p) = 0$, it is possible to write:

$$\Delta Z_r^\mu = - [R_{\rho\sigma\nu}^\mu Y^\nu Z^\rho X^\sigma]_p \delta\sigma \delta\tau + o(\delta^3) \quad (2.3.5)$$

It would be possible to evaluate the expression at r too, as it would differ only by higher order terms. Although the calculation was carried in a particular choice of coordinates, Eq. 2.3.5 is a tensor relation, therefore it must hold in all coordinate systems: the Riemann tensor thus determines the path dependence of parallel transport.

§2.3.2.1 Torsion

Consider two tangent vectors $X_p, Y_p \in T_p \mathcal{M}$ and a coordinate system x^μ such that $X_p = X^\mu \partial_\mu$ and $Y_p = Y^\mu \partial_\mu$. If $p : x^\mu$, as in Fig. 2.3 construct $r, s \in \mathcal{M}$ such that $r : x^\mu + \varepsilon X^\mu$ and

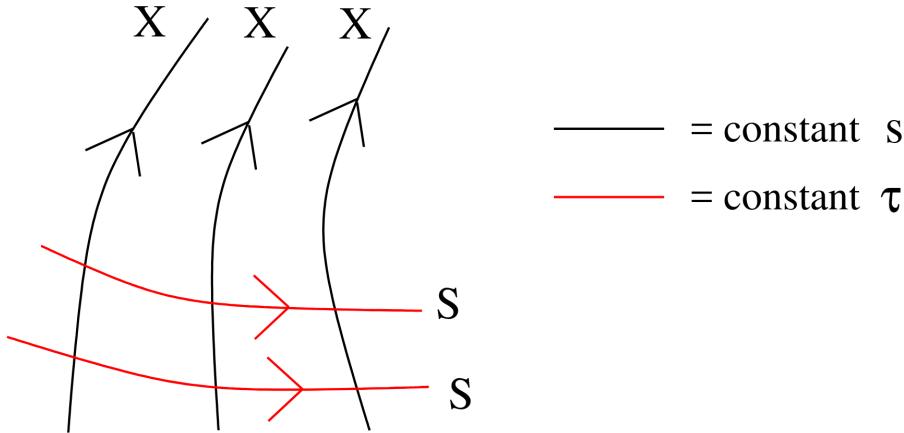


Figure 2.4: A one-parameter family of geodesics generated by X .

$s : x^\mu + \varepsilon Y^\mu$, with ε an infinitesimal parameter. Now, parallel transport $X_p \in T_p \mathcal{M}$ along the direction of Y_p to $X'_s \in T_s \mathcal{M}$ and $Y_p \in T_p \mathcal{M}$ along X_p to $Y'_r \in T_r \mathcal{M}$; their components will be:

$$X'_s = (X^\mu - \varepsilon \Gamma_{\nu\rho}^\mu Y^\nu X^\rho) \quad Y'_r = (Y^\mu - \varepsilon \Gamma_{\nu\rho}^\mu X^\nu Y^\rho)$$

Repeating this process, starting from point s and moving along the direction of X'_s , a new point $q \in \mathcal{M}$ is determined, with coordinates:

$$q : x^\mu + \varepsilon (X^\mu + Y^\mu) - \varepsilon^2 \Gamma_{\nu\rho}^\mu Y^\nu X^\rho$$

Analogously, starting at point r and moving along the direction of Y'_r , a new point $t \in \mathcal{M}$ is determined, with coordinates:

$$t : x^\mu + \varepsilon (X^\mu + Y^\mu) - \varepsilon^2 \Gamma_{\nu\rho}^\mu X^\nu Y^\rho$$

If the connection is torsion-free, then $q \equiv t$. On the other hand, if $T_{\nu\rho}^\mu \neq 0$, the parallelogram fails to close, as in Fig. 2.3.

§2.3.3 Geodesic deviation

Definition 2.3.3 (Tangent and deviation vector fields)

Given a one-parameter family of geodesics $\{x^\mu(\tau; s)\}_{s \in \mathbb{R}}$ on a manifold \mathcal{M} , the **tangent vector field** and the **deviation vector field** are defined as:

$$X^\mu := \frac{\partial x^\mu}{\partial \tau} \Big|_s \quad S^\mu := \frac{\partial x^\mu}{\partial s} \Big|_\tau \quad (2.3.6)$$

The meaning of these vector fields is evident: the tangent vector field fixes a particular geodesics (i.e. a particular s) and assigns at each point of the geodesic its tangent vector, while the deviation vector field fixes a particular value of the affine parameter τ and assigns at each point with this value a vector which takes to a nearby geodesic (at the same τ).

The family of geodesics sweeps a surface embedded in the manifold, so there is freedom in the choice of coordinates s and τ . In particular, it's always possible to pick them so that $X = \frac{\partial}{\partial \tau}$ and $S = \frac{\partial}{\partial s}$, in order for them to be linearly independent: $[X, S] = 0$, as in Fig. 2.4. Consider a torsion-free connection, so that:

$$\nabla_X S - \nabla_S X = [X, S] = 0 \implies \nabla_X \nabla_X S = \nabla_X \nabla_S X = \nabla_S \nabla_X X + \mathbf{R}(X, S)X$$

But X is tangent to geodesics, so from Eq. 2.3.3 $\nabla_X X = 0$ and:

$$\nabla_X \nabla_X S = R(X, S)X \quad (2.3.7)$$

Restricting to an integral curve γ of the vector field X , i.e. $X^\mu|_\gamma = \frac{dx^\mu}{d\tau}$, the covariant derivative along γ becomes:

$$\nabla_X|_\gamma = X^\mu|_\gamma \nabla_\mu = \frac{dx^\mu}{d\tau} \nabla_\mu \equiv \frac{D}{D\tau}$$

Hence, in index notation, the change of the deviation vector along the geodesic is expressed as:

$$\frac{D^2 S^\mu}{D\tau^2} = R^\mu_{\nu\rho\sigma} X^\rho S^\sigma X^\nu \quad (2.3.8)$$

This can be interpreted as the relative acceleration of neighbouring geodesics, and it is determined by the Riemann tensor. Experimentally, such geodesic deviations are called **tidal forces**, which are the non-local manifestation of a gravitational field as a curvature of spacetime.

§2.4 Riemann tensor

Recall Eq. 2.2.13 for the components of the Riemann tensor $R^\sigma_{\rho\mu\nu}$: it is manifestly anti-symmetric in its last two indices, but there are other subtle symmetries when using the Levi–Civita connection.

Proposition 2.4.1 (Symmetries of the Riemann tensor)

On a metric manifold with a Levi–Civita connection:

$$R_{\sigma\rho\mu\nu} = -R_{\sigma\rho\nu\mu} = -R_{\rho\sigma\mu\nu} = R_{\mu\nu\sigma\rho} \quad R_{\sigma[\rho\mu\nu]} = 0 \quad (2.4.1)$$

Proof. Set normal coordinates centered at a point p : then, $\Gamma^\mu_{\nu\rho} = 0$ and $\partial_\mu g^{\lambda\sigma} = 0$ at that point. At p , the Riemann tensor can be written as:

$$\begin{aligned} R_{\sigma\rho\mu\nu} &= g_{\sigma\lambda} R^\lambda_{\rho\mu\nu} = g_{\sigma\lambda} [\partial_\mu \Gamma^\lambda_{\nu\rho} - \partial_\nu \Gamma^\lambda_{\mu\rho}] \\ &= \frac{1}{2} [\partial_\mu (\partial_\nu g_{\sigma\rho} + \partial_\rho g_{\nu\sigma} - \partial_\sigma g_{\nu\rho}) - \partial_\nu (\partial_\mu g_{\sigma\rho} + \partial_\rho g_{\mu\sigma} - \partial_\sigma g_{\mu\rho})] \\ &= \frac{1}{2} [\partial_\mu \partial_\rho g_{\nu\sigma} - \partial_\mu \partial_\sigma g_{\nu\rho} - \partial_\nu \partial_\rho g_{\mu\sigma} + \partial_\nu \partial_\sigma g_{\mu\rho}] \end{aligned}$$

The symmetries are then manifest, and being these tensor equations they are valid in all coordinate systems. \square

An important computation tool is the Bianchi identity.

Theorem 2.4.1 (Bianchi identity)

On a metric manifold with a Levi–Civita connection:

$$\nabla_{[\lambda} R_{\sigma\rho]\mu\nu} = 0 \iff R^\sigma_{\rho[\mu\nu;\lambda]} = 0 \quad (2.4.2)$$

Proof. The two equations are equivalent, so the proof is of the first one. In normal coordinates $\nabla_\mu = \partial_\mu$ at p , so schematically: $R = \partial\Gamma + \Gamma\Gamma$, thus $\nabla R = \partial R + \Gamma\partial\Gamma = \partial^2\Gamma$. Explicitly:

$$\partial_\lambda R_{\sigma\rho\mu\nu} = \frac{1}{2} \partial_\lambda [\partial_\mu \partial_\rho g_{\nu\sigma} - \partial_\mu \partial_\sigma g_{\nu\rho} - \partial_\nu \partial_\rho g_{\mu\sigma} + \partial_\nu \partial_\sigma g_{\mu\rho}]$$

Anti-symmetrizing the appropriate indices yields the result. \square

Note that Eq. 2.4.1–2.4.2 do not require that the connection is a Levi–Civita connection, but are valid for general torsion-free connections.

§2.4.1 Ricci and Einstein tensors

Definition 2.4.1 (Ricci tensor and scalar)

On a metric manifold, the **Ricci tensor** is defined as:

$$R_{\mu\nu} := R^\rho_{\mu\rho\nu} \quad (2.4.3)$$

The **Ricci scalar** is defined as:

$$R := g^{\mu\nu} R_{\mu\nu} \quad (2.4.4)$$

Proposition 2.4.2 (Symmetry)

On a metric manifold with a Levi–Civita connection:

$$R_{\mu\nu} = R_{\nu\mu} \quad (2.4.5)$$

Proof. Using Eq. 2.4.1: $R_{\mu\nu} = g^{\sigma\rho} R_{\sigma\mu\rho\nu} = g^{\rho\sigma} R_{\rho\nu\sigma\mu} = R_{\nu\mu}$. \square

Proposition 2.4.3

On a metric manifold:

$$\nabla^\mu R_{\mu\nu} = \frac{1}{2} \nabla_\nu R \quad (2.4.6)$$

Proof. Writing explicitly Bianchi identity:

$$\nabla_\lambda R_{\sigma\rho\mu\nu} + \nabla_\sigma R_{\rho\lambda\mu\nu} + \nabla_\rho R_{\lambda\sigma\mu\nu} = 0$$

Contracting with $g^{\mu\lambda} g^{\rho\nu}$:

$$\nabla^\mu R_{\mu\nu} - \nabla_\nu R + \nabla^\nu R_{\nu\mu} = 0$$

which yields the thesis. \square

An important tensor obtained as a combination of the Ricci tensor and scalar is the **Einstein tensor**, which encodes information on the curvature of the metric manifold on which it is defined:

$$G_{\mu\nu} := R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \quad (2.4.7)$$

Most importantly, the Einstein tensor is covariantly conserved, as of Eq. 2.4.6:

$$\nabla^\mu G_{\mu\nu} = 0 \quad (2.4.8)$$

§2.4.2 Connection and curvature forms

This section focuses on a Lorentzian manifold, but the discussion is equivalent on a Riemannian one: it is sufficient to swap η_{ab} with δ_{ab} as the flat metric.

§2.4.2.1 Vielbeins

Although typically calculations are carried on a coordinate basis $\{e_\mu\} = \{\partial_\mu\}$, there are possible bases without such an interpretation. For example, a linear combination of a coordinate basis is not in general a coordinate basis itself:

$$\hat{e}_a = e_a^\mu \partial_\mu \quad (2.4.9)$$

A particularly useful non-coordinate basis is one such that:

$$g(\hat{e}_a, \hat{e}_b) = g_{\mu\nu} e_a^\mu e_b^\nu = \eta_{ab} \quad (2.4.10)$$

The components e_a^μ are called **vielbeins** or *tetrads*. In this non-coordinate system, the manifold looks flat (or, at least, its patch covered by the given chart). In the following computations, Greek indices are raised/lowered by the metric $g_{\mu\nu}$, while Latin indices by the metric η_{ab} . The vielbeins are not unique, for given a set of vielbeins e_a^μ it is always possible to find a new one:

$$\tilde{e}_a^\mu = e_b^\mu (\Lambda^{-1})_a^b \quad (2.4.11)$$

The transformation matrix must satisfy the condition imposed by Eq. 2.4.10, i.e.:

$$\Lambda_a^c \Lambda_b^d \eta_{cd} = \eta_{ab} \quad (2.4.12)$$

These are **local Lorentz transformations**, because the condition is that of Lorentz transformation, but Λ is now allowed to vary over the manifold. The dual basis of one-forms $\{\hat{\theta}^a\}$ is defined by $\hat{\theta}^a(\hat{e}_b) = \delta_b^a$. The relation to the coordinate basis is:

$$\hat{\theta}^a = e_a^\mu dx^\mu \quad (2.4.13)$$

where the coefficients satisfy $e_a^\mu e_b^\nu = \delta_b^a$ and $e_a^\mu e_a^\nu = \delta_\mu^\nu$. The metric is a tensor, therefore $g = g_{\mu\nu} dx^\mu \otimes dx^\nu = \eta_{ab} \hat{\theta}^a \otimes \hat{\theta}^b$, thus it is related to the vielbeins by:

$$g_{\mu\nu} = e_a^\mu e_\nu^b \eta_{ab} \quad (2.4.14)$$

§2.4.2.2 Connection 1-form

On a non-coordinate basis $\{\hat{e}_a\}$, connection components are computed in the usual way:

$$\nabla_{\hat{e}_c} \hat{e}_b = \Gamma_{cb}^a \hat{e}_a \quad (2.4.15)$$

However, these are not the same components $\Gamma_{\nu\rho}^\mu$ as in a coordinate basis.

Definition 2.4.2 (Connection 1-form)

On a metric manifold with vielbeins, the **connection 1-form** is defined as:

$$\omega_b^a := \Gamma_{cb}^a \hat{\theta}^c \quad (2.4.16)$$

Note that these are really n^2 1-forms, according to values of $a, b = 1, \dots, n \equiv \dim_{\mathbb{R}} \mathcal{M}$. This is also known as the *spin connection*, due to its relationship to spinors in curved spacetime.

Proposition 2.4.4 (Transformation law)

Given a local Lorentzian transformation Λ :

$$\tilde{\omega}_b^a = \Lambda_c^a \omega_d^c (\Lambda^{-1})_b^d + \Lambda_c^a (d\Lambda^{-1})_b^c \quad (2.4.17)$$

The second term reflects the second term in Prop. 2.2.1, which involves the derivative of the coordinate transformation.

The curvature of a metric manifold can also be studied through Cartan's structure equations, which encode the curvature information into forms.

Theorem 2.4.2 (First Cartan structure equation)

On a metric manifold with a torsion-free connection:

$$d\hat{\theta}^a + \omega^a{}_b \wedge \hat{\theta}^b = 0 \quad (2.4.18)$$

Proof. From Eq. 2.4.15 $\Gamma_{cb}^a = e^a{}_\rho e_c{}^\mu \nabla_\mu e_b{}^\rho$, thus, remembering $\Gamma_{[\mu\nu]}^\rho = 0$ (torsion-free):

$$\begin{aligned} \omega^a{}_b \wedge \hat{\theta}^b &= \Gamma_{cb}^a (e^c{}_\mu dx^\mu) \wedge (e^b{}_\nu dx^\nu) \\ &= e^a{}_\rho e_c{}^\mu (\partial_\mu e_b{}^\rho + e_b{}^\nu \Gamma_{\mu\nu}^\rho) (e^c{}_\mu dx^\mu) \wedge (e^b{}_\nu dx^\nu) \\ &= e^a{}_\rho \underbrace{e_c{}^\lambda e_c{}_\mu}_{\delta_\mu^\lambda} e^b{}_\nu (\partial_\lambda e_b{}^\rho + e_b{}^\sigma \Gamma_{\lambda\sigma}^\rho) dx^\mu \wedge dx^\nu = e^a{}_\rho e^b{}_\nu \partial_\mu e_b{}^\rho dx^\mu \wedge dx^\nu \end{aligned}$$

But $e^b{}_\nu e_b{}^\rho = \delta_\nu^\rho$, so $e^b{}_\nu \partial_\mu e_b{}^\rho = -e_b{}^\rho \partial_\mu e^n{}_\nu$, hence:

$$\omega^a{}_b \wedge \hat{\theta}^b = -e^a{}_\rho e_b{}^\rho \partial_\mu e^b{}_\nu dx^\mu \wedge dx^\nu = -\partial_\mu e^a{}_\nu dx^\mu \wedge dx^\nu = -d\hat{\theta}^a$$

which is the structure equation. \square

For a Levi–Civita connection, a stronger result holds.

Proposition 2.4.5

On a metric manifold with a Levi–Civita connection:

$$\omega_{ab} = -\omega_{ba} \quad (2.4.19)$$

Proof. Being the Levi–Civita connection compatible with the metric:

$$\begin{aligned} \Gamma_{abc} &= \eta_{ad} e^d{}_\rho e_b{}^\mu \nabla_\mu e_c{}^\rho = -\eta_{ad} e_c{}^\rho e_b{}^\mu \nabla_\mu e^d{}_\rho = -\eta_{cf} e^f{}_\sigma e_b{}^\mu \nabla_\mu (\eta_{ad} g^{\rho\sigma} e^d{}_\rho) \\ &= -\eta_{cf} e^f{}_\rho e_b{}^\mu \nabla_\mu e_a{}^\rho = -\Gamma_{cba} \end{aligned}$$

From Eq. 2.4.16 $\omega_{ab} = \Gamma_{acb} \hat{\theta}^c$, thus completing the proof. \square

Eq. 2.4.18–2.4.19 allow to quickly compute the spin connection, as they uniquely define it. Indeed, ω_{ab} being anti-symmetric means that there are $\frac{1}{2}n(n-1)$ independent 1-forms, i.e. $\frac{1}{2}n^2(n-1)$ independent components. The Cartan structure equation relates two 2-forms, each with $\frac{1}{2}n(n-1)$ independent components, thus posing $\frac{1}{2}n^2(n-1)$ constraints (as there are n equations) and uniquely fixing the spin connection.

§2.4.2.3 Curvature 2-form

Recall Eq. 2.4.1, which holds for Levi–Civita connections. Computing the Riemann tensor in the non-coordinate basis $R^a{}_{bcd} = R(\hat{\theta}^a, \hat{e}_b, \hat{e}_c, \hat{e}_d)$, the anti-symmetry of the last two indices persists: $R^a{}_{bcd} = -R^a{}_{bdc}$.

Definition 2.4.3 (Curvature 2-form)

On a metric manifold with a Levi–Civita connection, the **curvature 2-form** is defined as:

$$\mathcal{R}^a{}_b := \frac{1}{2} R^a{}_{bcd} \hat{\theta}^c \wedge \hat{\theta}^d \quad (2.4.20)$$

Again, these are really n^2 2-forms. A second Cartan structure relation holds.

Theorem 2.4.3 (Second Cartan structure equation)

On a metric manifold with a Levi–Civita connection:

$$\mathcal{R}^a{}_b = d\omega^a{}_b + \omega^a{}_c \wedge \omega^c{}_b \quad (2.4.21)$$

Connection and curvature forms make computing the Riemann tensor less tedious, as exterior derivatives take significantly less effort than covariant derivatives.

Part II

General Relativity

Chapter 3

External Gravitational Fields

In classical field theories, two distinct classes of objects are considered: particles and fields. Fields determine the motion of particles, while particles induce oscillations in fields.

§3.1 Geodetic motion

Free particles moving on an n -dimensional metric manifold (\mathcal{M}, g) have geodesics as trajectories, which are found as extrema of the action functional.

§3.1.1 Non-relativistic particles

The action of a free particle is composed of the sole kinetic term, which, for a non-relativistic particle, is $\frac{1}{2}m\dot{\mathbf{x}} \cdot \dot{\mathbf{x}}$, hence:

$$L = \frac{m}{2}g_{ij}(\mathbf{x})\dot{x}^i\dot{x}^j \quad (3.1.1)$$

where $\{x^i\}_{i=1,\dots,n}$ are the coordinates of the chart(s) which cover the portion of \mathcal{M} spanned by the trajectory.

Proposition 3.1.1 (Non-relativistic motion)

The equations of motion of the Lagrangian Eq. 3.1.1 are:

$$\ddot{x}^i + \Gamma_{jk}^i \dot{x}^j \dot{x}^k = 0 \quad (3.1.2)$$

where Γ_{jk}^i are the Christoffel symbols of the Levi–Civita connection on \mathcal{M} .

Proof. Varying the action of Eq. 3.1.1 on trajectories between $\mathbf{x}_1 \equiv \mathbf{x}(t_1), \mathbf{x}_2 \equiv \mathbf{x}(t_2) \in \mathcal{M}$:

$$\delta S = \frac{m}{2} \int_{\mathbf{x}_1}^{\mathbf{x}_2} d^n x \sqrt{g} \delta [g_{ij} \dot{x}^i \dot{x}^j] = \frac{m}{2} \int_{\mathbf{x}_1}^{\mathbf{x}_2} d^n x \sqrt{g} \delta x^k \left[g_{ij,k} \dot{x}^i \dot{x}^j - 2 \frac{d}{dt} (g_{ik} \dot{x}^i) \right]$$

Setting the term in parentheses to zero yields:

$$\frac{1}{2}g_{ij,k} \dot{x}^i \dot{x}^j - g_{ik,j} \dot{x}^i \dot{x}^j - g_{ik} \ddot{x}^i = 0$$

Symmetrizing $\frac{1}{2}g_{ij,k} - g_{ik,j}$ results in $-g_{il}\Gamma_{jk}^l$, hence the geodesic equation is recovered. \square

This is precisely the geodesic equation on \mathcal{M} .

§3.1.2 Relativistic particles

Consider now a relativistic particle moving on an n -dimensional Lorentzian manifold (\mathcal{M}, g) . In this case, the free action is the distance between the two endpoints $x_1, x_2 \in \mathcal{M}$ of the trajectory:

$$\mathcal{S} = -mc \int_{x_1}^{x_2} \sqrt{-ds^2} = -mc \int_{\sigma_1}^{\sigma_2} d\sigma \sqrt{-g_{\mu\nu} \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma}} \quad (3.1.3)$$

where the negative sign clarifies that particles can only move on timelike and null geodesics, and $\sigma \in [\sigma_1, \sigma_2] \subseteq \mathbb{R}$ parametrizes the trajectory. Note that $g_{\mu\nu} = g_{\mu\nu}(x(\sigma))$, and the Lagrangian of the relativistic free particle is:

$$L = -mc \sqrt{-g_{\mu\nu} \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma}} \quad (3.1.4)$$

This action is clearly reparametrization-invariant, since:

$$\tilde{\mathcal{S}} = -mc \int_{\tilde{\sigma}_1}^{\tilde{\sigma}_2} d\tilde{\sigma} \sqrt{-g_{\mu\nu} \frac{dx^\mu}{d\tilde{\sigma}} \frac{dx^\nu}{d\tilde{\sigma}}} = -mc \int_{\sigma_1}^{\sigma_2} d\sigma \frac{d\tilde{\sigma}}{d\sigma} \sqrt{-g_{\mu\nu} \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma}} \left(\frac{d\sigma}{d\tilde{\sigma}} \right)^2 = \mathcal{S}$$

By this invariance, the value of the action between two points on a trajectory has a well-defined meaning: it is the elapsed **proper time** measured by the particle along the motion:

$$\tau(\sigma) \equiv \frac{1}{c} \int_0^\sigma d\sigma' \sqrt{-g_{\mu\nu} \frac{dx^\mu}{d\sigma'} \frac{dx^\nu}{d\sigma'}} \quad (3.1.5)$$

Being this a monotone function on the trajectory, it is possible to parametrize the action in terms of proper time:

$$\mathcal{S} = -mc \int_{\tau_1}^{\tau_2} d\tau \sqrt{-g_{\mu\nu} \dot{x}^\mu \dot{x}^\nu} \quad (3.1.6)$$

where it is conventional to set $\dot{x} \equiv \frac{dx}{d\tau}$.

Proposition 3.1.2 (Relativistic motion)

The equations of motion of the action Eq. 3.1.3 are:

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma_{\nu\rho}^\mu \frac{dx^\nu}{d\tau} \frac{dx^\rho}{d\tau} = 0 \quad (3.1.7)$$

where $\Gamma_{\nu\rho}^\mu$ are the Christoffel symbols of the Levi–Civita connection on \mathcal{M} .

Proof. Varying the action yields the following equations of motion:

$$-\frac{1}{2L} g_{\mu\nu,\rho} \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma} - \frac{d}{d\sigma} \left(-\frac{1}{L} g_{\rho\nu} \frac{dx^\nu}{d\sigma} \right) = 0$$

Performing the differentiation and using the fact that $L \neq 0$ (for a massive particle):

$$g_{\mu\rho} \frac{d^2x^\rho}{d\sigma^2} + \frac{1}{2} (g_{\mu\rho,\nu} + g_{\mu\nu,\rho} - g_{\nu\rho,\mu}) \frac{dx^\mu}{d\sigma} \frac{dx^\rho}{d\sigma} = \frac{1}{L} \frac{dL}{d\sigma} g_{\mu\rho} \frac{dx^\rho}{d\sigma}$$

This is the equation for a non-affinely-parametrized geodesic. To see this, note that from

Eq. 3.1.5:

$$c \frac{d\tau}{d\sigma} = L(\sigma) \implies L(\tau) = \sqrt{-g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau}} = \frac{d\sigma}{d\tau} L(\sigma) = c$$

Therefore, the Lagrangian can be made constant with any affine parametrization $\sigma = a\tau + b$, with $a \in \mathbb{R} - \{0\}$, $b \in \mathbb{R}$. This completes the proof. \square

It is straightforward to compute the conjugate momentum from Eq. 3.1.4:

$$p_\mu = \frac{\partial L}{\partial \dot{x}^\mu} = -mc \frac{\partial}{\partial \dot{x}^\mu} \sqrt{-g_{\nu\rho} \dot{x}^\nu \dot{x}^\rho} = -\frac{m^2 c^2}{L} \dot{x}_\mu$$

which results into the **mass-shell condition**:

$$p^2 = -m^2 c^2 \quad (3.1.8)$$

Therefore, the time component of p^μ is found to be $(p^0)^2 = m^2 c^2 + \mathbf{p}^2$, which has two important consequences: first of all, $p^0 \neq 0$, which means that the particle is never at rest in the temporal direction (which was to be expected); secondly, p^μ only has $n - 1$ independent components, which is a consequence of the reparametrization invariance of the action¹.

§3.1.3 Interactions

Once the relativistic motion of a (massive) free particle on a general n -dimensional Lorentzian manifold has been clarified, focus now on Minkowski spacetime $\mathbb{R}^{1,3}$ with the Lorentz–Minkowski metric $\eta_{\mu\nu} = \text{diag}(-1, +1, +1, +1)$. The action then takes the form:

$$\mathcal{S} = -mc \int_{\sigma_1}^{\sigma_2} d\sigma \sqrt{-\eta_{\mu\nu} \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma}}$$

Clearly, there are two possibilities to introduce an interaction term: either inside or outside the square root.

§3.1.3.1 Electromagnetism

The addition of a term $+d\sigma V(x)$ outside the square root is not reparametrization-invariant. Indeed, the simplest interaction term which satisfies both reparametrization invariance and Lorentz invariance is:

$$\mathcal{S} = \int_{\sigma_1}^{\sigma_2} d\sigma \left[-mc \sqrt{-\eta_{\mu\nu} \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma}} - qA_\mu(x) \frac{dx^\mu}{d\sigma} \right] \quad (3.1.9)$$

where $A_\mu(x)$ is the potential of the field and $q \in \mathbb{R}$ is the charge of the coupling. To determine the potential A_μ for the electromagnetic interaction, compare the classical limit of Eq. 3.1.9 to the action of a non-relativistic particle in an external electromagnetic field:

$$\mathcal{S} \xrightarrow{c \rightarrow \infty} \int_{t_1}^{t_2} dt \left[-mc^2 + \frac{m}{2} \dot{\mathbf{x}}^2 - qA_0(x)c - q\mathbf{A}(x) \cdot \dot{\mathbf{x}} \right]$$

¹The n equations of motion are $x^\mu = x^\mu(\sigma)$, but σ cannot contain any information about the system, thus one equation must be used to eliminate the σ dependence, resulting in $n - 1$ degrees of freedom. This was to be expected from the non-relativistic limit, which has $n - 1$ degrees of freedom too (since time is absolute in this limit).

$$\mathcal{S}_{\text{em}} = \int_{t_1}^{t_2} dt \left[\frac{m}{2} \dot{\mathbf{x}}^2 - q\phi(x) - q\mathbf{A}(x) \cdot \dot{\mathbf{x}} \right]$$

Comparing the individual terms, the electromagnetic interaction is described by the potential $A_\mu(x) = (\phi(x)/c, \mathbf{A}(x))$.

§3.1.3.2 Gravity

While the addition of an interaction outside the square root forces it to be described by a vector field (at least in its simplest form), doing so inside the square root requires a tensor field, in order to satisfy both reparametrization invariance and Lorentz invariance. The resulting action then is:

$$\mathcal{S} = -mc \int_{\sigma_1}^{\sigma_2} d\sigma \sqrt{-(\eta_{\mu\nu} + V_{\mu\nu}(x)) \frac{dx^\mu}{d\sigma} \frac{dx^\nu}{d\sigma}}$$

However, this can be interpreted² as the action of a relativistic free particle on a curved 4-dimensional Lorentzian manifold with metric $g_{\mu\nu}(x) \equiv \eta_{\mu\nu} + V_{\mu\nu}(x)$. To compute the non-relativistic limit of this action, expand the term in the square root:

$$\mathcal{S} = -mc^2 \int_{t_1}^{t_2} dt \sqrt{-g_{00}(x) - \frac{1}{c} \left(2g_{0i}\dot{x}^i + \frac{1}{c} g_{ij}\dot{x}^i\dot{x}^j \right)}$$

The only interesting case is $V_{\mu\nu}(x) \equiv -c^{-2}V(x)\delta_{\mu,0}\delta_{\nu,0}$, in which case:

$$\mathcal{S} \xrightarrow{c \rightarrow \infty} \int_{t_1}^{t_2} dt \left[-mc^2 - m\frac{V(x)}{2} + \frac{m}{2} \dot{\mathbf{x}}^2 \right]$$

Now, compare this to the action of a non-relativistic particle in an external gravitational field:

$$\mathcal{S}_g = \int_{t_1}^{t_2} dt \left[-mc^2 + \frac{m}{2} \dot{\mathbf{x}}^2 - m\Phi(x) \right]$$

The comparison gives $V(x) = 2\Phi(x)$: therefore, the gravitational interaction naturally induces the notion of a curved spacetime. Moreover, this is the condition that a metric describing the gravitational interaction must satisfy in the weak-field limit (i.e. the non-relativistic limit):

$$g_{00}(x) \approx - \left[1 + \frac{2\Phi(x)}{c^2} \right] \quad (3.1.10)$$

with $\Phi(x)$ the Newtonian potential (indeed, this is seldom called the “Newtonian limit”).

§3.2 Equivalence principle

As per the discussion in §3.1.3.2, the particle’s mass m can be regarded as the “charge” which couples it to the gravitational field. This fact is known as the **weak equivalence principle** (WEP), which states the equivalence between inertial mass $m_i : \mathbf{F}_{\text{tot}} = m_i \ddot{\mathbf{x}}$ and gravitational mass m_g :

$$m_i = m_g \quad (3.2.1)$$

This equivalence has been experimentally verified with relative error $\sim 10^{-13}$.

²To be precise, this is true assuming that the tensor $g_{\mu\nu}(x)$ satisfies the conditions in Def. 2.1.1.

§3.2.1 Kottler–Möller metric

A direct consequence of the WEP is the indistinguishability of a uniform gravitational field from a constant acceleration.

Consider a massive particle with a constant acceleration $\mathbf{a} = a\hat{\mathbf{e}}_x$ with respect to an inertial reference frame \mathcal{O} in Minkowski spacetime. A constant acceleration means that the rapidity grows linearly in proper time as $\varphi(\tau) = \frac{a\tau}{c}$, so that the particle's velocity (along $\hat{\mathbf{e}}_x$) is:

$$v(\tau) = c \operatorname{sech} \frac{a\tau}{c} \quad (3.2.2)$$

Since $\frac{dt}{d\tau} = \gamma(\tau)$ and $\frac{dx}{d\tau} = v(\tau)$, inserting the expression for $v(\tau)$ in the Lorentz factor yields:

$$t(\tau) = \frac{c}{a} \sinh \frac{a\tau}{c} \quad x(\tau) = \frac{c^2}{a} \cosh \frac{a\tau}{c} - \frac{c^2}{a} \quad (3.2.3)$$

where the integration constants have been chosen so that $t(0) = 0$ and $x(0) = 0$. The worldline (i.e. trajectory) of the particle is thus described by a hyperbola in \mathcal{O} :

$$\left(x + \frac{c^2}{a} \right)^2 - c^2 t^2 = \frac{c^4}{a^2} \quad (3.2.4)$$

The coordinate transformation from the inertial reference frame \mathcal{O} , with coordinates (t, x) , to the non-inertial comoving reference frame of the particle, with coordinates (τ, ρ) , is given by:

$$ct = \left(\rho + \frac{c^2}{a} \right) \sinh \frac{a\tau}{c} \quad x = \left(\rho + \frac{c^2}{a} \right) \cosh \frac{a\tau}{c} - \frac{c^2}{a} \quad (3.2.5)$$

Indeed, the particle's spatial trajectory is correctly described by $\rho = 0$. Note that the comoving coordinates (τ, ρ) do not cover the whole Minkowski spacetime: this shows that there are regions of spacetime which are causally disconnected from the particle. The pull-back of the metric in the comoving frame is:

$$ds^2 = - \left(1 + \frac{a\rho}{c^2} \right)^2 c^2 d\tau^2 + d\rho^2 + dy^2 + dz^2 \quad (3.2.6)$$

This is known as the **Kottler–Möller metric**. This metric tensor has precisely the form discussed in §3.1.3.2, since in the sub-relativistic limit its temporal component is $g_{00} \approx -(1 + \frac{2a\rho}{c^2})$, and it describes a gravitational field with potential $\Phi(\rho) = a\rho$, which is a uniform gravitational field.

§3.2.2 Einstein's equivalence principle

A corollary of the WEP is that a uniform gravitational field can be nullified by choosing a particular non-inertial reference frame, the *free-fall reference frame*, which is comoving with the considered particle.

Einstein generalized the WEP to a stronger principle, **Einstein's equivalence principle**, which states that *any* gravitational field can be locally nullified by a suitable choice of (non-inertial) reference frame. Formally, this is equivalent to Th. 2.3.1: one can always choose normal coordinates in a given point so that spacetime locally looks flat, i.e. $g_{\mu\nu}(p) = \eta_{\mu\nu}$.

Although locally nullifiable, the effects of a non-uniform gravitational field become measurable when it is possible to conduct non-local measurements, due to the presence of **tidal forces**.

Example 3.2.1 (Einstein's elevator Gedankenexperiment)

Consider an observer in a closed box which is falling towards the Earth. Although in the free-fall reference frame the observer cannot establish whether it is fluctuating in space or falling towards the Earth, they can do so with a non-local experiment: taking two test masses (i.e. not affected by their mutual gravitational attraction), if the box is fluctuating in space they will remain in their initial position due to inertia, meanwhile if the box is falling towards the Earth they will have a horizontal displacement, due to the fact that the Earth's gravitational field is radial (and thus non-uniform). This is the case of a tidal force.

§3.2.3 Gravitational time dilation

Consider a spherically-symmetric body of uniform mass M . In the weak-field approximation, the gravitational field generated by the body is given by a metric such that $g_{00}(x) = -1 - \frac{2\Phi(x)}{c^2}$, with $\Phi(x) = -\frac{GM}{r}$ and $r \equiv \|x\|$. Therefore, a stationary observer at a distance r from the body will measure time intervals given by (using $d\tau^2 := -ds^2$):

$$d\tau^2 = \left(1 - \frac{2GM}{rc^2}\right) dt^2$$

Defining the Schwarzschild radius of a mass M as $R_s \equiv 2GM/c^2$ and integrating, the **gravitational time dilation** is found:

$$\Delta\tau(r) = \Delta t \sqrt{1 - \frac{R_s}{r}} \quad (3.2.7)$$

which means that time flows slower near massive objects.

Proposition 3.2.1 (Relative gravitational time dilation)

Given two stationary observers at distances r and $r + \Delta r$ from a mass M , with $\Delta r \ll r$ and $R_s \ll r$, then:

$$\frac{\Delta\tau(r + \Delta r)}{\Delta\tau(r)} = 1 + \frac{R_s \Delta r}{2r^2} \quad (3.2.8)$$

Proof. Using Eq. 3.2.7:

$$\begin{aligned} \Delta\tau(r + \Delta r) &= \Delta t \sqrt{1 - \frac{R_s}{r + \Delta r}} \approx \Delta t \sqrt{1 - \frac{R_s}{r} + \frac{R_s \Delta r}{r^2}} \\ &\approx \Delta t \sqrt{1 - \frac{R_s}{r}} \left(1 + \frac{R_s \Delta r}{2r^2}\right) = \Delta\tau(r) \left(1 + \frac{R_s \Delta r}{2r^2}\right) \end{aligned}$$

which is the thesis. \square

Accounting for the relative gravitational time dilation is extremely important in precision applications (like the GPS): for example, with an elevation difference of $\Delta r \sim 10^3$ m with respect to sea level $r \sim 6 \cdot 10^6$ m, the measured time difference is $\sim 10^{-12}$ s.

Proposition 3.2.2 (Gravitational redshift)

Consider a photon emitted with frequency ω_1 at a distance r_1 from the center of a weak gravitational potential $\Phi(r)$. An observer at distance r_2 will measure a shifted frequency ω_2

given by:

$$\frac{\omega_2}{\omega_1} = \left[1 + \frac{\Phi(r_2) - \Phi(r_1)}{c^2} \right]^{-1} \quad (3.2.9)$$

Proof. Since $\omega = \frac{2\pi}{\Delta\tau}$, where $\Delta\tau$ is the time elapsed during an oscillation of a full wavelength, by Eq. 3.1.10:

$$\frac{\Delta\tau_2}{\Delta\tau_1} = \sqrt{\frac{1 + 2\Phi(r_2)/c^2}{1 + 2\Phi(r_1)/c^2}} \approx 1 + \frac{\Phi(r_2) - \Phi(r_1)}{c^2}$$

This is equivalent to the thesis. \square

In the case of the (weak) gravitational field generated by a mass M , since $\Phi(r) \propto -r^{-1}$, then $r_2 > r_1 \implies \omega_2 < \omega_1$ (redshift) and vice versa (blueshift).

Chapter 4

Geometrodynamics

The force of gravity is mediated by a gravitational field, which is identified with a metric $g_{\mu\nu}(x)$ on a 4-dimensional Lorentzian manifold called spacetime. This metric is a dynamical object, as all other fields in Nature, thus the laws governing its dynamics must be studied.

§4.1 Einstein-Hilbert action

Differential Geometry places strict limits to the possible actions that can be formulated, ensuring it is something intrinsic to the metric and independent on the choice of coordinates.

Spacetime is a Lorentzian manifold \mathcal{M} , thus, recalling the canonical volume form in Eq. 2.1.14, there needs to be a metric-dependent scalar function on \mathcal{M} : the obvious non-trivial choice is the Ricci scalar. The resulting action is the **Einstein–Hilbert action**:

$$\mathcal{S} = \int d^4x \sqrt{-g} R \quad (4.1.1)$$

where the negative sign makes it manifest that the metric has signature $(-, +, +, +)$. Schematically, the Riemann tensor is $R \sim \partial\Gamma + \Gamma\Gamma$, while $\Gamma \sim \partial g$, thus the Einstein-Hilbert action is second order in derivatives of the metric, like most other actions (for bosonic fields) in physics.

§4.1.1 Equations of motion

To determine the Euler-Lagrange equations of the Einstein-Hilbert action, consider a fixed initial metric $g_{\mu\nu}(x)$ and a perturbation $g_{\mu\nu}(x) \mapsto g_{\mu\nu}(x) + \delta g_{\mu\nu}(x)$. For the inverse metric:

$$g_{\mu\nu}g^{\nu\rho} = \delta_\mu^\rho \implies \delta g_{\mu\nu}g^{\nu\rho} + g_{\mu\nu}\delta g^{\nu\rho} = 0 \implies \delta g^{\mu\nu} = -g^{\mu\rho}g^{\nu\sigma}\delta g_{\rho\sigma}$$

Lemma 4.1.1

The variation of $\sqrt{-g}$ is:

$$\delta\sqrt{-g} = -\frac{1}{2}\sqrt{-g}g_{\mu\nu}\delta g^{\mu\nu} \quad (4.1.2)$$

Proof. Recall that for invertible matrices $\log \det A = \text{tr} \log A$, i.e. $(\det A)^{-1}\delta(\det A) = \text{tr}(A^{-1}\delta A)$. Applying this to the metric:

$$\delta\sqrt{-g} = \frac{1}{2}\frac{1}{\sqrt{-g}}\delta(-g) = \frac{1}{2}\frac{1}{\sqrt{-g}}(-g)g^{\mu\nu}\delta g_{\mu\nu} = \frac{1}{2}\sqrt{-g}g^{\mu\nu}\delta g_{\mu\nu} = -\frac{1}{2}\sqrt{-g}g_{\mu\nu}\delta g^{\mu\nu}$$

which is the thesis. □

Lemma 4.1.2

The variation of the Christoffel symbols is:

$$\delta\Gamma_{\mu\nu}^\rho = \frac{1}{2}g^{\rho\sigma}(\nabla_\mu\delta g_{\sigma\nu} + \nabla_\nu\delta g_{\sigma\mu} - \nabla_\sigma\delta g_{\mu\nu}) \quad (4.1.3)$$

Proof. First note that, although $\Gamma_{\mu\nu}^\rho$ is not a tensor, $\delta\Gamma_{\mu\nu}^\rho$ is, for it is the difference of two Christoffel symbols, one computed with $g_{\mu\nu}$ and the other with $g_{\mu\nu} + \delta g_{\mu\nu}$, but the extra term in the transformation law of $\Gamma_{\mu\nu}^\rho$ is independent of the metric, thus cancels out. This observation allows working in normal coordinates at $p \in \mathcal{M}$, so that $\partial_\rho g_{\mu\nu} = 0$ and $\Gamma_{\mu\nu}^\rho = 0$ at that point. Then, to linear order in the variation, at p :

$$\delta\Gamma_{\mu\nu}^\rho = \frac{1}{2}g^{\rho\sigma}(\partial_\mu\delta g_{\sigma\nu} + \partial_\nu\delta g_{\sigma\mu} - \partial_\sigma\delta g_{\mu\nu}) = \frac{1}{2}g^{\rho\sigma}(\nabla_\mu\delta g_{\sigma\nu} + \nabla_\nu\delta g_{\sigma\mu} - \nabla_\sigma\delta g_{\mu\nu})$$

This is a tensor equation, hence valid in all coordinate system and, being p arbitrary, on all \mathcal{M} . \square

Lemma 4.1.3

The variation of the Ricci tensor is:

$$\delta R_{\mu\nu} = \nabla_\rho\delta\Gamma_{\mu\nu}^\rho - \nabla_\nu\delta\Gamma_{\rho\mu}^\rho \quad (4.1.4)$$

Proof. Working in normal coordinates, the Riemann tensor becomes $R^\sigma_{\mu\rho\nu} = \partial_\rho\Gamma_{\nu\mu}^\sigma - \partial_\nu\Gamma_{\rho\mu}^\sigma$, so:

$$\delta R^\sigma_{\mu\rho\nu} = \partial_\rho\delta\Gamma_{\nu\mu}^\sigma - \partial_\nu\delta\Gamma_{\rho\mu}^\sigma = \nabla_\rho\delta\Gamma_{\nu\mu}^\sigma - \nabla_\nu\delta\Gamma_{\rho\mu}^\sigma$$

This is a tensor equation, hence valid in all coordinates systems and on all \mathcal{M} . Contracting indices σ, ρ and working to leading order yields the result. \square

Proposition 4.1.1 (Einstein field equations in vacuum)

The Euler-Lagrange equations of the Einstein-Hilbert action are:

$$G_{\mu\nu} = 0 \quad (4.1.5)$$

Proof. Varying the Einstein-Hilbert action:

$$\begin{aligned} \delta\mathcal{S} &= \delta \int d^4x \sqrt{-g} g^{\mu\nu} R_{\mu\nu} \\ &= \int d^4x [\delta(\sqrt{-g})g^{\mu\nu} R_{\mu\nu} + \sqrt{-g}(\delta g^{\mu\nu})R_{\mu\nu} + \sqrt{-g}g^{\mu\nu}(\delta R_{\mu\nu})] \\ &= \int d^4x \sqrt{-g} \left[\left(-\frac{1}{2}Rg_{\mu\nu} + R_{\mu\nu} \right) \delta g^{\mu\nu} + g^{\mu\nu} (\nabla_\rho\delta\Gamma_{\mu\nu}^\rho - \nabla_\nu\delta\Gamma_{\rho\mu}^\rho) \right] \\ &= \int d^4x \sqrt{-g} \left[\left(R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} \right) \delta g^{\mu\nu} + \nabla_\mu (g^{\rho\nu}\delta\Gamma_{\rho\nu}^\mu - g^{\mu\nu}\delta\Gamma_{\rho\nu}^\rho) \right] \end{aligned}$$

The last term is a total derivative, hence by the divergence theorem Eq. 2.2.19 it yields

a boundary term which can be ignored (Gibbons-Hawking boundary term). The Euler-Lagrange equations are then found imposing $\delta\mathcal{S} = 0$ for arbitrary $\delta g_{\mu\nu}$, so recalling the definition of the Einstein tensor Eq. 2.4.7 the proof is completed. \square

These equations are the Einstein field equations in vacuum, which govern the geometrodynamics of spacetime in the absence of any matter. For this reason, they can be further simplified: by contracting with $g^{\mu\nu}$ one finds $R = 0$, thus:

$$R_{\mu\nu} = 0 \quad (4.1.6)$$

§4.1.1.1 Dimensional analysis

The Einstein-Hilbert action in Eq. 4.1.1 does not have the right dimensions, which will be necessary when considering the presence of matter. If x^μ has dimension of length, then $g_{\mu\nu}$ is dimensionless and the Ricci scalar is $[R] = L^{-2}$. Including the integration measure, $[\mathcal{S}] = L^2$, but it must have the same dimension of $[\hbar] = ML^2T^{-1}$ (i.e. energy \times times). Thus, the action with the right dimension is:

$$\mathcal{S} = \frac{c^3}{16\pi G} \int d^4x \sqrt{-g} R \quad (4.1.7)$$

In the following, natural units are adopted: $c = \hbar = 1$.

§4.1.1.2 Cosmological constant

In reality, Eq. 4.1.7 is not the simplest action allowed by Differential Geometry: in fact, a constant term could be added to the Ricci scalar. The action then becomes:

$$\mathcal{S} = \frac{1}{16\pi G} \int d^4x \sqrt{-g} (R - 2\Lambda) \quad (4.1.8)$$

Λ is a **cosmological constant** and has dimension $[\Lambda] = L^{-2}$. The resulting field equations are:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = -\Lambda g_{\mu\nu} \quad (4.1.9)$$

Contracting with $g^{\mu\nu}$ yields $R = 4\Lambda$, thus in the absence of matter:

$$R_{\mu\nu} = \Lambda g_{\mu\nu} \quad (4.1.10)$$

§4.1.2 Diffeomorphisms

Being the metric a symmetric $\mathbb{R}^{4\times 4}$ matrix, it should have $\frac{1}{2} \times 4 \times 5 = 10$ degrees of freedom. However, not all the components of $g_{\mu\nu}$ are physical. Indeed, two metrics related by a change of coordinates $x^\mu \mapsto \tilde{x}^\mu(x)$ describe the same physical spacetime, thus there is a redundancy in any given representation of the metric, which removes precisely 4 degrees of freedom, leaving only 6 physical degrees of freedom.

Mathematically, given a diffeomorphism $\phi : \mathcal{M} \rightarrow \mathcal{M}$, it maps all fields on \mathcal{M} to a new set of fields on \mathcal{M} : the result is physically indistinguishable from the original, describing the same spacetime but in different coordinates. Such diffeomorphisms are analogous to gauge symmetries in Yang-Mills theory.

Consider a diffeomorphism $x^\mu \mapsto \tilde{x}^\mu = x^\mu + \delta x^\mu$: this can be viewed as an active change, where points in spacetime are mapped from one another, or as a passive change, in which only the

coordinates of each point are affected, but the two interpretations are equivalent. This change of coordinates can be thought as generated by an infinitesimal vector field $X^\mu : \delta x^\mu = -X^\mu(x)$, so that the metric transforms as:

$$\begin{aligned}\tilde{g}_{\mu\nu}(\tilde{x}) &= \frac{\partial x^\rho}{\partial \tilde{x}^\mu} \frac{\partial x^\sigma}{\partial \tilde{x}^\nu} g_{\rho\sigma}(x) = (\delta_\mu^\rho + \partial_\mu X^\rho) (\delta_\nu^\sigma + \partial_\nu X^\sigma) g_{\rho\sigma}(x) \\ &= g_{\mu\nu}(x) + g_{\mu\rho}(x)\partial_\nu X^\rho + g_{\nu\rho}(x)\partial_\mu X^\rho\end{aligned}$$

Meanwhile, the Taylor expansion around $\tilde{x} = x + \delta x$ is:

$$\tilde{g}_{\mu\nu}(\tilde{x}) = \tilde{g}_{\mu\nu}(x + \delta x) = \tilde{g}_{\mu\nu}(x) - X^\lambda \partial_\lambda \tilde{g}_{\mu\nu}(x)$$

Comparing the metrics at the same point, it is understood that it undergoes an infinitesimal change:

$$\delta g_{\mu\nu}(x) = \tilde{g}_{\mu\nu}(x) - g_{\mu\nu}(x) = X^\lambda \partial_\lambda g_{\mu\nu}(x) + g_{\mu\rho}(x)\partial_\nu X^\rho + g_{\nu\rho}(x)\partial_\mu X^\rho$$

By Eq. 1.3.8, this is the Lie derivative of the metric:

$$\delta g_{\mu\nu} = (\mathcal{L}_X g)_{\mu\nu} \quad (4.1.11)$$

Thus, an infinitesimal diffeomorphism along $X \in \mathfrak{X}(\mathcal{M})$ makes the metric change by an infinitesimal amount given by its Lie derivative along X , which can be viewed as the leading term in the Taylor expansion along X . These equations can be rewritten in a simpler form:

$$\delta g_{\mu\nu} = X^\rho \partial_\rho g_{\mu\nu} + \partial_\nu X_\mu - X^\rho \partial_\nu g_{\mu\rho} + \partial_\mu X_\nu - X^\rho \partial_\mu g_{\nu\rho} = \partial_\mu X_\nu + \partial_\nu X_\mu + 2g_{\rho\sigma} \Gamma_{\mu\nu}^\sigma X^\rho$$

Therefore:

$$\delta g_{\mu\nu} = \nabla_\mu X_\nu + \nabla_\nu X_\mu \quad (4.1.12)$$

This equation can be used in the path integral. In fact, insisting that $\delta\mathcal{S} = 0$ for any $\delta g_{\mu\nu}$ gives the equations of motion; on the other hand, those variations $\delta g_{\mu\nu}$ for which $\delta\mathcal{S} = 0$ for any metric are the *symmetries* of the action. From Prop. 4.1.1:

$$\delta\mathcal{S} = \int d^4x \sqrt{-g} G^{\mu\nu} \delta g_{\mu\nu} = 2 \int d^4x \sqrt{-g} G^{\mu\nu} \nabla_\mu X_\nu$$

Invariance for $x^\mu \mapsto x^\mu - X^\mu$ means that $\delta\mathcal{S} = 0$ for all $X \in \mathfrak{X}(\mathcal{M})$, hence, integrating by parts:

$$\nabla_\mu G^{\mu\nu} = 0 \quad (4.1.13)$$

This is the Bianchi identity: although it is a result from Differential Geometry, it follows from diffeomorphism invariance of the Einstein-Hilbert action. This identity contains 4 equations, which make the 10 Einstein equations not completely independent: in fact, only 6 of them are independent, the same number of independent components of the metric, thus ensuring that the field equations are not overdetermined.

§4.2 Simple solutions

The Einstein equations in the absence of matter are $R_{\mu\nu} = \Lambda g_{\mu\nu}$, with $\Lambda \in \mathbb{R}$.

§4.2.1 Minkowski space

The simplest case is $\Lambda = 0$. The Einstein equations then reduce to $R_{\mu\nu} = 0$, with the condition of $g_{\mu\nu}$ being non-degenerate, since it is a metric and the field equations require the existence of the inverse metric $g^{\mu\nu}$. This restriction is physically unusual: it is not a holonomic constraint on the physical degrees of freedom, but an inequality $\det g < 0$ (with required signature $(-, +, +, +)$) which is not found for other fields of the Standard Model.

The simplest Ricci flat metric is Minkowski space $ds^2 = -dt^2 + dx^2$, but it is not the only metric obeying $R_{\mu\nu} = 0$. Another example is Schwarzschild metric.

§4.2.2 de Sitter space

Consider $\Lambda > 0$. A possible ansatz is:

$$ds^2 = -f(r)^2 dt^2 + f(r)^{-2} dr^2 + r^2(d\vartheta^2 + \sin^2 \vartheta d\varphi^2)$$

§4.2.2.1 Riemann tensor

First, one needs to compute the Ricci tensor for this metric: a simple way is using the curvature form. The non-coordinate 1-forms satisfy $ds^2 = \eta_{ab}\hat{\theta}^a \otimes \hat{\theta}^b$, thus:

$$\hat{\theta}^0 = f dt \quad \hat{\theta}^1 = f^{-1} dr \quad \hat{\theta}^2 = r d\vartheta \quad \hat{\theta}^3 = r \sin \vartheta d\varphi$$

$$d\hat{\theta}^0 = f' dr \wedge dt \quad d\hat{\theta}^1 = 0 \quad d\hat{\theta}^2 = dr \wedge d\vartheta \quad d\hat{\theta}^3 = \sin \vartheta dr \wedge d\varphi + r \cos \vartheta d\vartheta \wedge d\varphi$$

The first Cartan structure relation Eq. 2.4.18, together with Eq. 2.4.19, allow to determine the connection 1-form. For example, the first equation is $\omega_1^0 = f' f dt = f' d\hat{\theta}^0$, and then $\omega_0^1 = \omega_{10} = -\omega_{01} = \omega_1^0$. Using the other structure equation, the only non-vanishing components of the connection 1-form are:

$$\omega_1^0 = \omega_0^1 = f' \hat{\theta}^0 \quad \omega_1^2 = -\omega_2^1 = \frac{f}{r} \hat{\theta}^2 \quad \omega_1^3 = -\omega_3^1 = \frac{f}{r} \hat{\theta}^3 \quad \omega_2^3 = -\omega_3^2 = \frac{\cot \theta}{r} \hat{\theta}^3$$

The curvature 2-form can be computed from the second Cartan structure relation Eq. 2.4.21. For example, $\mathcal{R}_1^0 = d\omega_1^0 + \omega_c^0 \wedge \omega_1^c$, but $\omega_c^0 \wedge \omega_1^c = \omega_1^0 \wedge \omega_1^c = 0$, thus $\mathcal{R}_1^0 = d\omega_1^0 = ((f')^2 + f'' f) dr \wedge dt$. From the curvature 2-form, the Riemann tensor can be computed via Eq. 2.4.20 (recall the anti-symmetries of the Riemann tensor), finding the only non-vanishing components:

$$R_{0101} = ff'' + (f')^2 \quad R_{0202} = R_{0303} = \frac{ff'}{r} \quad R_{1212} = R_{1313} = -\frac{ff'}{r} \quad R_{2323} = \frac{1-f^2}{r^2}$$

To convert back to $x^\mu = (t, r, \vartheta, \varphi)$, use $R_{\mu\nu\rho\sigma} = e_a^\mu e_b^\nu e_c^\rho e_d^\sigma R_{abcd}$, which is particularly easy given that the matrices e_a^μ which define the non-coordinate 1-forms are diagonal:

$$R_{trtr} = f(r)f''(r) + f'(r)^2 \quad R_{t\vartheta t\vartheta} = f(r)^3 f'(r)r \quad R_{t\varphi t\varphi} = f(r)^3 f'(r)r \sin^2 \vartheta$$

$$R_{r\vartheta r\vartheta} = -\frac{f'(r)r}{f(r)} \quad R_{r\varphi r\varphi} = -\frac{f'(r)r}{f(r)} \sin^2 \vartheta \quad R_{\vartheta\varphi\vartheta\varphi} = (1-f(r)^2)r^2 \sin^2 \vartheta$$

§4.2.2.2 Ricci tensor

Given the Riemann tensor, it is easy to check that the Ricci tensor is diagonal with components:

$$R_{tt} = -f(r)^4 R_{rr} = f(r)^3 \left[f''(r) + \frac{2f'(r)}{r} + \frac{f'(r)^2}{f(r)} \right]$$

$$R_{\varphi\varphi} = \sin^2 \vartheta R_{\vartheta\vartheta} = (1 - f(r)^2 - 2f(r)f'(r)r) \sin^2 \vartheta$$

Imposing $R_{\mu\nu} = \Lambda g_{\mu\nu}$ determines two constraints on $f(r)$:

$$f''(r) + \frac{2f'(r)}{r} + \frac{f'(r)^2}{f(r)} = -\frac{\Lambda}{f(r)} \quad 1 - 2f(r)f'(r)r - f(r)^2 = \Lambda r^2$$

A solution is:

$$f(r) = \sqrt{1 - \frac{r^2}{R^2}} \quad R^2 \equiv \frac{3}{\Lambda}$$

This determines the metric of **de Sitter space**:

$$ds^2 = -\left(1 - \frac{r^2}{R^2}\right) dt^2 + \left(1 - \frac{r^2}{R^2}\right)^{-1} dr^2 + r^2(d\vartheta^2 + \sin^2 \vartheta d\varphi^2) \quad (4.2.1)$$

More precisely, this is the *static patch* of de Sitter space.

§4.2.2.3 de Sitter geodesics

To interpret a metric, it's useful to study its geodesics. First, note that a non-trivial $g_{tt}(r)$ term means that a particle won't remain at rest at $r \neq 0$, but it will be pushed to smaller values of $g_{tt}(r)$, i.e. larger values of r . The action of a particle in the general $f(r)$ metric, parametrized by its proper time σ , is:

$$\mathcal{S} = \int d\sigma \left[-f(r)^2 \dot{t}^2 + f(r)^{-2} \dot{r}^2 + r^2(\dot{\vartheta}^2 + \sin^2 \vartheta \dot{\varphi}^2) \right] \quad (4.2.2)$$

where $\dot{x}^\mu \equiv \frac{dx^\mu}{d\sigma}$. This Lagrangian has two ignorable degrees of freedom which lead to conserved quantities: $t(\sigma)$ and $\varphi(\sigma)$, as they appear only with time derivatives. The first one has energy as the conserved quantity, while the second has angular momentum:

$$E = -\frac{1}{2} \frac{\partial L}{\partial \dot{t}} = f(r)^2 \dot{t} \quad (4.2.3)$$

$$\ell = \frac{1}{2} \frac{\partial L}{\partial \dot{\varphi}} = r^2 \sin^2 \vartheta \dot{\varphi} \quad (4.2.4)$$

The $\frac{1}{2}$ are due to the absence of the usual factor in the kinetic term. The equations of motion from the action Eq. 4.2.2 should be supplemented with a constraint to distinguish whether a particle is massive or massless. For a massive particle, the trajectory must be timelike, so:

$$-f(r)^2 \dot{t}^2 + f(r)^{-2} \dot{r}^2 + r^2(\dot{\vartheta}^2 + \sin^2 \vartheta \dot{\varphi}^2) = -1$$

WLOG consider geodesics that lie in the $\vartheta = \frac{\pi}{2}$ plane, so that $\dot{\vartheta} = 0$ and $\sin^2 \vartheta = 1$. Replacing (t, φ) with (E, ℓ) , the constraint becomes:

$$\dot{r}^2 + V_{\text{eff}}(r) = E^2 \quad V_{\text{eff}}(r) = \left(1 + \frac{\ell^2}{r^2}\right) f(r)^2 \quad (4.2.5)$$

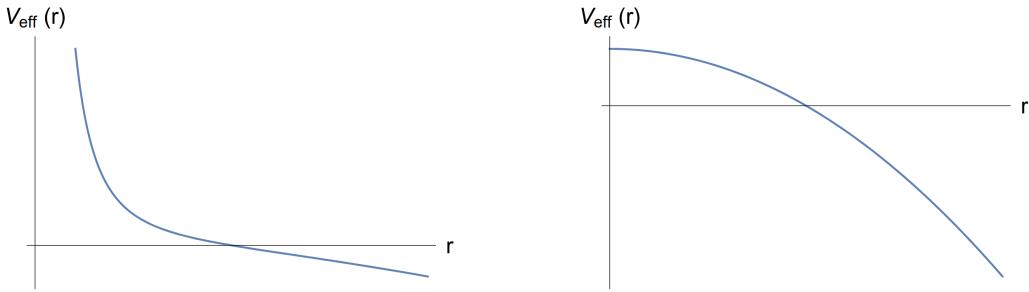


Figure 4.1: Effective potential for de Sitter geodesics, with $\ell \neq 0$ and $\ell = 0$ respectively.

For de Sitter space:

$$V_{\text{eff}}(r) = \left(1 + \frac{\ell^2}{r^2}\right) \left(1 - \frac{r^2}{R^2}\right) \quad (4.2.6)$$

This effective potential is plotted in Fig. 4.1: immediately, one sees that the potential pushes the particle to larger values of r . Focusing on geodesics with $\ell = 0$, the potential is a harmonic repulsor: a particle stationary at $r = 0$ is an unstable geodesic, since if it has non-zero initial velocity it will follow the trajectory:

$$r(\sigma) = R\sqrt{E^2 - 1} \sinh \frac{\sigma}{R} \quad (4.2.7)$$

Note that the metric is singular at $r = R$, a fact not manifest in the geodesic Eq. 4.2.7 which shows that any observer reaches $r = R$ in finite proper time. Problems arise when studying the coordinate time t , which has the interpretation of the time experienced by some stationary at $r = 0$; from Eq. 4.2.3:

$$\frac{dt}{d\sigma} = E \left(1 - \frac{r^2}{R^2}\right)^{-1}$$

This shows that $t(\sigma) \rightarrow \infty$ as $r(\sigma) \rightarrow R$: in fact, if $r(\sigma^*) = R$, then for $\sigma = \sigma^* - \varepsilon$ one has $\frac{dt}{d\sigma} = -\frac{\alpha}{\varepsilon}$, i.e. $t(\varepsilon) = -\log(\varepsilon/R)$, so $t(\varepsilon) \rightarrow \infty$ as $\varepsilon \rightarrow 0$. This means that a particle moving along the geodesic Eq. 4.2.7 will reach $r = R$ in finite proper time, but a stationary observer at $r = 0$ will measure an infinite amount of time.

This strange behaviour is similar to what happens at the horizon of a black hole ($r = 2GM$): however, the Schwarzschild metric has a singularity at $r = 0$, while de Sitter metric looks flat at $r = 0$. de Sitter space seems like an inverted black hole in which particles are pushed outwards to $r = R$.

§4.2.2.4 de Sitter embeddings

de Sitter space can be nicely embedded as a submanifold of $\mathbb{R}^{1,4}$ with metric:

$$ds^2 = -(dX^0)^2 + \sum_{i=1}^4 (dX^i)^2 \quad (4.2.8)$$

In particular, de Sitter metric Eq. 4.2.1 is a metric on the submanifold of $\mathbb{R}^{1,4}$ defined by the timelike hyperboloid:

$$-(X^0)^2 + \sum_{i=1}^4 (X^i)^2 = R^2 \quad (4.2.9)$$

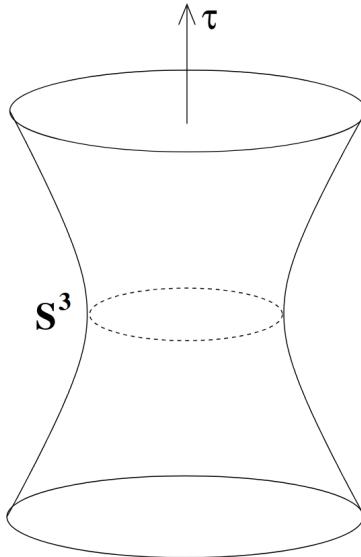


Figure 4.2: de Sitter space visualization with global coordinates.

A way of parametrizing this constraint is by imposing that $r^2 = (X^1)^2 + (X^2)^2 + (X^3)^2$, so that:

$$R^2 - r^2 = (X^4)^2 - (X^0)^2$$

The solutions are parametrized as:

$$X^0 = \sqrt{R^2 - r^2} \sinh \frac{t}{R} \quad X^4 = \sqrt{R^2 - r^2} \cosh \frac{t}{R}$$

Computing the respective variations, along with $\sum_{i=1}^3 (dX^i)^2 = dr^2 + r^2 d\Omega_2^2$, where $d\Omega_n^2$ is the metric element on S^n , allows to show that the pull-back of the Minkowski metric Eq. 4.2.8 on the hyperboloid Eq. 4.2.9 so parametrized gives the de Sitter metric Eq. 4.2.1 in the static patch coordinates.

The coordinates $\{X^i\}_{i=0,\dots,4}$ so defined are not the most intuitive: they single out X^4 as special, when the constraint does no such thing, and they do not cover the whole hyperboloid, as they are limited to $X^4 \geq 0$. A better choice of parametrization is:

$$X^0 = R \sinh \frac{\tau}{R} \quad X^i = y^i R \cosh \frac{\tau}{R} : \sum_{i=1}^4 (y^i)^2 = 1$$

Given this constraint, $\{y^i\}_{i=1,2,3,4}$ parametrize S^3 . These coordinates retain more of the symmetry of de Sitter space and cover the whole space, thus are a better parametrization. The pull-back of Minkowski metric, however, gives a rather different metric on de Sitter space:

$$ds^2 = -d\tau^2 + R^2 \cosh^2 \frac{\tau}{R} d\Omega_3^2 \quad (4.2.10)$$

These are known as *global coordinates*, as they cover the whole space (except for usual singularities at the poles for any choice of coordinates on S^3). Since this metric is related to that in Eq. 4.2.1 by a change of coordinates, it too must obey the Einstein equations. Global coordinates also show that the singularity which happens at $X^4 = 0$, i.e. at $r = R$, is nothing but a coordinate singularity.

These coordinates provide a clearer intuition for the physics of de Sitter space: it is a time-dependent solution of the field equations in which a spatial S^3 first shrinks to a minimal radius R and then expands, as shown in Fig. 4.2. The expansionary phase is a good approximation to our current universe on large scales.

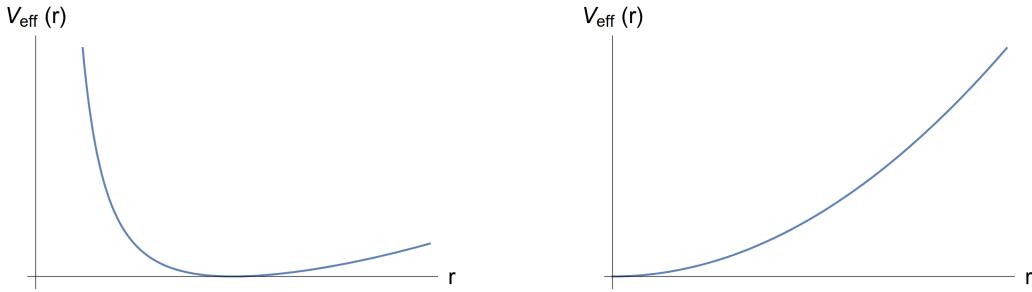


Figure 4.3: Effective potential for anti-de Sitter geodesics, with $\ell \neq 0$ and $\ell = 0$ respectively.

§4.2.3 Anti-de Sitter space

Consider $\Lambda < 0$. With the same ansatz as for de Sitter space, it is easy to find the metric for **anti-de Sitter space** (AdS):

$$ds^2 = -\left(1 + \frac{r^2}{R^2}\right) dt^2 + \left(1 + \frac{r^2}{R^2}\right)^{-1} dr^2 + r^2 d\Omega_2^2 \quad R^2 \equiv -\frac{3}{\Lambda} \quad (4.2.11)$$

Equivalent coordinates are found setting $r = R \sinh \rho$:

$$ds^2 = -\cosh^2 \rho dt^2 + R^2 d\rho^2 + R^2 \sinh^2 \rho d\Omega_2^2 \quad (4.2.12)$$

In AdS there is no coordinate singularity in r , and indeed coordinates cover the whole space.

§4.2.3.1 Anti-de Sitter geodesics

AdS has the same action as in Eq. 4.2.2, thus the radial trajectory of a massive particle moving along a geodesic in the $\vartheta = \frac{\pi}{2}$ plane is $\dot{r}^2 + V_{\text{eff}}(r) = E^2$, with:

$$V_{\text{eff}}(r) = \left(1 + \frac{\ell^2}{r^2}\right) \left(1 + \frac{r^2}{R^2}\right) \quad (4.2.13)$$

Plotting it in Fig. 4.3, the geodesics' behaviour is clear: with no angular momentum, anti-de Sitter space acts like a harmonic oscillator, pulling the particle towards the origin and making it oscillate around $r = 0$. If the particle has non-zero angular momentum, then the potential has a minimum at $r_*^2 = R\ell$, thus particles oscillate around r_* : importantly, particles with finite energy cannot escape to $r \rightarrow \infty$, but are constrained by the spacetime within some finite distance from the origin.

This picture of AdS as a harmonic trap which pulls particle to the origin clashes with the fact that AdS is a homogeneous space (roughly, all points are the same). To understand how these two facts are compatible, consider a stationary observer at $r = 0$: this is a geodesic and, from its perspective, other observers (with $\ell = 0$) will oscillate around $r = 0$ along geodesics. However, since these observers move along geodesics, in their reference frame they are stationary at $r = 0$, with all other observers oscillating. Thus, while in dS each observer views themselves at the center of the universe, with other observers moving away from them, in AdS each observer views themselves as the center of the universe, with other observer oscillating around them.

It's possible to study geodesics for a massless particle too. This time, the constraint to the action is:

$$-f(r)^2 t^2 + f(r)^{-2} \dot{r}^2 + r^2 (\dot{\vartheta}^2 + \sin^2 \vartheta \dot{\varphi}^2) = 0$$

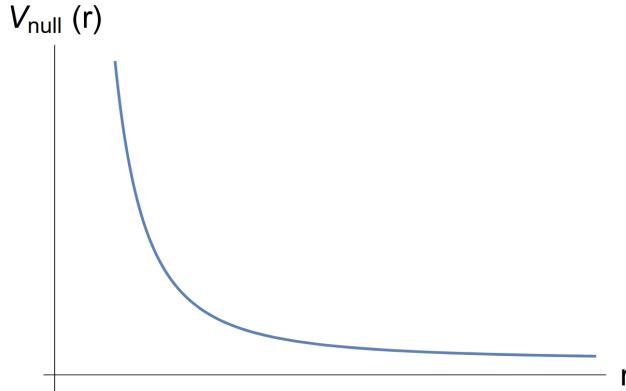


Figure 4.4: Potential experienced by massless particles with \$\ell \neq 0\$ in AdS.

This means that the particle follows a null geodesic. Its equation of motion is:

$$\dot{r}^2 + V_{\text{null}}(r) = E^2 \quad V_{\text{null}}(r) = \frac{\ell^2}{r^2} \left(1 + \frac{r^2}{R^2} \right)$$

Its plot in Fig. 4.4 makes it clear that any massless particle can escape to \$r \rightarrow \infty\$, as the null potential is asymptotically constant, and it will experience the usual gravitational redshift: AdS only confines massive particles. To solve for the trajectory, it's easier to work with \$r = R \sinh \rho\$ and \$\ell = 0\$, so that:

$$R \dot{\rho} = \pm \frac{E}{\cosh \rho} \quad \Rightarrow \quad R \sinh \rho(\sigma) = E(\sigma - \sigma_0)$$

One sees that \$\rho \rightarrow \infty\$ only in infinite affine time (i.e. \$\sigma \rightarrow \infty\$). However, in coordinate time, recalling Eq. 4.2.3, \$E = \cosh^2 \rho \dot{t}\$, so:

$$R \tan \frac{t}{R} = E(\sigma - \sigma_0)$$

Hence, as affine time \$\sigma \rightarrow \infty\$, coordinate time \$t \rightarrow \frac{\pi}{2}R\$. This means that light rays escape to infinity in a finite amount of coordinate time: to make sense of dynamics in AdS, one must specify some boundary conditions at infinity to dictate how massless particles and field behave. AdS does not appear to be of any cosmological interest.

§4.2.3.2 Anti-de Sitter embeddings

AdS too has a natural embedding in a 5-dimensional spacetime: it is a submanifold of \$\mathbb{R}^{2,3}\$ with metric:

$$ds^2 = -(dX^0)^2 - (dx^4)^2 + \sum_{i=1}^3 (dX^i)^2 \tag{4.2.14}$$

In particular, anti-de Sitter metric Eq. 4.2.11 is a metric on the hyperboloid:

$$-(X^0)^2 - (X^4)^2 + \sum_{i=1}^3 (X^i)^2 = -R^2 \tag{4.2.15}$$

This constraint can be solved via the parametrization:

$$X^0 = R \cosh \rho \sin \frac{t}{R} \quad X^4 = R \cosh \rho \cos \frac{t}{R} \quad X^i = y^i R \sinh \rho : \sum_{i=1}^3 (y^i)^2 = 1$$

Given this constraint, $\{y^i\}_{i=1,2,3}$ parametrize \mathbb{S}^2 . The pull-back of the metric Eq. 4.2.14 on the hyperboloid so parametrized gives anti-de Sitter metric Eq. 4.2.11.

Note a small subtlety: the embedding hyperboloid has topology $\mathbb{S}^1 \times \mathbb{R}^3$, not \mathbb{R}^4 : this corresponds to a compact time direction, as $t \in [0, 2\pi R)$. However, AdS metric does not have such restriction, with $t \in \mathbb{R}$: this is a universal covering of the hyperboloid.

Another useful parametrization of the hyperboloid is the following:

$$X^4 - X^3 = r \quad X^4 + X^3 = \frac{R^2}{r} + \frac{r}{R^2} \eta_{ij} x^i x^j \quad X^i = \frac{r}{R} x^i, i = 0, 1, 2$$

with $r \in [0, \infty)$. With these coordinates, the metric takes the form:

$$ds^2 = R^2 \frac{dr^2}{r^2} + \frac{r^2}{R^2} \eta_{ij} dx^i dx^j$$

These coordinates do not cover the whole AdS, but only one-half of the hyperboloid, restricted to $X^4 - X^3 > 0$: this is known as the *Poincaré patch* of AdS. Moreover, x^0 cannot be further extended beyond $x^0 \in (-\infty, +\infty)$, thus in global coordinates the restriction $t \in [0, 2\pi R)$ remains. In this case, massive particles fall towards $r = 0$.

Finally, there are two more possible coordinate systems on the Poincaré patch, obtained by setting $z = \frac{R^2}{r}$ and $r = Re^\rho$:

$$ds^2 = \frac{R^2}{z^2} (dz^2 + \eta_{ij} dx^i dx^j) \quad ds^2 = R^2 d\rho^2 + e^{2\rho} \eta_{ij} dx^i dx^j$$

In each case, massive particles fall towards $z = \infty$ or $\rho = -\infty$.

§4.3 Symmetries

What makes Minkowski, dS and AdS special solutions to the Einstein equations are their symmetries.

The symmetries of Minkowski spacetime are familiar: translations and rotations of spacetime, with the latter splitting between proper rotations and Lorentz boosts. These symmetries are responsible for the existence of energy, momentum and angular momentum on a fixed Minkowski background.

It is important to characterize the symmetries of a general metric.

§4.3.1 Isometries

Recall Def. 1.2.5 of flow on a manifold: by Eq. 1.2.9, it is possible to identify a flow with a vector field $K \in \mathfrak{X}(\mathcal{M})$ such that it is tangent to the flow at each point of the manifold: $K^\mu = \frac{dx^\mu}{dt}$, with $x^\mu(t) \equiv x^\mu(\sigma_t)$.

Definition 4.3.1 (Isometry)

Given a flow associated to $K \in \mathfrak{X}(\mathcal{M})$, it is said to be an **isometry** if:

$$\mathcal{L}_K g = 0 \iff \nabla_\mu K_\nu + \nabla_\nu K_\mu = 0 \quad (4.3.1)$$

Recall Eq. 4.1.11-4.1.12 for the equivalence. This condition means that the metric doesn't change along flow lines: this is called *Killing equation*, and a vector field which satisfied it is a **Killing vector field**. Eq. 1.2.17 can be generalized to hold for the Lie derivative of arbitrary tensor fields, thus Killing vectors too form a Lie algebra: it is the Lie algebra of the isometry group of the manifold.

Example 4.3.1 (Ignorable degrees of freedom)

Given a metric $g_{\mu\nu}(x)$, if it doesn't depend on $x^1 \equiv y$, then $X = \partial_y$ is a Killing vector field, for $(\mathcal{L}_{\partial_y} g)_{\mu\nu} = \partial_y g_{\mu\nu} = 0$. This is the case of ignorable degrees of freedom, as Eq. 4.2.3-4.2.4.

§4.3.1.1 Minkowski spacetime

Consider Minkowski spacetime $\mathbb{R}^{1,3}$ with $g_{\mu\nu} = \eta_{\mu\nu}$. Killing equation is:

$$\partial_\mu K_\nu + \partial_\nu K_\mu = 0$$

There are two forms of solution. The first one is:

$$K_\mu = c_\mu$$

for any constant vector c_μ . These generate translations. Alternatively:

$$K_\mu = \omega_{\mu\nu} x^\nu : \omega_{\mu\nu} = -\omega_{\nu\mu}$$

These generate rotations and Lorentz boosts (see §1.2 of [1]). The emergence of the Lie algebra structure can be elucidated defining the Killing vectors:

$$P_\mu := \partial_\mu \quad M_{\mu\nu} := \eta_{\mu\rho} x^\rho \partial_\nu - \eta_{\nu\rho} x^\rho \partial_\mu$$

These are 10 Killing vectors: 4 for translations and 6 for rotations and boosts.

Proposition 4.3.1 (Poincaré algebra)

Killing vectors of Minkowski spacetime obey:

$$[P_\mu, P_\nu] = 0 \quad [M_{\mu\nu}, P_\sigma] = -\eta_{\mu\sigma}P_\nu + \eta_{\sigma\nu}P_\mu$$

$$[M_{\mu\nu}, M_{\rho\sigma}] = \eta_{\mu\rho}M_{\nu\sigma} + \eta_{\nu\rho}M_{\mu\sigma} - \eta_{\mu\rho}M_{\nu\sigma} - \eta_{\nu\sigma}M_{\mu\rho}$$

Proof. By direct calculation. □

This is precisely the Lie algebra of the Poincaré group $\text{ISO}(1, 3) := \mathbb{R}^4 \rtimes \text{SO}(1, 3)$, i.e. the isometry group of Minkowski spacetime.

§4.3.1.2 de Sitter and anti-de Sitter space

The isometries of dS and AdS are best seen from their embeddings. The constraint Eq. 4.2.9 which defines de Sitter space is invariant under the rotations of $\mathbb{R}^{1,4}$, thus dS inherits the $\text{SO}(1, 4)$ isometry group. Similarly, the constraint Eq. 4.2.15 which defines anti-de Sitter space is invariant under rotations of $\mathbb{R}^{2,3}$, so AdS inherits the $\text{SO}(2, 3)$ isometry group. Note that both these groups are 10-dimensional, as $\mathbb{R}^4 \rtimes \text{SO}(1, 3)$.

It is simple to determine the 10 Killing spinors of 5-dimensional spacetime:

$$M_{AB} = \eta_{AC}X^C\partial_B - \eta_{BC}X^C\partial_A$$

where $X^A, A = 0, 1, 2, 3, 4$ are 5-dimensional coordinates and η_{AB} is the appropriate Minkowski metric, with signature $(-, +, +, +, +)$ for dS and $(-, -, +, +, +)$ for AdS. In either case, the Lie algebra is that of the appropriate Lorentz group:

$$[M_{AB}, M_{CD}] = \eta_{AD}M_{BC} + \eta_{BC}M_{AD} - \eta_{AC}M_{BD} - \eta_{BD}M_{AC}$$

The embedding hyperbolae are both invariant under these Killing vectors: flows generated by them map points on the hyperbolae to other points on the hyperbolae. Therefore, these Killing vectors are inherited by dS and AdS respectively.

Energy Energy is defined by timelike Killing vectors. In anti-de Sitter space there is no problem finding a timelike Killing vector, as the metric in global coordinates Eq. 4.2.11 is time-independent, so $K = \partial_t$. However, this is not the case in de Sitter space.

Considering dS in the static patch with coordinates $r^2 = (X^1)^2 + (X^2)^2 + (X^3)^2$, $X^0 = \sqrt{R^2 - r^2}\sinh \frac{t}{R}$ and $X^4 = \sqrt{R^2 - r^2}\cosh \frac{t}{R}$, the metric Eq. 4.2.1 is time-independent, thus $K = \partial_t$ is a Killing vector; pushing it forward to the 5-dimensional space:

$$\partial_t = \frac{\partial X^A}{\partial t}\partial_A = \frac{1}{R}(X^4\partial_0 + X^0\partial_4)$$

On the static patch, this Killing vector is timelike and the energy Eq. 4.2.3 follows from it. The problem is that the static patch is only one-half of the hyperboloid: when considering the whole AdS, one must account for the case $X^0 = 0, X^4 < 0$, in which the Killing vector points in the past direction, or $X^0 \neq 0, X^4 = 0$, in which it is spacelike. Therefore, the Killing vector can be both timelike (future-directed and past-directed) or spacelike when spanning the whole manifold: for this reason, it's not possible to define a global positive conserved energy on the total de Sitter space. The same conclusion follows by noting that the metric in global coordinates Eq. 4.2.10 is time-dependent.

§4.3.2 Conserved quantities

It is possible to reframe Noether's theorem in the context of Killing vectors.

Theorem 4.3.1 (Noether's theorem)

Given a massive particle moving on a geodesic $x^\mu(\tau)$ in a spacetime with metric g which admits a Killing vector field K , then the **Noether charge** is conserved along the geodesic:

$$Q := K_\mu \frac{dx^\mu}{d\tau} \quad (4.3.2)$$

Proof. First, show that Q is indeed conserved along the geodesic, recalling Eq. 4.3.1:

$$\frac{dQ}{d\tau} = \partial_\nu K_\mu \frac{dx^\nu}{d\tau} \frac{dx^\mu}{d\tau} + K_\mu \frac{d^2 x^\mu}{d\tau^2} = \partial_\nu K_\mu \frac{dx^\nu}{d\tau} \frac{dx^\mu}{d\tau} - K_\mu \Gamma_{\rho\sigma}^\mu \frac{dx^\rho}{d\tau} \frac{dx^\sigma}{d\tau} = \nabla_\nu K_\mu \frac{dx^\nu}{d\tau} \frac{dx^\mu}{d\tau} = 0$$

To show that Q follows from Noether's theorem, consider the action of the massive particle and introduce an infinitesimal transformation $\delta x^\mu = K^\mu(x)$:

$$\begin{aligned} \delta S &= \delta \int d\tau g_{\mu\nu}(x) \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = \int d\tau \left[\partial_\rho g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} \delta x^\rho + 2g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{d\delta x^\nu}{d\tau} \right] \\ &= \int d\tau \left[\partial_\rho g_{\mu\nu} \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} K^\rho + 2 \frac{dx^\mu}{d\tau} \left(\frac{dK_\mu}{d\tau} - K^\nu \frac{dg_{\mu\nu}}{d\tau} \right) \right] \\ &= \int d\tau [\partial_\rho g_{\mu\nu} K^\rho - 2K^\rho \partial_\nu g_{\mu\rho} + 2\partial_\nu K_\mu] \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} = \int d\tau 2\nabla_\nu K_\mu \frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau} \end{aligned}$$

The transformation is a symmetry if $\delta S = 0$, thus, by the symmetry of $\frac{dx^\mu}{d\tau} \frac{dx^\nu}{d\tau}$, the resulting equation is $\nabla_\nu K_\mu = 0$, i.e. Killing equation. \square

Example 4.3.2

Energy and angular momentum defined on de Sitter geodesics by Eq. 4.2.3-4.2.4 are Noether charges associated to Killing vectors ∂_t and ∂_φ respectively.

§4.3.3 Komar integrals

Definition 4.3.2 (Komar form)

Given a Killing vector $K = K^\mu \partial_\mu$ and the 1-form $K^\flat \equiv K_\mu dx^\mu$, the associated **Komar form** is the 2-form defined as:

$$F := dK^\flat \quad (4.3.3)$$

Proposition 4.3.2 (Komar forms and Killing vectors)

Given a Killing vector $K = K^\mu \partial_\mu$, the associated Komar form is:

$$F_{\mu\nu} = \nabla_\mu K_\nu - \nabla_\nu K_\mu \quad (4.3.4)$$

Proof. $F = \frac{1}{2}F_{\mu\nu}dx^\mu \wedge dx^\nu = dK = \nabla_\mu K_\nu dx^\mu \wedge dx^\nu$. □

Theorem 4.3.2 (Generalized Maxwell equations)

If the vacuum Einstein equations $R_{\mu\nu} = 0$ hold, then a Komar form obeys the vacuum Maxwell equations:

$$d \star F = 0 \iff \nabla^\mu F_{\mu\nu} = 0 \quad (4.3.5)$$

Proof. Recall Prop. 2.2.6 for the equivalence. From Ricci identity Eq. 2.2.15:

$$(\nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu)K^\sigma = R^\sigma_{\rho\mu\nu}K^\rho \implies (\nabla_\mu \nabla_\nu - \nabla_\nu \nabla_\mu)K^\mu = R_{\rho\nu}K^\rho$$

From Killing equation $\nabla_{(\mu} K_{\nu)} = 0 \implies \nabla_\mu K^\mu = 0$, thus $\nabla_\mu \nabla_\nu K^\mu = R_{\rho\nu}K^\rho$ and:

$$\nabla^\mu F_{\mu\nu} = \nabla^\mu \nabla_\mu K_\nu - \nabla^\mu \nabla_\nu K_\mu = -2\nabla^\mu \nabla_\nu K_\mu = -2R_{\rho\nu}K^\rho$$

If $R_{\rho\nu} = 0$, then $\nabla^\mu F_{\mu\nu} = 0$. □

Definition 4.3.3 (Komar integral)

Given a Komar form F associated to a Killing vector K , the associated **Komar charge** on a spatial submanifold Σ is defined as:

$$Q_{\text{Komar}} := -\frac{1}{8\pi G} \int_{\Sigma} d \star F = -\frac{1}{8\pi G} \int_{\partial\Sigma} \star F = -\frac{1}{8\pi G} \int_{\partial\Sigma} \star dK^\flat \quad (4.3.6)$$

Proposition 4.3.3

If Eq. 4.3.5 holds, then Q_{Komar} is conserved.

Proof. Recall Eq. 2.2.27: the proof is identical. □

As for Noether integrals of a particle, Komar integrals of spacetime are interpreted based on the defining Killing vector. For example, if K^μ is everywhere timelike, i.e. $g_{\mu\nu}K^\mu K^\nu < 0$, then its Komar integral can be identified with the energy (or the mass) of spacetime; if the Killing vector is related to rotations, instead, the conserved charge is the angular momentum of spacetime.

§4.4 Asymptotics of spacetime

The three special solution (Minkowski, dS, AdS) not only have different spacetime curvature and symmetries, but, more fundamentally, they have different behaviour at infinity. Their importance lies in the fact that, however complicated the metric might be, if fields are suitably localized, then they will asymptote to one of these three symmetric spaces.

§4.4.1 Conformal transformations

Definition 4.4.1 (Conformal transformation)

Given a metric manifold (\mathcal{M}, g) and a non-vanishing $\Omega \in \mathcal{C}^\infty(\mathcal{M})$, a **conformal transformation** is defined as:

$$\tilde{g}_{\mu\nu}(x) = \Omega^2(x)g_{\mu\nu}(x) \quad (4.4.1)$$

Typically, g and \tilde{g} describe different spacetime with considerably warped distances. However, conformal transformations preserve angles: in Lorentzian spacetime, the two metrics have the same causal structure, i.e. null/timelike/spacelike vector fields in one metric remain null/timelike/spacelike in the other too.

Proposition 4.4.1

Only conformal transformations of the metric preserve its causal structure.

Although timelike particle trajectories necessarily remain timelike under a conformal transformation, the same needs not be true for geodesics, as distances get warped.

Proposition 4.4.2 (Invariance of null geodesics)

Null geodesics are mapped to null geodesics under a conformal transformation.

Proof. First compute the Christoffel symbols in the new metric:

$$\begin{aligned} \Gamma_{\rho\sigma}^\mu[\tilde{g}] &= \frac{1}{2}\tilde{g}^{\mu\nu}(\partial_\rho\tilde{g}_{\nu\sigma} + \partial_\sigma\tilde{g}_{\rho\nu} - \partial_\nu\tilde{g}_{\rho\sigma}) \\ &= \frac{1}{2}\Omega^{-2}g^{\mu\nu}(\partial_\rho(\Omega^2g_{\nu\sigma}) + \partial_\sigma(\Omega^2g_{\rho\nu}) - \partial_\nu(\Omega^2g_{\rho\sigma})) \\ &= \Gamma_{\rho\sigma}^\mu[g] + \Omega^{-1}(\delta_\sigma^\mu\nabla_\rho\Omega + \delta_\rho^\mu\nabla_\sigma\Omega - g_{\rho\sigma}\nabla^\mu\Omega) \end{aligned}$$

In the last line, recall that $\nabla = \partial$ on scalar functions. Suppose an affinely parametrized geodesic in the metric g :

$$\begin{aligned} \frac{d^2x^\mu}{d\tau^2} + \Gamma_{\rho\sigma}^\mu[g]\frac{dx^\rho}{d\tau}\frac{dx^\sigma}{d\tau} &= 0 \\ \implies \frac{d^2x^\mu}{d\tau^2} + \Gamma_{\rho\sigma}^\mu[\tilde{g}]\frac{dx^\rho}{d\tau}\frac{dx^\sigma}{d\tau} &= \Omega^{-1}(\delta_\sigma^\mu\nabla_\rho\Omega + \delta_\rho^\mu\nabla_\sigma\Omega - g_{\rho\sigma}\nabla^\mu\Omega)\frac{dx^\rho}{d\tau}\frac{dx^\sigma}{d\tau} \end{aligned}$$

For a null geodesic $g_{\rho\sigma}\frac{dx^\rho}{d\tau}\frac{dx^\sigma}{d\tau} = 0$, thus:

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma_{\rho\sigma}^\mu[\tilde{g}]\frac{dx^\rho}{d\tau}\frac{dx^\sigma}{d\tau} = 2\frac{dx^\mu}{d\tau}\frac{1}{\Omega}\frac{d\Omega}{d\tau}$$

This is the equation of non-affinely parametrized geodesic, hence the thesis. \square

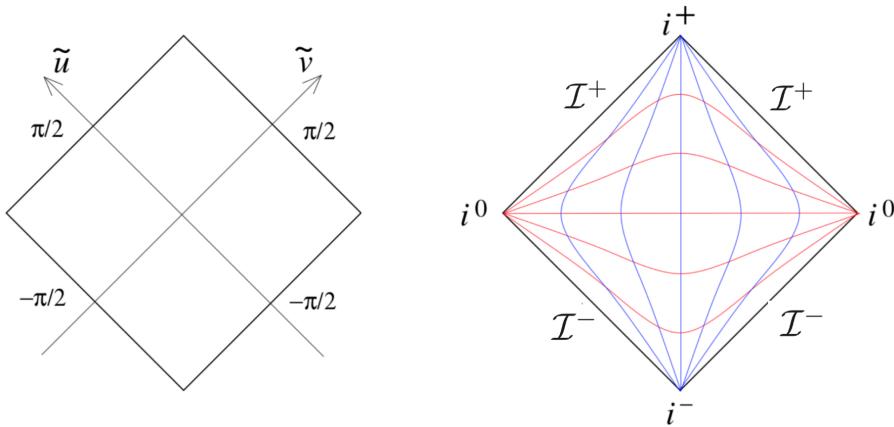


Figure 4.5: Penrose diagram for $d = 1 + 1$ Minkowski spacetime.

Usual curvature tensors are not invariant under conformal transformations. A curvature tensor that is indeed invariant is the **Weyl tensor**:

$$C_{\mu\nu\rho\sigma} := R_{\mu\nu\rho\sigma} - \frac{2}{n-2} (g_{\mu[\rho} R_{\sigma]\nu} - g_{\nu[\rho} R_{\sigma]\mu}) + \frac{2}{(n-1)(n-2)} R g_{\mu[\rho} g_{\sigma]\nu} \quad (4.4.2)$$

where $n \equiv \dim_{\mathbb{R}} \mathcal{M}$. The Weyl tensor has all the symmetries of the Riemann tensor, with the additional property that it vanishes when contracting any pair of indices with the metric: it can be viewed as the trace-free part of the Riemann tensor.

§4.4.2 Penrose diagrams

To study the asymptotic behaviour of spacetime, one needs to perform a conformal transformation that maps infinity to a finite distance: the resulting causal structure can then be drawn on a finite area and the resulting picture is called a *Penrose diagram*.

§4.4.2.1 Minkowski spacetime

It is simplest to study $\mathbb{R}^{1,1}$, where the Minkowski metric is $ds^2 = -dt^2 + dx^2$. First, introduce **lightcone coordinates**:

$$u = t - x \quad v = t + x \quad \Rightarrow \quad ds^2 = -du dv$$

These coordinates range in $u, v \in \mathbb{R}$. To work with finite quantities, define:

$$u = \tan \tilde{u} \quad v = \tan \tilde{v} \quad \Rightarrow \quad ds^2 = -\frac{1}{\cos^2 \tilde{u} \cos^2 \tilde{v}} d\tilde{u} d\tilde{v}$$

where now $\tilde{u}, \tilde{v} \in (-\frac{\pi}{2}, +\frac{\pi}{2})$. Note that the metric diverges when approaching the boundary of Minkowski spacetime. However, a conformal transformation is possible with $\Omega(\tilde{u}, \tilde{v}) \equiv \cos \tilde{u} \cos \tilde{v}$:

$$d\tilde{s}^2 = (\cos^2 \tilde{u} \cos^2 \tilde{v}) ds^2 = -d\tilde{u} d\tilde{v}$$

It is now possible to add $\tilde{u} = \pm \frac{\pi}{2}$ and $\tilde{v} = \pm \frac{\pi}{2}$: this operation is called **conformal compactification**.

Penrose diagrams, as other relativistic diagrams, present light-rays travelling at 45° , time in the vertical direction and space in the horizontal one. Fig. 4.5 shows the Penrose diagram for Minkowski spacetime $\mathbb{R}^{1,1}$ with \tilde{u}, \tilde{v} coordinates, both with coordinate axis and geodesics. Moreover, the various infinities of Minkowski space are shown:

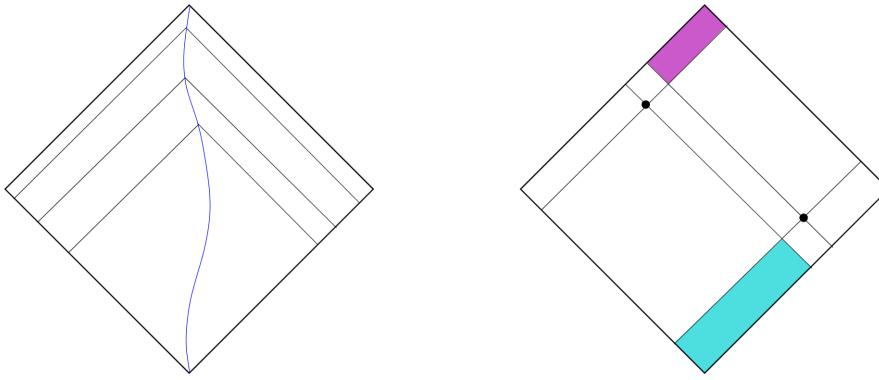


Figure 4.6: Lightcones in $d = 1 + 1$ Minkowski spacetime.

- timelike geodesics (blue) start at $i^-(\tilde{u}, \tilde{v}) = (-\frac{\pi}{2}, -\frac{\pi}{2})$ and end at $i^+(\tilde{u}, \tilde{v}) = (+\frac{\pi}{2}, +\frac{\pi}{2})$, called respectively **past** and **future timelike infinity**;
- spacelike geodesics (red) start and end at $i^0(\tilde{u}, \tilde{v}) = (\mp\frac{\pi}{2}, \pm\frac{\pi}{2})$, both called **spacelike infinity**;
- all null curves (not shown) start at the boundary $\mathcal{I}^- \equiv \{\tilde{u} = -\frac{\pi}{2}\} \cup \{\tilde{v} = -\frac{\pi}{2}\}$ and end at the boundary $\mathcal{I}^+ \equiv \{\tilde{u} = +\frac{\pi}{2}\} \cup \{\tilde{v} = +\frac{\pi}{2}\}$, called respectively **past** and **future null infinity**.

It is clear then that in Minkowski spacetime there are more ways reach infinity along a null direction than in a timelike or spacelike direction.

Penrose diagrams allow to immediately visualize the causal structure of spacetime. As shown in Fig. 4.6, given a particle moving along a timelike curve, as it approaches i^+ its past lightcone encompasses progressively more of spacetime: thus, an observer in Minkowski spacetime can in principle see everything, as long as they wait long enough. Relatedly, given any two points in Minkowski spacetime, they are causally connected both in the past and in the future, as both cones intersect (see Fig. 4.6): thus, there was always an event in the past that could have influenced both, and there will always be an event in the future that can be influenced by both.

4d Minkowski spacetime The analysis can be repeated for $\mathbb{R}^{1,3}$ with $ds^2 = -dt^2 + dr^2 + r^2 d\Omega_2^2$. Lightcone coordinates are again:

$$u = t - r \quad v = t + r \quad \Rightarrow \quad ds^2 = -du dv + \frac{1}{4} (u - v)^2 d\Omega_2^2$$

Finite range coordinates are again:

$$u = \tan \tilde{u} \quad v = \tan \tilde{v} \quad \Rightarrow \quad ds^2 = \frac{1}{4 \cos^2 \tilde{u} \cos^2 \tilde{v}} (-4d\tilde{u} d\tilde{v} + \sin^2(\tilde{u} - \tilde{v}) d\Omega_2^2)$$

Finally, the conformal transformation with $\Omega(\tilde{u}, \tilde{v}) = 2 \cos \tilde{u} \cos \tilde{v}$ leads to:

$$d\tilde{s}^2 = -4d\tilde{u} d\tilde{v} + \sin^2(\tilde{u} - \tilde{v}) d\Omega_2^2$$

In 4d Minkowski spacetime there is an additional constraint: $r \geq 0$, so $v \geq u$. Conformal compactification leads to:

$$-\frac{\pi}{2} \leq \tilde{u} \leq \tilde{v} \leq +\frac{\pi}{2}$$

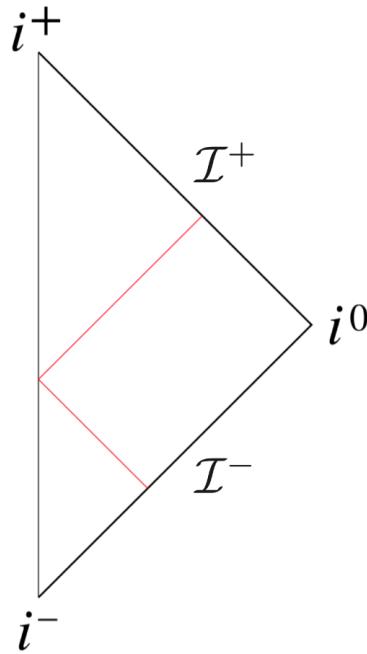


Figure 4.7: Penrose diagram for $d = 1 + 3$ Minkowski spacetime.

The corresponding Penrose diagram is drawn in Fig. 4.7: the spatial \mathbb{S}^2 is not shown for simplicity, but every point on the diagram corresponds to an \mathbb{S}^2 of radius $|\sin(\tilde{u} - \tilde{v})|$. The line $\tilde{u} = \tilde{v}$ is not a boundary of Minkowski spacetime, but it is simply the origin $r = 0$ (at which \mathbb{S}^2 shrinks to a point): to illustrate this, a null geodesic is drawn.

In general, Penrose diagrams are only useful for spacetimes which contain an obvious \mathbb{S}^2 , i.e. those with $\text{SO}(3)$ isometry: luckily, these are the simplest and most important in physics.

§4.4.2.2 de Sitter space

Recall global coordinates on dS: from Eq. 4.2.10, $ds^2 = -d\tau^2 + R^2 \cosh^2 \frac{\tau}{R} d\Omega_3^2$. To draw the Penrose diagram, define *conformal time* as:

$$\frac{d\eta}{d\tau} = \frac{1}{R \cosh(\tau/R)} \implies \cos \eta = \frac{1}{\cosh(\tau/R)}$$

Given that $\tau \in \mathbb{R}$, then $\eta \in (-\frac{\pi}{2}, +\frac{\pi}{2})$. In conformal time, de Sitter space has the metric:

$$ds^2 = \frac{R^2}{\cos^2 \eta} (-d\eta^2 + d\Omega_3^2)$$

Writing $d\Omega_3^2 = d\chi^2 + \sin^2 \chi d\Omega_2^2$, with $\chi \in [0, \pi]$, de Sitter metric is conformally equivalent to:

$$d\tilde{s}^2 = -d\eta^2 + d\chi^2 + \sin^2 \chi d\Omega_2^2$$

After conformal compactification, $\eta \in [-\frac{\pi}{2}, +\frac{\pi}{2}]$ and $\chi \in [0, \pi]$. The Penrose diagram is drawn in Fig. 4.8: the two vertical lines are not boundaries of dS, but simply the north and south poles of \mathbb{S}^3 . The boundaries of the spacetime are the top and bottom lines, labelled both i^\pm and \mathcal{I}^\pm as they are where both timelike and null geodesics originate and terminate.

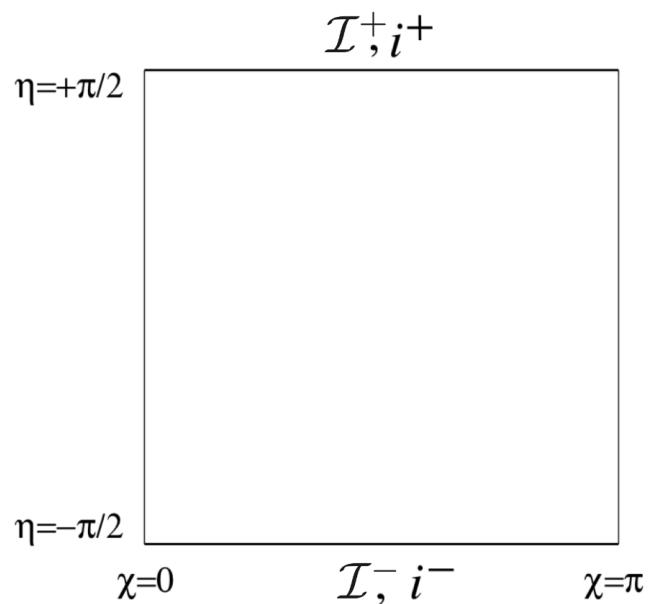


Figure 4.8: Penrose diagram for de Sitter space.

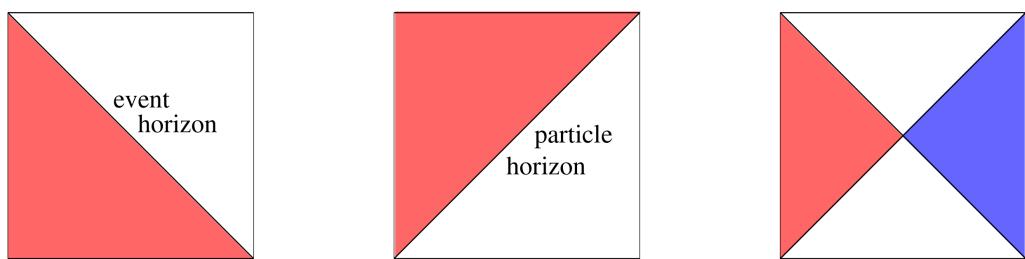


Figure 4.9: Event and particle horizons for an observer at the north pole in dS, and the causal diamonds for an observer at the north and south pole.

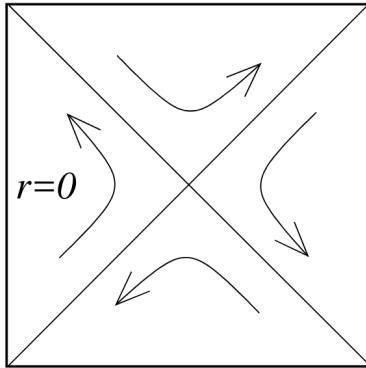


Figure 4.10: $K = \partial_t$ Killing vector field in de Sitter spacetime.

Proposition 4.4.3 (dS boundary)

de Sitter spacetime has a spacelike \mathbb{S}^3 boundary with timelike normal vector.

The causal structure of dS is very different from that of Minkowski spacetime: it is no longer true that any observer can see everything by waiting long enough. For example, as shown in Fig. 4.9, an observer at the north pole will eventually see only exactly half the spacetime: the boundary of this half-space is the observer's *event horizon*, in the sense that signals from beyond the horizon cannot reach them. It is also clear that this event horizon is observer-dependent: in this context, these are referred to as **cosmological horizons**.

Furthermore, an observer at the north pole will only be able to communicate with another half of the spacetime, as shown in Fig. 4.9: the boundary of this region of influence is known as the *particle horizon* and it represents the furthest distance light can travel since the beginning of time. Its intersection with the event horizon determines the (nothern) **causal diamond**: usually, nothern and southern causal diamonds are causally disconnected.

Penrose diagrams can also be used to explain the divergence at $r = R$ of the metric Eq. 4.2.1 on the static patch of dS. Recalling the embedding of the static patch in $\mathbb{R}^{1,4}$, parametrized as $X^0 = \sqrt{R^2 - r^2} \sinh \frac{t}{R}$ and $X^4 = \sqrt{R^2 - r^2} \cosh \frac{t}{R}$. Naively, the surface $r = R$ corresponds to $X^0 = X^4 = 0$, but writing $r = R(1 - \varepsilon^2/2)$ yields that $X^0 \sim R\varepsilon \sinh \frac{t}{R}$ and $X^4 \sim R\varepsilon \cosh \frac{t}{R}$, so $\varepsilon \rightarrow 0$ can be obtained keeping $X^0, X^4 \neq 0$ and finite, provided that $t \rightarrow \pm\infty$: to do this, $\varepsilon \exp(\pm t/R)$ must be kept finite, thus the surface $r = R$ is identified with the lines $X^0 = \pm X^4$. Translation to polar coordinates is done by $X^0 = R \sinh \frac{\tau}{R}$ and $X^4 = R \cosh \frac{\tau}{R} \cos \chi$, with χ the polar angle on \mathbb{S}^3 ; after mapping to conformal time, one finds that:

$$X^0 = \pm X^4 \iff \sin \eta = \pm \cos \chi \iff \chi = \pm \left(\eta - \frac{\pi}{2} \right)$$

These are precisely the lines determining the polar causal diamonds. It can also be checked that $r = R$ on the static patch corresponds to the north pole $\chi = 0$ in global coordinates and that $t = \tau$ along this line. Therefore, the static patch of dS provides coordinates that cover only the nothern causal diamond of dS spacetime, with the coordinate singularity at $r = R$ corresponding to the past and future observer-dependent horizons.

Finally, Penrose diagrams help to understand the nature of the $K = \partial_t$ Killing vector field exhibited by the static patch metric. As shown in Fig. 4.10, there's no global timelike Killing vector field in dS spacetime: extending the Killing vector beyond the static patch, i.e. the nothern causal diamond, it is timelike but past-oriented on the souther causal diamond and spacelike on the upper and lower quadrants.

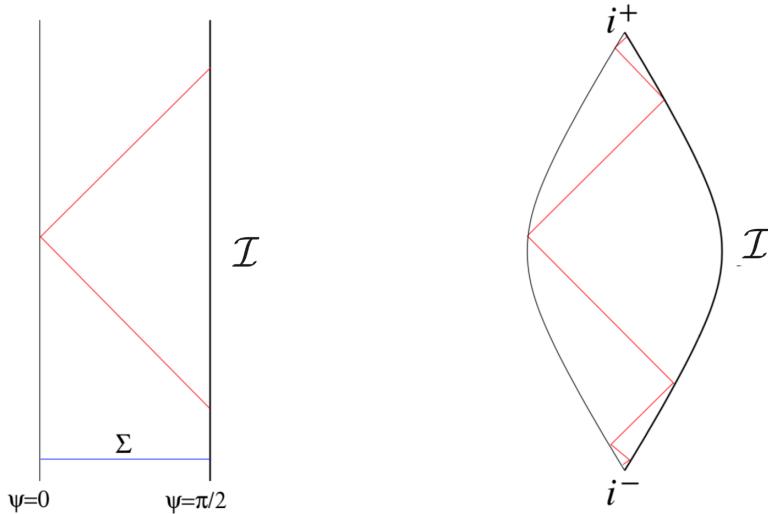


Figure 4.11: Penrose diagrams for AdS, without and with conformally compactified time coordinate.

§4.4.2.3 Anti-de Sitter space

The global coordinates on AdS are, from Eq. 4.2.11, $ds^2 = -\cosh^2 \rho dt^2 + R^2 d\rho^2 + R^2 \sinh^2 \rho d\Omega_2^2$, with $\rho \in [0, \infty)$. Introduce a conformal radial coordinate:

$$\frac{d\psi}{d\rho} = \frac{1}{\cosh \rho} \quad \Rightarrow \quad \cos \psi = \frac{1}{\cosh \rho}$$

with $\psi \in [0, \frac{\pi}{2})$. With a dimensionless coordinate time $\tilde{t} = \frac{t}{R}$, the metric becomes:

$$ds^2 = \frac{R^2}{\cos^2 \psi} (-d\tilde{t}^2 + d\psi^2 + \sin^2 \psi d\Omega_2^2) = \frac{R^2}{\cos^2 \psi} (-d\tilde{t}^2 + d\Omega_3^2)$$

AdS metric is therefore conformally equivalent to:

$$d\tilde{s}^2 = -d\tilde{t}^2 + d\psi^2 + \sin^2 \psi d\Omega_2^2$$

After conformal compactification, $\tilde{t} \in \mathbb{R}$ and $\psi \in [0, \frac{\pi}{2}]$ and the resulting Penrose diagram is the infinite strip in Fig. 4.11. The edge at $\psi = 0$ is not a boundary, but the spatial origin where \mathbb{S}^2 shrinks to a point; in contrast, $\psi = \frac{\pi}{2}$ is a boundary of the spacetime, labelled \mathcal{I} , which should be viewed as a combination of \mathcal{I}^- , \mathcal{I}^+ and i^0 , since null and spacelike geodesics begin and end there.

Proposition 4.4.4 (AdS boundary)

AdS spacetime has a timelike $\mathbb{R} \times \mathbb{S}^2$ boundary with spacelike normal vector.

Note that \mathbb{R} is the time factor. The Penrose diagram clearly shows that light rays reach the boundary in finite conformal time. To study physics in AdS, one needs to specify boundary conditions at \mathcal{I} : for example, reflecting boundary conditions make light rays bounce back and forth forever, rendering AdS a “box” spacetime in which massive particles are confined in the interior and massless particles bounce off the boundary.

Another characteristic of AdS space is that it is not *globally hyperbolic*: there exists no Cauchy surface on which initial data can be specified. Consider for example the 3-dimensional spacelike

hypersurface Σ in Fig. 4.11. Specifying initial data on Σ it's not sufficient to solve for their time evolution: in AdS, there exist points in the future of Σ which are in causal contact with the boundary, thus the time evolution depends on boundary conditions too.

To make the Penrose diagram for AdS not stretch to infinity, the time coordinate can be restricted to finite values by:

$$\tilde{t} = \tan \tau \quad \Rightarrow \quad ds^2 = \frac{R^2}{\cos^2 \psi \cos^4 \tau} (-d\tau^2 + \cos^4 \tau d\Omega_3^2)$$

This metric is conformally equivalent to:

$$d\tilde{s}^2 = -d\tau^2 + \cos^4 \tau (d\psi^2 + \sin^2 \psi d\Omega_2^2)$$

with conformally compactified $\tau \in [-\frac{\pi}{2}, +\frac{\pi}{2}]$. Ignoring the spatial \mathbb{S}^2 , the resulting Penrose diagram is drawn in Fig. 4.11: the spatial \mathbb{S}^3 grows and shrinks in time, the timelike boundary \mathcal{I} is still present and now the past and future timelike infinities i^\pm are also shown. This diagram makes it clear that a lightray bounces back and forth off the boundary of AdS an infinite number of times.

§4.5 Matter coupling

Spacetime is not merely the background on which matter exists, but it is dynamically influenced by the matter distribution on it. It is therefore necessary to study how matter couples to the spacetime metric.

§4.5.1 Field theories in curved spacetime

The simplest way to describe matter is by fields governed by a Lagrangian. Consider a scalar field $\phi(x)$. In flat Minkowski spacetime, its action is:

$$\mathcal{S}_{\text{scalar}} := \int d^4x \left[-\frac{1}{2}\eta^{\mu\nu}\partial_\mu\phi\partial_\nu\phi - V(\phi) \right] \quad (4.5.1)$$

The negative sign of the kinetic term follows from the signature choice $(-, +, +, +)$. The generalization to curved spacetime is straightforward:

$$\mathcal{S}_{\text{scalar}} = \int d^4x \sqrt{-g} \left[-\frac{1}{2}g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi - V(\phi) \right]$$

Note the (useful) redundancy $\nabla_\mu\phi = \partial_\mu\phi$. Curved spacetime also introduces the possibility to add new terms to the Lagrangian. For example:

$$\mathcal{S}_{\text{scalar}} = \int d^4x \sqrt{-g} \left[-\frac{1}{2}g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi - V(\phi) - \frac{1}{2}\xi R\phi^2 \right] \quad (4.5.2)$$

for some $\xi \in \mathbb{R}$. This theory correctly reduces to Eq. 4.5.1 on flat spacetime, as $R = 0$. To derive the equation of motion, vary the action keeping the metric fixed:

$$\begin{aligned} \delta\mathcal{S}_{\text{scalar}} &= \int d^4x \sqrt{-g} \left[-g^{\mu\nu}\nabla_\mu\delta\phi\nabla_\nu\phi - \frac{\partial V}{\partial\phi}\delta\phi - \xi R\phi\delta\phi \right] \\ &= \int d^4x \sqrt{-g} \left[\left(g^{\mu\nu}\nabla_\mu\nabla_\nu\phi - \frac{\partial V}{\partial\phi} - \xi R\phi \right) \delta\phi - \nabla_\mu(\delta\phi\nabla^\mu\phi) \right] \end{aligned}$$

where integration by parts was possible due to $\nabla_\mu g_{\rho\sigma} = 0$ for the Levi-Civita connection. The last term is, by divergence theorem, a boundary term, thus the equation of motion for a scalar field theory in curved spacetime is:

$$g^{\mu\nu}\nabla_\mu\nabla_\nu\phi - \frac{\partial V}{\partial\phi} - \xi R\phi = 0 \quad (4.5.3)$$

Now covariant derivatives are necessary, as $\nabla_\mu\nabla_\nu \neq \partial_\mu\partial_\nu$.

§4.5.2 Einstein equations with matter

To understand how matter fields back-react on spacetime, consider the combined action:

$$\mathcal{S} = \frac{1}{16\pi G} \int d^4x \sqrt{-g} (R - 2\Lambda) + \mathcal{S}_M \quad (4.5.4)$$

where \mathcal{S}_M is the action for matter fields, which in general depends on both the fields and the metric.

Definition 4.5.1 (Energy-momentum tensor)

Given a field theory for matter described by the action \mathcal{S}_M , the **energy-momentum tensor** is defined as:

$$T_{\mu\nu} := -\frac{2}{\sqrt{-g}} \frac{\delta \mathcal{S}_M}{\delta g^{\mu\nu}} \quad (4.5.5)$$

Clearly, the energy-momentum tensor is symmetric, since it inherits the symmetry of the metric.

Proposition 4.5.1 (Einstein field equations)

The equations of motion derived from the action Eq. 4.5.4 are:

$$G_{\mu\nu} + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu} \quad (4.5.6)$$

Proof. Varying the full metric, by Def. 4.5.1:

$$\delta \mathcal{S} = \frac{1}{16\pi G} \int d^4x \sqrt{-g} [G_{\mu\nu} + \Lambda g_{\mu\nu}] \delta g^{\mu\nu} - \frac{1}{2} \int d^4x \sqrt{-g} T_{\mu\nu} \delta g^{\mu\nu} = 0$$

which is the thesis. \square

These are the full Einstein field equations, describing gravity coupled to matter. It is possible to rewrite them by observing that the cosmological constant can be absorbed in the energy-momentum tensor as an additive component:

$$(T_\Lambda)_{\mu\nu} = -\frac{\Lambda}{8\pi G} g_{\mu\nu} \quad (4.5.7)$$

This is justified by the fact that matter fields often mimic a cosmological constant. Contracting the remaining equation with $g^{\mu\nu}$ (i.e. taking the trace) then gives $-R = 8\pi G T$, where $T \equiv g^{\mu\nu} T_{\mu\nu}$, hence:

$$R_{\mu\nu} = 8\pi G \left(T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu} \right) \quad (4.5.8)$$

Remember that the cosmological constant is present inside the energy-momentum tensor.

§4.5.3 Energy-momentum tensor

The action \mathcal{S}_M is, by hypothesis, diffeomorphism-invariant, thus, recalling the argument which lead to Bianchi identity Eq. 4.1.13, given $\delta g_{\mu\nu} = (\mathcal{L}_X g)_{\mu\nu} = 2\nabla_{(\mu} X_{\nu)}$:

$$\delta \mathcal{S}_M = -\frac{1}{2} \int d^4x \sqrt{-g} T_{\mu\nu} \delta g^{\mu\nu} = - \int d^4x \sqrt{-g} T_{\mu\nu} \nabla^\mu X^\nu$$

Diffeomorphism invariance means that $\delta \mathcal{S}_M = 0$ for all $X \in \mathfrak{X}(\mathcal{M})$, hence, integrating by parts:

$$\nabla_\mu T^{\mu\nu} = 0 \quad (4.5.9)$$

Of course, this was necessary to make Einstein equations consistent, as $\nabla_\mu G^{\mu\nu} = 0$. Anyway, this equation hints to the more profound nature of the energy-momentum tensor, which has nothing to do with gravity: it can be shown that the energy-momentum tensor is linked to Noether currents associated to translational invariance in space and time. Trivially, Eq. 4.5.9 reduces in flat spacetime to $\partial_\mu T^{\mu\nu} = 0$, which is the usual conservation law of Noether currents.

Consider a translation $x^\mu \mapsto x^\mu + \delta x^\mu$, with $\delta x^\mu = X^\mu(x)$. The action restricted to flat spacetime is not invariant under such shift, but one which is invariant can be constructed coupling the matter fields to a background metric and allowing this to vary. The change of the action in flat space, where the metric is fixed, must be equal and opposite to the change of the action where the metric can vary but x^μ is fixed, thus:

$$\begin{aligned}\delta\mathcal{S}_{\text{flat}} &= - \int d^4x \frac{\delta\mathcal{S}_M}{\delta g^{\mu\nu}} \Big|_{g_{\mu\nu}=\eta_{\mu\nu}} \delta g^{\mu\nu} = - \int d^4x \frac{\delta\mathcal{S}_M}{\delta g^{\mu\nu}} \Big|_{g_{\mu\nu}=\eta_{\mu\nu}} \partial^{(\mu} X^{\nu)} \\ &= -2 \int d^4x \frac{\delta\mathcal{S}_M}{\delta g^{\mu\nu}} \Big|_{g_{\mu\nu}=\eta_{\mu\nu}} \partial^\mu X^\nu = -2 \int d^4x \partial^\mu \left[\frac{\delta\mathcal{S}_M}{\delta g^{\mu\nu}} \right]_{g_{\mu\nu}=\eta_{\mu\nu}} X^\nu\end{aligned}$$

But $\delta\mathcal{S}_{\text{flat}} = 0$ for all constant X^μ , as this is the definition of a translationally-invariant theory, hence the conserved Noether current in flat space is:

$$T_{\mu\nu} = -2 \frac{\delta\mathcal{S}_M}{\delta g^{\mu\nu}} \Big|_{g_{\mu\nu}=\eta_{\mu\nu}}$$

i.e. the flat version of Eq. 4.5.5.

§4.5.3.1 Field theories

It is straightforward to compute $T_{\mu\nu}$ for a scalar field $\phi(x)$. Recall Eq. 4.5.2 (with $\xi = 0$) and Lemma 4.1.1:

$$\delta\mathcal{S}_{\text{scalar}} = \int d^4x \sqrt{-g} \left[\frac{1}{4} g_{\mu\nu} \nabla^\rho \phi \nabla_\rho \phi + \frac{1}{2} g_{\mu\nu} V(\phi) - \frac{1}{2} \nabla_\mu \phi \nabla_\nu \phi \right] \delta g^{\mu\nu}$$

This gives the energy-momentum tensor:

$$T_{\mu\nu} = \nabla_\mu \phi \nabla_\nu \phi - g_{\mu\nu} \left(\frac{1}{2} \nabla^\rho \phi \nabla_\rho \phi + V(\phi) \right) \quad (4.5.10)$$

Restricting to flat Minkowski spacetime:

$$T_{00} = \frac{1}{2} \dot{\phi}^2 + \frac{1}{2} (\nabla \phi)^2 + V(\phi)$$

which is the energy density of a scalar field.

Maxwell theory Varying Maxwell action Eq. 2.2.22:

$$\delta\mathcal{S}_{\text{Maxwell}} = -\frac{1}{4} \int dx^4 \sqrt{-g} \left[-\frac{1}{2} g_{\mu\nu} F^{\rho\sigma} F_{\rho\sigma} + 2g^{\rho\sigma} F_{\mu\rho} F_{\nu\sigma} \right] \delta g^{\mu\nu}$$

So the energy-momentum tensor for Maxwell theory is:

$$T_{\mu\nu} = g^{\rho\sigma} F_{\mu\rho} F_{\nu\sigma} - \frac{1}{4} g_{\mu\nu} F^{\rho\sigma} F_{\rho\sigma} \quad (4.5.11)$$

In flat Minkowski spacetime:

$$T_{00} = \frac{1}{2} \mathbf{E}^2 + \frac{1}{2} \mathbf{B}^2$$

which is the energy density of the electromagnetic field.

§4.5.3.2 Perfect fluids

A perfect fluid is described by its energy density $\rho(\mathbf{x}, t)$, pressure $p(\mathbf{x}, t)$ and velocity 4-vector field $u^\mu(\mathbf{x}, t) : u^\mu u_\mu = -1$. Pressure and energy density are related by an *equation of state* $p = p(\rho)$.

Example 4.5.1 (Dust)

Dust is a fluid of massive particles floating around very slowly, so that the equation of state is $p = 0$.

Example 4.5.2 (Radiation)

Radiation is a fluid of photons with $p = \rho/3$.

The energy-momentum tensor of a perfect fluid is:

$$T^{\mu\nu} = (\rho + p)u^\mu u^\nu + pg^{\mu\nu} \quad (4.5.12)$$

A fluid at rest ($u^\mu = \delta^{\mu,0}$) in flat Minkowski spacetime has $T^{\mu\nu} = \text{diag}(\rho, p, p, p)$, thus T_{00} is yet again the energy density, as expected. Generally, $\rho = T_{\mu\nu}u^\mu u^\nu$ is the energy density measured by an observer co-moving with the fluid.

Bianchi identity $\nabla_\mu T^{\mu\nu} = 0$ determines two constraints. The first is:

$$u^\mu \nabla_\mu \rho + (\rho + p) \nabla_\mu u^\mu = 0 \quad (4.5.13)$$

which is a generalization of mass conservation (where mass is identified with ρ). The first term calculates how fast ρ changes along u^μ , while the second expresses it depending on the rate of flow out of the region $\nabla_\mu u^\mu$. The second constraint is:

$$(\rho + p)u^\mu \nabla_\mu u^\nu = -(g^{\mu\nu} + u^\mu u^\nu) \nabla_\mu p \quad (4.5.14)$$

which is a generalization of Euler equation, i.e. the fluid equivalent of Newton's second law.

§4.5.4 Energy conservation

There is a difference between the energy-momentum tensor and the current which arises from a global symmetry. Consider a conserved current $J^\mu : \nabla_\mu J^\mu = 0$. Invoking the divergence theorem Eq. 2.2.19:

$$0 = \int_V d^4x \sqrt{-g} \nabla_\mu J^\mu = \int_{\partial V} d^3x \sqrt{\gamma} n_\mu J^\mu$$

where V is a spatial volume with boundary $\partial V = \Sigma_1 \cup \Sigma_2 \cup B$, with Σ_1, Σ_2 past and future spacelike boundaries and B timelike (lateral) boundary. If no current flows out of the region, i.e. $n_\mu J^\mu|_B = 0$, then this expression becomes the conservation $Q(\Sigma_1) = Q(\Sigma_2)$ of the charge associated to the current:

$$Q(\Sigma) \equiv \int_\Sigma d^3x \sqrt{\gamma} n_\mu J^\mu$$

Thus, for a vector field, covariant conservation is equivalent to actual conservation. However, the same argument does not apply to the energy-momentum tensor: the problem arises from generalizing $\nabla_\mu J^\mu = \frac{1}{\sqrt{-g}} \partial_\mu (\sqrt{-g} J^\mu)$ to a higher-order tensor field, which is necessary in order to have a divergence theorem like Eq. 2.2.19. Indeed:

$$\nabla_\mu T^{\mu\nu} = \partial_\mu T^{\mu\nu} + \Gamma_{\mu\rho}^\mu T^{\rho\nu} + \Gamma_{\mu\rho}^\nu T^{\mu\rho} = \frac{1}{\sqrt{-g}} \partial_\mu (\sqrt{-g} T^{\mu\nu}) + \Gamma_{\mu\rho}^\nu T^{\mu\rho}$$

The last term doesn't allow to convert the integral of $\nabla_\mu T^{\mu\nu}$ to a boundary term. Instead:

$$\partial_\mu(\sqrt{-g} T^{\mu\nu}) = -\sqrt{-g} \Gamma_{\mu\rho}^\nu T^{\mu\rho} \quad (4.5.15)$$

Therefore, for higher-order tensors, covariant conservation is not equivalent to actual conservation.

§4.5.4.1 Conserved energy from Killing vectors

Given a Killing vector K , it is possible to define a conserved current associated to the energy-momentum tensor as:

$$J_T^\nu := -K_\mu T^{\mu\nu} \quad (4.5.16)$$

This current is covariantly conserved, as:

$$\nabla_\nu J_T^\nu = -(T^{\mu\nu} \nabla_\nu K_\mu + K_\mu \nabla_\nu T^{\mu\nu}) = -2T^{\mu\nu} \nabla_{(\nu} K_{\mu)} = 0$$

Its associated conserved charge on a spatial hypersurface Σ is defined as:

$$Q_T(\Sigma) := \int_{\Sigma} d^3x \sqrt{\gamma} n_\mu J_T^\mu \quad (4.5.17)$$

The interpretation of this charge depends on the properties of the Killing vector: if K is globally timelike, then the charge is the energy of matter $E = Q_T(\Sigma)$, meanwhile if K is globally spacelike, then it is the momentum of matter.

Absence of Killing vectors The problem of energy conservation becomes subtle when dealing with spacetimes which do not have any globally timelike Killing vector.

For example, a system comprised of two orbiting stars is modelled by a spacetime which does not have such a Killing vector: however, the problem of energy conservation does not arise in this case, as the stars, while orbiting each other, emit gravitational waves, thus losing energy and eventually spiraling towards each other. Nonetheless, a meaningful question is that of energy conservation of the total system, i.e. the two stars and the gravitational field. A guess would be to consider a total energy-momentum tensor defined similarly to Eq. 4.5.5, but:

$$T_{\mu\nu}^{\text{total}} = -\frac{2}{\sqrt{-g}} \left[\frac{1}{16\pi G} \frac{\delta S_{\text{EH}}}{\delta g^{\mu\nu}} + \frac{\delta S_M}{\delta g^{\mu\nu}} \right] = -\frac{1}{8\pi G} G_{\mu\nu} + T_{\mu\nu} = 0$$

by Einstein field equations. This equation has no physical significance other than expressing the subtlety of energy conservation in General Relativity.

Clearly, one could try to understand the energy carried by the gravitational field alone. Unfortunately, there are compelling arguments that there exists no tensor which can be thought as the local energy density of the gravitational field: roughly speaking, the energy density of the Newtonian gravitational field Φ is proportional to $(\nabla\Phi)^2$, so the relativistic equivalent should be proportional to the first derivatives of the metric, which can be made locally vanishing by normal coordinates due to the equivalence principle, and a tensor which vanishes in one coordinate system does so in all of them.

§4.5.5 Energy conditions

To study the general properties of a spacetime without explicitly referencing the specific characteristics of matter, one needs to place certain restrictions on the kinds of energy-momentum tensor which are considered physical: these are the *energy conditions* and express the idea that energy should be positive.

§4.5.5.1 Weak energy condition

This condition states that, for any timelike vector field X :

$$T_{\mu\nu}X^\mu X^\nu \geq 0$$

This quantity is the energy measured by an observer moving along the timelike integral curves of X , and it should be non-negative. A timelike curve can be arbitrarily close to a null curve, thus by continuity:

$$T_{\mu\nu}X^\mu X^\nu \geq 0 \quad \forall X \in \mathfrak{X}(\mathcal{M}) : X_\mu X^\mu \leq 0 \quad (4.5.18)$$

Fluids Recall the energy-momentum tensor for a fluid Eq. 4.5.12 and impose the weak energy condition (WLOG $X \cdot X = -1$):

$$(\rho + p)(u \cdot X)^2 - p \geq 0$$

In the rest-frame $u^\mu = (1, 0, 0, 0)$, so considering a constant $X^\mu = (\cosh \varphi, \sinh \varphi, 0, 0)$, whose integral curves are world-lines of observers boosted with rapidity φ with respect to the fluid, then:

$$(\rho + p) \cosh^2 \varphi - p \geq 0 \quad \Rightarrow \quad \begin{cases} \rho \geq 0 & \varphi = 0 \\ p \geq -\rho & \varphi \rightarrow \infty \end{cases}$$

The first condition ensures that the energy density is positive, the second allows for negative pressure limited from below.

Note that there are situations in which negative energy density makes physical sense: viewing the cosmological constant as part of the energy-momentum density, then any $\Lambda < 0$ violates the weak energy condition. In this sense, AdS space violates the weak energy condition.

Scalar fields The weak energy condition for the energy-momentum tensor of a scalar field theory Eq. 4.5.10 reads:

$$(X^\mu \partial_\mu \phi)^2 + \frac{1}{2} \partial_\mu \phi \partial^\mu \phi + V(\phi) \geq 0$$

The sum of the first two terms is always positive: define $Y_\mu = \partial_\mu \phi + X_\mu X^\nu \partial_\nu \phi : X_\mu Y^\mu = 0$, i.e. orthogonal to X_μ , so it must be spacelike or null, hence $Y_\mu Y^\mu \geq 0$. Rewriting the above condition:

$$\frac{1}{2} (X^\mu \partial_\mu \phi)^2 + \frac{1}{2} Y_\mu Y^\mu + V(\phi) \geq 0 \quad \Rightarrow \quad V(\phi) \geq 0$$

The weak energy condition is clearly violated by any classical field theory with $V(\phi) \leq 0$.

§4.5.5.2 Strong energy condition

This condition states that:

$$R_{\mu\nu}X^\mu X^\nu \geq 0 \quad \forall X \in \mathfrak{X}(\mathcal{M}) : X_\mu X^\mu \leq 0 \quad (4.5.19)$$

The strong energy condition ensures that timelike geodesics converge, i.e. that gravity is attractive. Using Eq. 4.5.8, this condition can be rewritten as:

$$\left(T_{\mu\nu} - \frac{1}{2} T g_{\mu\nu} \right) X^\mu X^\nu \geq 0$$

Taking yet again $X \cdot X = -1$, the strong energy condition for the energy-momentum tensor of a fluid Eq. 4.5.12 in the rest-frame reads:

$$(\rho + p)(u \cdot X)^2 - p + \frac{1}{2}(3p - \rho) \geq 0 \implies (\rho + p) \cosh \varphi^2 - p + \frac{1}{2}(3p - \rho) \geq 0$$

whose solution is:

$$\begin{cases} p \geq -\rho/3 & \varphi = 0 \\ p \geq -\rho & \varphi \rightarrow \infty \end{cases}$$

It is possible to show that $\Lambda > 0$ violates the strong energy condition: dS space is incompatible with it, as timelike geodesics are pulled apart by the expansion of space.

Moreover, any classical scalar field theory with $V(\phi) \geq 0$ violates the strong energy condition.

§4.5.5.3 Null energy condition

This condition states that:

$$T_{\mu\nu}X^\mu X^\nu \geq 0 \quad \forall X \in \mathfrak{X}(\mathcal{M}) : X_\mu X^\mu = 0 \quad (4.5.20)$$

This condition is implied by both the weak and strong energy condition, but the converse is not true: it is weaker than both conditions. However, it is satisfied by any classical field theory and any perfect fluid with $\rho + p \geq 0$.

§4.5.5.4 Dominant energy condition

Given a future-directed timelike vector field X , it is possible to define the energy density current measured by an observer moving along integral curves of X as:

$$J^\mu \equiv -T^{\mu\nu}X_\nu$$

The dominant energy condition requires that, in addition to the WEC Eq. 4.5.18:

$$J_\mu J^\mu \leq 0 \quad (4.5.21)$$

This means that the energy density current is either timelike or null, so requiring that energy does not flow faster than light.

It is possible to check that this condition is always satisfied by a classical scalar field theory, while for a perfect fluid it imposes $\rho^2 \geq p^2$.

§4.6 Schwarzschild solution

The Schwarzschild solution is the unique static and spherically-symmetric vacuum solution to the Einstein field equations Eq. 4.1.5: it describes spacetime outside a spherically-symmetric body, and it is the most important exact solution of the field equations.

§4.6.1 Birkhoff theorem

To prove the uniqueness of the Schwarzschild solution, first a definition of “static” spacetime is required. Given that symmetries are defined through Killing vector fields (§4.3.1), it is straightforward to formulate precise definitions of “stationary” and “static” using them.

Definition 4.6.1 (Stationary spacetime)

A Lorentzian manifold (\mathcal{M}, g) is **stationary** if there exists a globally-timelike Killing vector field $K \in \mathfrak{X}(\mathcal{M})$.

This means that an observer moving along the flow of K , which is timelike, perceives no changes in spacetime, since the flow of K is an isometry group of \mathcal{M} . It may happen that K is timelike only in some open region of \mathcal{M} : then, only this part of spacetime is said stationary.

Proposition 4.6.1 (Stationary metric)

If (\mathcal{M}, g) is a stationary Lorentzian manifold, then $g_{\mu\nu,0} = 0 \forall \mu, \nu = 0, 1, 2, 3$.

Proof. Consider a spacelike hypersurface $\Sigma \subseteq \mathcal{M}$ and be $\phi : \mathbb{R} \times \mathcal{M} \rightarrow \mathcal{M}$ the flow of K on \mathcal{M} defined by Eq. 1.2.9. Then, define the following local coordinates on \mathcal{M} : if $p = \phi_t(p_0)$, with $p_0 \in \Sigma$ and $t \in \mathbb{R}$, then the coordinates of p are $(t, x^1(p_0), x^2(p_0), x^3(p_0))$, where x^μ are local coordinates on Σ . In terms of these coordinates, clearly $K = \partial_0$, hence using Eq. 1.3.18:

$$(\mathcal{L}_K g)_{\mu\nu} = g_{\mu\nu,\rho} K^\rho + g_{\rho\nu} K^\rho,_\mu + g_{\mu\rho} K^\rho,_\nu = g_{\mu\nu,0} + 0 + 0 = g_{\mu\nu,0}$$

Therefore, Killing equation $\mathcal{L}_K g = 0$ implies the thesis. \square

These coordinates are called *adapted coordinates* to the Killing field. To deduce a definition of static spacetime, which is a special case of stationary spacetime, consider adapted coordinates such that $g_{0i} = 0 \forall i = 1, 2, 3$, so that the Killing field is orthogonal to all spatial sections $\{t = \text{const.}\}$. Then, the 1-form $\omega = K^\flat$ associated to K is:

$$\omega = g_{\mu\nu} K^\mu dx^\nu = g_{00} dt = g(K, K) dt$$

since $g_{00} = g(\partial_0, \partial_0) = g(K, K)$. This trivially implies the **Frobenius condition** for ω :

$$\omega \wedge d\omega = 0 \tag{4.6.1}$$

Conversely, assume that the Frobenius condition holds for the Killing field K of a stationary Lorentzian manifold (\mathcal{M}, g) . Then, applying ι_K to Eq. 4.6.1 and using Th. 1.4.2 yields:

$$0 = \iota_K(\omega \wedge d\omega) = (\iota_K \omega) \wedge d\omega - \omega \wedge (\iota_K d\omega) = g(K, K) d\omega - \omega \wedge (\mathcal{L}_K \omega - dg(K, K))$$

To show that $\mathcal{L}_K \omega = 0$, expand both $\mathcal{L}_K \omega$ and $\mathcal{L}_K g$:

$$(\mathcal{L}_K \omega)(X) = K(\omega(X)) - \omega([K, X]) = K(g(K, X)) - g(K, [K, X])$$

$$0 = (\mathcal{L}_K g)(K, X) = K(g(K, X)) - g([K, K], X) - g([K, X], K) = K(g(K, X)) - g(K, [K, X])$$

Hence, setting $V \equiv g(K, K) \neq 0$, the above equation becomes:

$$0 = V d\omega + \omega \wedge dV = d(V^{-1}\omega)$$

Assuming \mathcal{M} to be simply connected, by the Poincaré Lemma $\omega = V df$, with $f \in C^\infty(\mathcal{M})$. Using this function as a time coordinate, then, the expression $\omega = g(K, K)dt$ is recovered, which shows that K is orthogonal to all spacelike sections $\{t = \text{const.}\}$, since taken $X \in \mathfrak{X}(\mathcal{M})$ tangential to one of these sections $g(K, X) = \omega(X) = V dt(X) = V X(dt) = 0$. Switching to adapted coordinates, $K = \partial_t$, so $g_{0i} = g(\partial_t, \partial_i) = g(K, \partial_i) = 0$. Therefore, this reasoning shows that, if the globally-timelike Killing vector field K satisfies the Frobenius condition, then the metric locally splits as:

$$ds^2 = g_{00}(\mathbf{x})dt^2 + g_{ij}(\mathbf{x})dx^i dx^j \quad (4.6.2)$$

This justifies the following definition.

Definition 4.6.2 (Static spacetime)

Given a stationary Lorentzian manifold (\mathcal{M}, g) with globally-timelike Killing vector field K , then it is said to be **static** if, setting $\omega \equiv K^\flat$, then $\omega \wedge d\omega = 0$ and locally $\omega = g(K, K)dt$ for an adapted time coordinate t , which is unique up to an additive constant.

In a static spacetime, the flow of K isometrically maps spacelike sections $\{t = \text{const.}\}$ onto each other, and an observer at rest moves along integral curves of K . In the following, K is assumed to be the only globally-timelike Killing vector field, so that there is a *distinguished time*.

Theorem 4.6.1 (Birkhoff's theorem)

A spherically-symmetric vacuum solution to the Einstein field equations must be static, and it is uniquely given by the **Schwarzschild solution**:

$$ds^2 = - \left(1 - \frac{2m}{r}\right) dt^2 + \left(1 - \frac{2m}{r}\right)^{-1} dr^2 + r^2 [d\vartheta^2 + \sin^2 \vartheta d\varphi^2] \quad (4.6.3)$$

on the manifold $\mathcal{M} = \mathbb{R} \times \mathbb{R}_{>2m}^+ \times \mathbb{S}^2$, with $m \in \mathbb{R}$.

Proof. First of all, spherical symmetry requires that the Lorentzian manifold considered has $\text{SO}(3)$ as an isometry group, with the conditions that the group orbits^a are 2-dimensional spacelike hypersurfaces. Thus, the manifold splits into $\mathcal{M} = \Sigma \times \mathbb{S}^2$, where Σ is a 2-dimensional Lorentzian submanifold. Since the metric on a 2-dimensional pseudo-Riemannian manifold can be always diagonalized^b, the most general spherically-symmetric metric takes the form:

$$ds^2 = -e^{2a(t,r)} dt^2 + e^{2b(t,r)} dr^2 + r^2 [d\vartheta^2 + \sin^2 \vartheta d\varphi^2]$$

The ϑ, φ are the standard coordinates on \mathbb{S}^2 , $t \in \mathbb{R}$ and the radial coordinate r is determined by the foliation of the spatial part of the manifold by invariant 2-spheres: in particular, each 2-sphere is characterized by a surface area A , and r is defined by $A = 4\pi r^2$.

With notation $f' \equiv \frac{df}{dr}$ and $f \equiv \frac{df}{dt}$, the only non-vanishing components of the Einstein

tensor are computed to be:

$$\begin{aligned} G^0_0 &= -\frac{1}{r^2} + e^{-2b} \left(\frac{1}{r^2} - \frac{2b'}{r} \right) & G^1_1 &= -\frac{1}{r^2} + e^{-2b} \left(\frac{1}{r^2} + \frac{2a'}{r} \right) & G^1_0 &= \frac{2\dot{b}}{r} e^{-a-b} \\ G^2_2 &= G^3_3 = e^{-2b} \left(a'^2 - a'b' + a'' + \frac{a' - b'}{r} \right) + e^{-2a} \left(-\dot{b}^2 + \dot{a}\dot{b} - \ddot{b} \right) \end{aligned}$$

Now, the vacuum field equations read $G^\mu_\nu = 0$. From $G^1_0 = 0$ it is clear that $b = b(r)$; then, $G_{11} + G_{00} = 0$ implies the partial differential equation $a' + b' = 0$, which has the general solution $a(t, r) = -b(r) + f(t)$. To determine $b(r)$, consider $G^0_0 = 0$:

$$-\frac{1}{r^2} + e^{-2b} \left(\frac{1}{r^2} - \frac{2b'}{r} \right) = 0 \implies (re^{-2b}) = 1 \implies e^{-2b} = 1 - \frac{2m}{r}$$

where $m \in \mathbb{R}$ is an integration constant. Then, the other vacuum field equations are satisfied, and the metric takes the form:

$$ds^2 = -e^{2f(t)} \left(1 - \frac{2m}{r} \right) dt^2 + \left(1 - \frac{2m}{r} \right)^{-1} + r^2 [d\vartheta^2 + \sin^2 \varphi d\varphi^2]$$

Introducing a new time coordinate $\tilde{t} = \int dt e^{f(t)}$, this metric reduces to the Schwarzschild solution, which is manifestly static with the timelike Killing vector field $K = \partial_t$. \square

^aGiven a group G which acts on a set X , the orbit of $x \in X$ is defined as $G(x) := \{gx \in X : g \in G\} \subseteq X$.

^bNeeds a proof.

The converse of Birkhoff's theorem is also true, and it is called Israel's theorem. The physical meaning of the parameter $m \in \mathbb{R}$ is clarified by studying the Newtonian limit for g_{00} at large distances:

$$-\left(1 - \frac{2m}{r} \right) \approx -\left(1 + \frac{2\Phi(r)}{c^2} \right)$$

Since $\Phi(r) = -GM/r$, where M is the total mass generating the gravitational field, then:

$$m = \frac{GM}{c^2} \tag{4.6.4}$$

Hence, m can be seen as the mass of the spherically-symmetric object which generates the Schwarzschild field: this restricts the physical significance of m to $m \in \mathbb{R}^+$, as $m = 0$ corresponds to flat Minkowski space. As a consequence, the Schwarzschild metric describes the spacetime outside any non-rotating spherically-symmetric object, even in presence of time dependence in the internal distribution of matter (e.g. a collapsing star).

Note that the solution is singular at $r = 2m \equiv R_s$: indeed, Birkhoff's theorem only applies in the $r > R_s$ region, while the “interior” region, as well as the nature of the singularity at $r = R_s$ and $r = 0$, remains to be studied (in §6.1.1). It is important to notice that r is not the “distance from the center”: indeed, given the foliated structure of the Schwarzschild spacetime, dr is not the distance between two 2-spheres with r and $r + dr$ (defined by their surface area), which is instead $dl = dr/\sqrt{1 - 2m/r}$. Therefore, it is clear that the $\{t = \text{const.}\}$ sections of spacetime are non-Euclidean.

§4.6.2 Motion in a Schwarzschild field

The Lagrangian for a test body in a Schwarzschild field, parametrized with proper time, is:

$$2L = - \left(1 - \frac{2m}{r}\right) \dot{t}^2 + \left(1 - \frac{2m}{r}\right)^{-1} \dot{r}^2 + r^2 [\dot{\vartheta}^2 + \sin^2 \vartheta \dot{\varphi}^2]$$

Note that, along the worldline $x^\mu(\tau)$, which is a geodetic on \mathcal{M} , the Lagrangian takes the value $2L = -1$, as $ds^2 = -d\tau^2$. The equation of motion for ϑ reads:

$$\frac{d}{d\tau}(r^2 \dot{\vartheta}) = r^2 \sin \vartheta \cos \vartheta \dot{\varphi}^2$$

Therefore, if the motion is initially in the equatorial plane, i.e. $\vartheta = \frac{\pi}{2}$ and $\dot{\vartheta} = 0$, it will remain so, hence WLOG $\vartheta \equiv \frac{\pi}{2}$ and the Lagrangian reduces to:

$$2L = - \left(1 - \frac{2m}{r}\right) \dot{t}^2 + \left(1 - \frac{2m}{r}\right)^{-1} \dot{r}^2 + r^2 \dot{\varphi}^2 \quad (4.6.5)$$

The variables t and φ are cyclic, and they correspond to two conserved quantities:

$$\frac{\partial L}{\partial \dot{\varphi}} = r^2 \dot{\varphi} \equiv \ell \quad - \frac{\partial L}{\partial \dot{t}} = \left(1 - \frac{2m}{r}\right) \dot{t} \equiv E \quad (4.6.6)$$

which are identified respectively with the angular momentum and the energy of the test body. Inserting these definitions into $2L = -1$, the equation of motion becomes:

$$\dot{r}^2 + V_{\text{eff}}(r) = E^2 \quad (4.6.7)$$

with the effective potential:

$$V_{\text{eff}}(r) = \left(1 - \frac{2m}{r}\right) \left(1 + \frac{\ell^2}{r^2}\right) \quad (4.6.8)$$

This potential always has asymptotics $V_{\text{eff}}(r) \rightarrow -\infty$ for $r \rightarrow 0$ and $V_{\text{eff}}(r) \rightarrow 1$ for $r \rightarrow +\infty$, but the intermediate behaviour depends on the ratio ℓ/m : in particular, for $\ell/m < 2\sqrt{3}$ the potential has no critical points, hence any incoming particle falls towards the singularity at $r = R_s$.

The main physically-interesting quantity is the orbit $r = r(\varphi)$, which can be determined from this equation of motion.

Proposition 4.6.2 (Quasi-Keplerian motion)

Defining $u(\varphi) \equiv \frac{1}{r(\varphi)}$ and $u' \equiv \frac{du}{d\varphi}$, then the orbital equation becomes:

$$u'' + u = \frac{m}{\ell^2} + 3mu^2 \quad (4.6.9)$$

Proof. First, note that by the chain rule:

$$\dot{r} = r' \dot{\varphi} = r' \frac{\ell}{r^2} \implies r'^2 \frac{\ell^2}{r^4} = E^2 - V_{\text{eff}}(r)$$

Then, introducing $r' = -u'/u^2$:

$$\ell^2 u'^2 = E^2 - (1 - 2mu) (1 + \ell^2 u^2)$$

Differentiating with respect to φ results in:

$$2u'u'' + 2uu' = \frac{2m}{\ell^2}u' + 6mu^2u'$$

Hence, either $u' = 0$, which is the equation for circular motion $r = \text{const.}$, or it solves the ODE Eq. 4.6.9. \square

In the Newtonian case $L = \frac{1}{2}\dot{\mathbf{x}}^2 + \frac{GM}{r^2}$, which results in the Keplerian orbital equation:

$$u'' + u = \frac{GM}{\ell^2} \quad (4.6.10)$$

Therefore, the motion in a Schwarzschild field is quasi-Keplerian, since the orbital equation contains an additional perturbative term $3mu^2$. For our planet system, this perturbation is relatively small, e.g. for Mercury:

$$\frac{3mu^2}{m/\ell^2} = 3u^2\ell^2 = \frac{3}{r^2} (r^2\dot{\varphi})^2 \simeq \frac{3}{c^2} \left(r \frac{d\varphi}{dt} \right)^2 \simeq \frac{3v_{\perp}^2}{c^2} \simeq 7.7 \cdot 10^{-8}$$

where v_{\perp} is the velocity component perpendicular to the radius vector.

§4.6.2.1 Freely-falling emitter

Consider a point-like source, emitting light in a constant and isotropic way in its comoving reference frame, which is radially falling from infinity. The emitted light is detected by an observer in a fixed position, along the same radial direction, at $r = R \gg R_s$.

Since light propagates on null curves with $ds^2 = 0$, from Eq. 4.6.3:

$$dt = \pm \frac{r}{r - R_s} dr \implies t = r + R_s \ln(r - R_s) + \text{const.}$$

where the positive solution has been chosen and the \mathbb{S}^2 component has been ignored, since the motion is radial (i.e. with fixed ϑ and φ). Now, fix an emission event (t_e, r_e) , which is detected at (t_r, R) . Clearly, by the above relation:

$$t_r - R - R_s \ln(R - R_s) = t_e - r_e - R_s \ln(r_e - R_s) \quad (4.6.11)$$

It is possible to make the relation between t_e and r_e explicit. Recall Eq. 4.6.8 and set $\ell = 0$ (radial motion), so that:

$$\dot{r}^2 = E^2 - 1 + \frac{R_s}{r} \equiv \frac{R_s}{r} - \frac{R_s}{r_0} \quad r_0 \equiv \frac{R_s}{1 - E^2}$$

where r_0 is the initial radial coordinate of the geodesic. Note that this is the same equation which can be derived from energy conservation in Newtonian mechanics, with $\Phi(r) = -GMm_s/r$. Since the source is falling from infinity, $E = 1$ (which justifies the choice of the positive solution for $t = t(r)$), hence the geodesic is described by:

$$r = R_s s^2 \quad \tau = -\frac{2}{3} R_s s^3 \quad (4.6.12)$$

where $s \in \mathbb{R}_{>1}^+$ is an auxiliary variable and the choice of the sign of τ is dictated by the fact that the source is falling, so s decreases along the trajectory while τ must always increase. Then, by Eq. 4.6.6:

$$\frac{r - R_s}{r} \frac{dt}{d\tau} = 1 \implies t = -\frac{2}{3}R_s s^3 - 2R_s s + R_s \ln \frac{s+1}{s-1}$$

Inserting this into Eq. 4.6.11 in place of (t_e, r_e) :

$$t_r = -\frac{2}{3}R_s s^3 - R_s s^2 - 2R_s s - 2R_s \ln(s-1) + R + R_s \ln \left(\frac{R}{R_s} - 1 \right)$$

In the limit $s \rightarrow 1$, i.e. $r \rightarrow R_s$, this equation becomes:

$$t_r \rightarrow -2R_s \ln(s-1) + \text{const.} \implies s \rightarrow 1 + ae^{-\frac{t_r}{2R_s}}$$

with $a \equiv \exp(\text{const.}/2R_s)$. Differentiating this equation and Eq. 4.6.12, then:

$$\frac{dt_r}{d\tau} \rightarrow a^{-1} e^{\frac{t_r}{2R_s}} \quad (4.6.13)$$

Note that τ is the proper time of the source, while t_r can be identified with the proper time of the observer, since $R \gg R_s$. This means that the rate of reception of photons by the observer is exponentially damped, resulting in an exponential decrease of the observed intensity of the source as it approaches $r = R_s$. Additionally, the light emitted by the source is exponentially redshifted¹, near the horizon: combining these effects, the source eventually becomes undetectable, but the radiation intensity never drops to zero, although the source crosses the horizon in a finite proper time (as per Eq. 4.6.12).

§4.6.3 Classical tests of General Relativity

§4.6.3.1 Gravitational redshift

Consider a stationary clock, i.e. a clock on a spacetime section $\{(r, \vartheta, \varphi) = \text{const.}\}$. Then, the time measured by it is:

$$d\tau = dt \sqrt{1 - \frac{2m}{r}} \quad (4.6.14)$$

Then, if two stationary clocks at $r_1 = r$ and $r_2 = r + \Delta r$ exchange light signals, the non-Euclidean nature of spacetime determines a gravitational redshift:

$$\frac{\nu_2}{\nu_1} = \sqrt{\frac{1 - R_s/r_1}{1 - R_s/r_2}} \simeq 1 - \frac{R_s \Delta r}{2r^2} \quad (4.6.15)$$

where the approximation holds for $\Delta r \ll r$ and $R_s \ll r$.

§4.6.3.2 Periastron precession

The perihelion shift of Mercury was, at the start of the 20th century, one of the main unsolved problem of celestial mechanics: even after the perturbing effects of other planets on Mercury's orbit had been accounted for, there remained in the data an unexplained advance in its perihelion. This precession is explained by General Relativity.

¹Requires proof.

To study in general the precession of the periastron of a generic planet, Eq. 4.6.9 is treated perturbatively. In the Newtonian limit, the Keplerian orbit is an ellipse:

$$u(\varphi) = \frac{1}{p} (1 + e \cos \varphi) \quad p = a (1 + e^2) = \frac{\ell^2}{m} \quad (4.6.16)$$

where e is the eccentricity of the orbit and a its greater semiaxis. Then, at first order of approximation, this solution is inserted in the perturbative term, yielding:

$$u'' + u = \frac{m}{\ell^2} + \frac{3m^3}{\ell^4} (1 + e \cos \varphi)^2$$

Lemma 4.6.1

For a constant $A \in \mathbb{R}$, the following ODEs have particular solutions:

$$v'' + v = \begin{cases} A \\ A \cos \varphi \\ A \cos^2 \varphi \end{cases} \implies v(\varphi) = \begin{cases} A \\ \frac{1}{2}A\varphi \sin \varphi \\ \frac{1}{2}A - \frac{1}{6}A \cos 2\varphi \end{cases} \quad (4.6.17)$$

Using Lemma 4.6.1, it is trivial to see that the perturbative term leads to the following relativistic correction to the Keplerian elliptic orbit:

$$u(\varphi) = \frac{m}{\ell^2} (1 + e \cos \varphi) + \frac{3m^3}{\ell^4} \left[1 + \frac{e^2}{2} - \frac{e^2}{6} \cos 2\varphi + e\varphi \sin \varphi \right] \quad (4.6.18)$$

with an initial periastron at $\varphi = 0$. The important term, in this perturbation, is the *secular term* proportional to $\varphi \sin \varphi$, which is responsible (at first-order approximation) for the periastron precession.

Proposition 4.6.3 (Periastron precession)

If $\varphi = 0$ is a periastron, then the next periastron is at $\varphi = 2\pi + \delta\varphi$, with:

$$\delta\varphi \simeq 3\pi \frac{R_s}{a(1-e^2)} \quad (4.6.19)$$

Proof. The periastron is determined by the condition $u' = 0$, i.e.:

$$\sin \varphi = \frac{3m^3}{\ell^2} \left[\frac{e}{3} \sin 2\varphi + \sin \varphi + \varphi \cos \varphi \right]$$

Introducing the periastron anomaly $\varphi = 2\pi + \delta\varphi$, this becomes:

$$\sin \delta\varphi = \frac{3m^3}{\ell^2} \left[\frac{e}{3} \sin 2\delta\varphi + \sin \delta\varphi + \delta\varphi \cos \delta\varphi + 2\pi \cos \delta\varphi \right]$$

Now, assume that $e \ll 1$ and $|\delta\varphi| \ll 1$: then, $|\sin \delta\varphi| \ll 1$ and $\cos \delta\varphi \simeq 1$, so that the only term which determines the anomaly is, approximately, the last one:

$$\sin \delta\varphi \simeq 6\pi \frac{m^3}{\ell^2} \cos \delta\varphi$$

Then, using $\tan \delta\varphi \simeq \delta\varphi$, $2m \equiv R_s$ and Eq. 4.6.16, the thesis is obtained. \square

Although an approximation valid for $e \ll 1$ and $|\delta\varphi| \ll 1$, this expression is incredibly accurate for predictions in the solar system. Defining the rate of periastron shift:

$$\dot{\omega} \equiv \frac{\delta\varphi}{T} \quad T = 2\pi\sqrt{\frac{a^3}{GM}} \quad (4.6.20)$$

where the orbital period is computed from Kepler's third law, the prediction for Mercury is $\dot{\omega} \simeq 42.98''$ per century, which is accurate at 0.1% error with respect to the observed data.

§4.6.3.3 Deflection of light

Recalling Eq. 4.6.5, the motion of lightrays is determined by $L = 0$, since for null curves $ds^2 = 0$:

$$\left(1 - \frac{2m}{r}\right)^{-1} E^2 - \left(1 - \frac{2m}{r}\right)^{-1} \dot{r}^2 - \frac{\ell^2}{r^2} = 0 \quad (4.6.21)$$

This equation results in a simpler orbital equation.

Proposition 4.6.4 (Motion of lightrays)

$$u'' + u = 3mu^2 \quad (4.6.22)$$

Proof. Eq. 4.6.21 can be rewritten, with $\dot{r} = -u'\ell$, as:

$$E^2 - (1 - 2mu) \ell^2 u^2 = \ell^2 u'^2 \implies u'^2 + u^2 = \frac{E^2}{\ell^2} + 2mu^3$$

Differentiating yields the thesis. \square

In the solar system, the perturbative term is relatively small, since:

$$\frac{3mu^2}{u} = \frac{3R_s}{2r} \leq \frac{R_s}{R_\odot} \simeq 4.24 \cdot 10^{-6}$$

The unperturbed solution, found neglecting the $3mu^2$ term, is a straight light-path:

$$u(\varphi) = \frac{1}{b} \sin \varphi \quad (4.6.23)$$

where b is the impact parameter, i.e. the distance between the asymptotic initial direction of the lightray and the center of the mass distribution which determines the curvature of spacetime. At first order in perturbation theory, then:

$$u'' + u = \frac{3m}{b^2} (1 - \cos^2 \varphi)$$

Solving this with unperturbed solution + particular solution yields the first-order solution:

$$u(\varphi) = \frac{1}{b} \sin \varphi + \frac{3m}{2b^2} \left(1 + \frac{1}{3} \cos 2\varphi\right) \quad (4.6.24)$$

For large r (small u), $|\varphi|$ is either small or close to π . For small φ , take $\sin \varphi \simeq \varphi$ and $\cos \varphi \simeq 1$, and define the asymptotic initial direction $\varphi_\infty \equiv \lim_{u \rightarrow 0} \varphi$, so that by Eq. 4.6.24:

$$\varphi_\infty = -\frac{2m}{b} \quad (4.6.25)$$

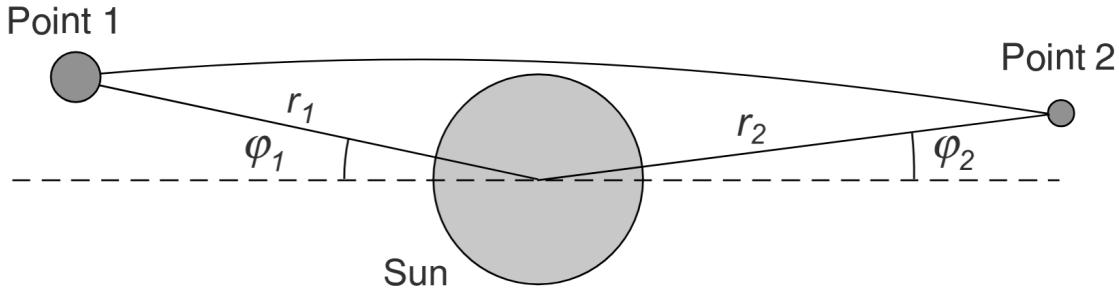


Figure 4.12: Path of a radar signal between two points in Schwarzschild spacetime.

For an observer at large distance, then, the **total deflection** is $\delta = 2|\varphi_\infty|$, that is:

$$\delta = \frac{2R_s}{b} \quad (4.6.26)$$

For the Sun, this results in $\delta \simeq 1.75''R_s/b$, which as of today is accurate within 0.02% from radioastronomical measurements of quasars: the experimental confirmation of this deviation was the first historical test of General Relativity.

§4.6.3.4 Gravitational time delay

Suppose that a signal is transmitted from a point (r_1, φ_1) to another point (r_2, φ_2) on the equatorial plane $\vartheta = \frac{\pi}{2}$. It is possible to compute the coordinate time required for this transmission.

Proposition 4.6.5 (Elapsed coordinate time)

Let $r_0 \in \mathbb{R}_{\geq R_s}^+$ be the distance of closest approach of a lightray. Then, the coordinate time elapsed during a transmission between r_0 and $r \geq r_0$ is:

$$\Delta t(r, r_0) = \int_{r_0}^r dr \left(1 - \frac{R_s}{r}\right)^{-1} \left[1 - \frac{1 - R_s/r}{1 - R_s/r_0} \left(\frac{r_0}{r}\right)^2\right]^{-\frac{1}{2}} \quad (4.6.27)$$

Assuming that $r, r_0 \gg R_s$, then:

$$\Delta t(r, r_0) \simeq \sqrt{r^2 - r_0^2} + R_s \ln \left[\frac{r + \sqrt{r^2 - r_0^2}}{r_0} \right] + \frac{R_s}{2} \sqrt{\frac{r - r_0}{r + r_0}} \quad (4.6.28)$$

Proof. First, express \dot{r} in terms of E :

$$\dot{r} = \frac{dr}{dt} = \frac{dr}{dt} \frac{E}{1 - R_s/r}$$

Then, recalling Eq. 4.6.21:

$$\left(1 - \frac{R_s}{r}\right)^{-3} \left(\frac{dr}{dt}\right)^2 = \left(1 - \frac{R_s}{r}\right)^{-1} - \frac{\ell^2}{E^2} \frac{1}{r^2}$$

If r_0 is the distance of closest approach, then $\frac{dr}{dt} \Big|_{r=r_0} = 0$, so that:

$$\frac{\ell^2}{E^2} = \frac{r_0^2}{1 - R_s/r_0}$$

Substituting this into the above equation and integrating yields the thesis. Then, if $r, r_0 \gg R_s$:

$$\Delta t(r, r_0) \simeq \int_{r_0}^r dr \left(1 + \frac{R_s}{r}\right) \left[1 - \left(\frac{R_s}{r_0} - \frac{R_s}{r}\right) \left(\frac{r_0}{r}\right)^2\right]^{-\frac{1}{2}}$$

This expression can be rearranged as:

$$\Delta t(r, r_0) \simeq \int_{r_0}^r dr \left(1 - \frac{r_0^2}{r^2}\right)^{-\frac{1}{2}} \left[1 + \frac{R_s}{r} + \frac{R_s r_0}{2r(r+r_0)}\right]$$

Integrating gives the thesis. \square

Note that the first term in the approximate expression is the non-relativistic term, while the other terms give the correction determined by the non-Euclidean nature of the Schwarzschild spacetime.

Now, if $|\varphi_1 - \varphi_2| > \frac{\pi}{2}$ as in Fig. 4.12, the total elapsed coordinate time is $\Delta t_{1,2} = \Delta t(r_1, r_0) + \Delta t(r_2, r_0)$, hence the time delay due to non-Euclideanity is given by the **Shapiro delay in coordinate time**:

$$\begin{aligned} \Delta t &:= 2 \left[\Delta t(r_1, r_0) + \Delta t(r_2, r_0) - \sqrt{r_1^2 - r_0^2} - \sqrt{r_2^2 - r_0^2} \right] \\ &\simeq 2R_s \ln \left[\frac{(r_1 + \sqrt{r_1^2 - r_0^2})(r_2 + \sqrt{r_2^2 - r_0^2})}{r_0^2} \right] + R_s \left[\sqrt{\frac{r_1 - r_0}{r_1 + r_0}} + \sqrt{\frac{r_2 - r_0}{r_2 + r_0}} \right] \end{aligned} \quad (4.6.29)$$

Although the coordinate time is not observable, since clocks measure their proper time, the Shapiro delay in coordinate time still gives a general idea of the magnitude of the general relativistic time delay.

Example 4.6.1 (Earth–Mars time delay)

Consider a radar transmission from Earth to Mars and then back to Earth. The Shapiro delay in coordinate time is (with $r_0 \ll r_1, r_2$):

$$\Delta t \simeq 2R_s \left[\ln \frac{4r_1 r_2}{r_0^2} + 1 \right] \lesssim 240 \mu s$$

§4.6.3.5 Gravitational microlensing

A direct consequence of the deflection of light due to a mass density is the lensing effect of the latter on the light emitted from objects further along the line of sight.

It is important to distinguish between two kinds of gravitational lensing: the cosmological lensing, where the source is a distant quasar and the lens is galaxy, and the gravitational microlensing, where the source is a distant star and the lens is a concentrated and relatively

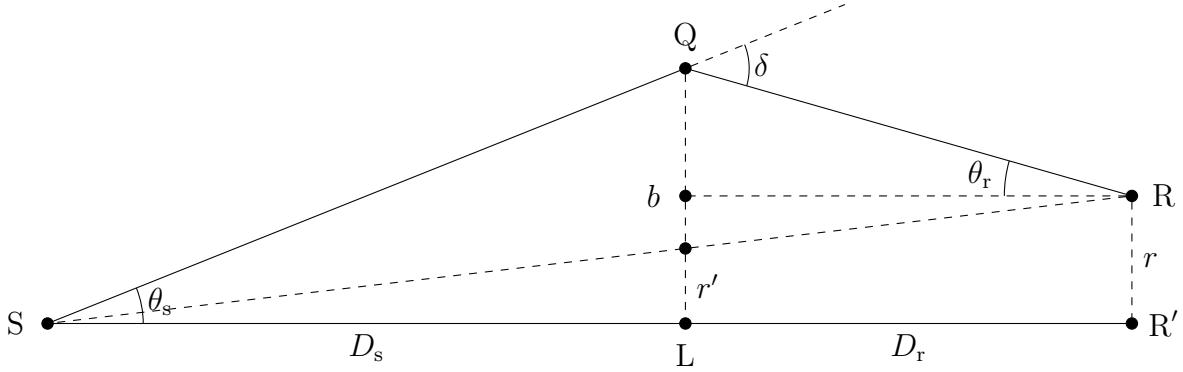


Figure 4.13: Schematics of the lens geometry.

small mass density. The former has a much more complex theory, due to the extended nature of the lens, hence only microlensing is considered here: the main experimental application of microlensing is the search for massive galactic halo objects (MACHOs, ~ 10 kpc), which act as lenses for stars in the Magellanic clouds (~ 50 kpc) and whose main candidates are brown dwarfs, inactive white dwarfs, neutron stars and black holes.

Optical schematics The optical scheme of a gravitational lens is composed of three points: the source S, the lens L and the observer R. Then, set R' the projection of R on the line SL and define the distances $D_s \equiv \overline{SL}$, $D_r \equiv \overline{LR'}$ and $r \equiv \overline{RR'}$, as in Fig. Fig. 4.13.

At first approximation, it is possible to identify the lightray emitted from the source and deflected by the lens with its two asymptotes, which are lines passing through a common point Q at a distance b from the lens and forming an angle δ given by Eq. 4.6.26, with $\overline{QL} \perp \overline{SL}$: these assumptions only induce errors of order $o(b^{-2})$, where $b, r \ll D_s, D_r$ is supposed.

Proposition 4.6.6 (Lensed image)

The image produced by the lens has dimension:

$$b_{\pm} = \frac{1}{2} \left[r' \pm \sqrt{r'^2 + 4b_E^2} \right] \quad (4.6.30)$$

with:

$$b_E \equiv \sqrt{2R_s \frac{D_s D_r}{D_s + D_r}} \quad r' \equiv \frac{D_s}{D_s + D_r} r \quad (4.6.31)$$

Proof. Geometrically, the following relations are clear (with $b \ll R_s$):

$$\theta_s + \theta_r = \delta \quad D_s \theta_s = b \quad D_r \theta_r = b - r$$

Substituting Eq. 4.6.26 and combining these equations, then:

$$\frac{b}{D_s} + \frac{b - r}{D_r} = \frac{2R_s}{b} \implies b^2 - \frac{D_s r}{D_s + D_r} b = 2R_s \frac{D_s D_r}{D_s + D_r}$$

Solving this equation yields the thesis. □

Note that the two solutions correspond to the lightray passing either “above” (like in Fig. 4.13)

or “below” the lens: this does not mean that the lens forms two images, since, for an image to form, it is necessary that a small beam of light passes through the same point after being lensed, but this does not happen in gravitational lenses, which are highly astigmatic.

While r' has a clear geometrical meaning, b_E corresponds to the case of a perfect alignment, i.e. $r = 0$. In this case, the lightrays which get lensed are not only the two already analyzed, which confined on the plane determined by S, L and R, but are infinitely many, and they lie on a cone with axis \overline{SR} and semiangle θ_r . In this case, the lens does not produce two images, but a whole circumference of radius b_E : an **Einstein ring**.

Note that these Einstein rings are nearly impossible to observe for relatively close sources: with $D_s = D_r = 1 \text{ kpc}$ and $R_s = 3 \text{ km}$, the angular semiopening of the ring is $\theta_r = b_E/D_r \simeq 0.002''$. On the other hand, these rings are fairly common on cosmological scales, e.g. with a galaxy as lens, but on these scales it is not possible to apply a Euclidean analysis.

Microlensing In the conditions of gravitational microlensing, observing lensed images or Einstein rings is impossible, as already seen. However, an observable effect is the magnification of the source. Assuming that the source emits isotropically a luminosity L , the power irradiated on an area $d\sigma$ in the absence of any lens is:

$$P_0 = \frac{L}{4\pi} \frac{d\sigma}{(D_s + D_r)^2}$$

On the other hand, if a lens is present, the power irradiated on $d\sigma$ is given by the luminosity emitted in the solid angle $d\Omega = \sin \theta_s d\theta_s d\varphi$, while $d\sigma = r dr d\varphi$ by axial symmetry, so that:

$$P = L \frac{d\Omega}{4\pi}$$

The magnification A is the ratio between P and P_0 , that is:

$$A = \frac{P}{P_0} = (D_s + D_r)^2 \frac{d\Omega}{d\sigma} \simeq (D_s + D_r)^2 \frac{\theta_s d\theta_s}{r dr} = \left(\frac{D_s + D_r}{D_s} \right)^2 \frac{b db}{r dr} = \frac{b db}{r' dr'}$$

Recalling Eq. 4.6.30, one must add the two contributions from each image, thus finding:

$$A = \frac{b_+^2}{r' |2b_+ - r'|} + \frac{b_-^2}{r' |2b_- - r'|}$$

where the absolute value is necessary in order to keep A positive. A trivial manipulation yields the final expression:

$$A = \frac{r'^2 + 2b_E^2}{2r' \sqrt{r'^2 + 4b_E^2}} \quad (4.6.32)$$

This expression seems to have a pole for $r' \rightarrow 0$: this is easily avoided by considering the source as an extended object, rather than a point.

Searching for these kinds of magnifications in the light of distant stars allows to study MACHOs: to distinguish the resulting luminous variability from variable stars, it is important to note that the microlensing effect is exactly symmetrical in time and non-dispersive, i.e. equal for all wavelengths.

Chapter 5

Weak Gravity

Although Einstein field equations are extremely difficult to solve, a possible ansatz is to consider an almost-flat metric with $\Lambda = 0$, which in the so called *almost-inertial coordinates* x^μ takes the form:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \quad (5.0.1)$$

where the perturbation of the metric is assumed to be small: $|h_{\mu\nu}| \ll 1$.

§5.1 Linearized gravity

The aim is to expand the field equations to linear order in $h_{\mu\nu}$: at this order, gravity can be thought as a symmetric spin 2 field $h_{\mu\nu}$ propagating through flat Minkowski spacetime. Therefore, indices are raised and lowered by Minkowski metric $\eta_{\mu\nu} = \text{diag}(-1, +1, +1, +1)$, and the field theory inherits Lorentz invariance:

$$x^\mu \mapsto \Lambda^\mu_\nu x^\nu \implies h^{\mu\nu} \mapsto \Lambda^\mu_\rho \Lambda^\nu_\sigma h^{\rho\sigma}(\Lambda^{-1}x)$$

where $h^{\mu\nu} = \eta^{\mu\rho} \eta^{\nu\sigma} h_{\rho\sigma}$. To leading order, the inverse metric is $g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu}$, thus:

$$\Gamma^\sigma_{\nu\rho} = \frac{1}{2} \eta^{\sigma\lambda} (\partial_\nu h_{\lambda\rho} + \partial_\rho h_{\nu\lambda} - \partial_\lambda h_{\nu\rho}) \quad (5.1.1)$$

Recalling Eq. 2.2.13, the $\Gamma\Gamma \sim h^2$ terms of the Riemann tensor are negligible to first order, so:

$$R^\sigma_{\rho\mu\nu} = \frac{1}{2} \eta^{\sigma\lambda} (\partial_\mu \partial_\rho h_{\nu\lambda} - \partial_\mu \partial_\lambda h_{\nu\rho} - \partial_\nu \partial_\rho h_{\mu\lambda} + \partial_\nu \partial_\lambda h_{\mu\rho}) \quad (5.1.2)$$

Contracting (σ, ρ) , the Ricci tensor is:

$$R_{\mu\nu} = \frac{1}{2} (\partial^\rho \partial_\mu h_{\nu\rho} + \partial^\rho \partial_\nu h_{\mu\rho} - \square h_{\mu\nu} - \partial_\mu \partial_\nu h) \quad (5.1.3)$$

where $\square := \partial^\mu \partial_\mu$ and $h \equiv h^\mu_\mu$ is the trace of $h_{\mu\nu}$. The Ricci scalar is:

$$R = \partial^\mu \partial^\nu h_{\mu\nu} - \square h \quad (5.1.4)$$

Finally, the Einstein tensor can be expressed as:

$$G_{\mu\nu} = \frac{1}{2} [\partial^\rho \partial_\mu h_{\nu\rho} + \partial^\rho \partial_\nu h_{\mu\rho} - \square h_{\mu\nu} - \partial_\mu \partial_\nu h - (\partial^\rho \partial^\sigma h_{\rho\sigma} - \square h) \eta_{\mu\nu}] \quad (5.1.5)$$

For linearized gravity, the Bianchi identity $\nabla^\mu G_{\mu\nu} = 0$ becomes $\partial^\mu G_{\mu\nu} = 0$, which is indeed obeyed by Eq. 5.1.5. Einstein field equations with a source $T_{\mu\nu}$, which, for consistency, must be suitably small, are then a set of linear PDEs:

$$\partial^\rho \partial_\mu h_{\nu\rho} + \partial^\rho \partial_\nu h_{\mu\rho} - \square h_{\mu\nu} - \partial_\mu \partial_\nu h - (\partial^\rho \partial^\sigma h_{\rho\sigma} - \square h) \eta_{\mu\nu} = 16\pi G T_{\mu\nu} \quad (5.1.6)$$

This can be thought as $\mathcal{L}(h_{\mu\nu}) = 16\pi G T_{\mu\nu}$, where \mathcal{L} is a linear differential operator known as **Lichnerowicz operator**.

Proposition 5.1.1 (Fierz–Pauli action)

Eq. 5.1.6 are the equations of motion derived from the **Fierz–Pauli action**:

$$\mathcal{S}_{\text{FP}} = \frac{1}{8\pi G} \int d^4x \left[-\frac{1}{4} \partial_\rho h_{\mu\nu} \partial^\rho h^{\mu\nu} + \frac{1}{2} \partial_\rho h_{\mu\nu} \partial^\nu h^{\rho\mu} + \frac{1}{4} \partial_\mu h \partial^\mu h - \frac{1}{2} \partial_\nu h^{\mu\nu} \partial_\mu h \right] \quad (5.1.7)$$

Proof. Varying the action:

$$\begin{aligned} \delta \mathcal{S}_{\text{FP}} &= \frac{1}{8\pi G} \int d^4x \left[\frac{1}{2} \partial_\rho \partial^\rho h_{\mu\nu} - \partial^\rho \partial_\nu h_{\rho\mu} - \frac{1}{2} \eta_{\mu\nu} \partial^\rho \partial_\rho h + \frac{1}{2} \partial_\nu \partial_\mu h + \frac{1}{2} \eta_{\mu\nu} \partial_\rho \partial_\sigma h^{\rho\sigma} \right] \delta h^{\mu\nu} \\ &= \frac{1}{8\pi G} \int d^4x [-G_{\mu\nu} \delta h^{\mu\nu}] \end{aligned}$$

Hence, $G_{\mu\nu} = 0$. To get the matter coupling, add $T_{\mu\nu} h^{\mu\nu}$ to the action. \square

§5.1.1 Gauge symmetry

Linearized gravity inherits a useful gauge symmetry from the diffeomorphism invariance of the full theory. Under a change of coordinates $x^\mu \mapsto x^\mu - \xi^\mu(x)$, where $\xi(x)$ is assumed to be small, the metric changes by Eq. 4.1.11 as $\delta g_{\mu\nu} = (\mathcal{L}_\xi g)_{\mu\nu} = \nabla_\mu \xi_\nu + \nabla_\nu \xi_\mu$; for the linearized metric Eq. 5.0.1, being both h and ξ small, the covariant derivatives take the vanishing connection of Minkowski spacetime, thus:

$$h_{\mu\nu} \mapsto h_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu \quad (5.1.8)$$

This is similar to the gauge transformation of Maxwell theory $A_\mu \mapsto A_\mu + \partial_\mu \alpha$: just like $F_{\mu\nu} = 2\partial_{[\mu} A_{\nu]}$ is gauge-invariant, so is the linearized Riemann tensor Eq. 5.1.2.

Proposition 5.1.2

The Fierz–Pauli action is invariant under the gauge symmetry Eq. 5.1.8.

Proof. Recalling the linearized Bianchi identity $\partial^\mu G_{\mu\nu} = 0$:

$$\delta \mathcal{S}_{\text{FP}} = -\frac{1}{8\pi G} \int d^4x 2G_{\mu\nu} \partial^\mu \xi^\nu = \frac{1}{4\pi G} \int d^4x \partial^\mu G_{\mu\nu} \xi^\nu = 0$$

which shows the invariance. \square

As in Electromagnetism, it's useful to impose a gauge fixing condition.

Proposition 5.1.3 (de Donder gauge)

It is always possible to pick **de Donder gauge**:

$$\partial^\mu h_{\mu\nu} - \frac{1}{2}\partial_\nu h = 0 \quad (5.1.9)$$

Proof. Suppose that the metric perturbation does not obey de Donder condition, but WLOG $\partial^\mu h_{\mu\nu} - \frac{1}{2}\partial_\nu h = f_\nu$ for some functions f_ν . After the gauge transformation Eq. 5.1.8, this becomes $\partial^\mu h_{\mu\nu} - \frac{1}{2}\partial_\nu h + \square\xi_\nu = f_\nu$, thus one only needs to find $\xi_\nu : \square\xi_\nu = f_\nu$, which always has a solution. \square

In de Donder gauge, the field equations Eq. 5.1.6 are greatly simplified:

$$\square h_{\mu\nu} - \frac{1}{2}\eta_{\mu\nu}\square h = -16\pi GT_{\mu\nu} \quad (5.1.10)$$

To simplify these equations even more, it is useful to define:

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

as $\bar{h} = -h$. With this choice, the linearized Einstein equations in de Donder gauge reduce to a set of wave equations:

$$\square\bar{h}_{\mu\nu} = -16\pi GT_{\mu\nu} \quad (5.1.11)$$

Non-linear theory de Donder gauge can be extended to the full non-linear theory as the condition:

$$g^{\mu\nu}\Gamma_{\mu\nu}^\rho = 0 \quad (5.1.12)$$

Note that this is not a tensor equation, since $\Gamma_{\mu\nu}^\rho$ is not a tensor, and indeed the point of gauge fixing is to set a preferred choice of coordinates. This gauge condition simplifies the expression of the d'Alembertian $\square := \nabla^\mu\nabla_\mu = g^{\mu\nu}(\partial_\mu\partial_\nu - \Gamma_{\mu\nu}^\rho\partial_\rho)$, which simply becomes $\square = g^{\mu\nu}\partial_\mu\partial_\nu$; moreover, the same applies to 1-forms: $\nabla^\mu\omega_\mu = g^{\mu\nu}\nabla_\mu\omega_\nu = g^{\mu\nu}(\partial_\mu\omega_\nu - \Gamma_{\mu\nu}^\rho\omega_\rho) = \partial^\mu\omega_\mu$.

§5.1.2 Newtonian limit

In the presence of a low-density, slowly-moving distribution of matter, the linearized field equations reduce to the Newtonian theory of gravity. For a stationary matter configuration, the only non-vanishing component of the energy-momentum tensor is $T_{00} = \rho(\mathbf{x})$; moreover, since there is no time-dependence, $\square = -\partial_t^2 + \nabla^2 = \nabla^2$. The Einstein equations become:

$$\nabla^2\bar{h}_{00} = -16\pi G\rho(\mathbf{x}) \quad \nabla^2\bar{h}_{0i} = \nabla^2\bar{h}_{ij} = 0$$

With suitable boundary conditions, the solutions to these equations are:

$$\bar{h}_{00} = -4\Phi(\mathbf{x}) \quad \bar{h}_{0i} = \bar{h}_{ij} = 0$$

where $\Phi : \nabla^2\Phi = 4\pi G\rho$ is the Newtonian gravitational potential. Then $\bar{h} = 4\Phi$ and:

$$h_{\mu\nu} = -2\Phi(\mathbf{x})\delta_{\mu\nu}$$

The full metric $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$ is thus expressed as:

$$ds^2 = -(1 + 2\Phi(\mathbf{x}))dt^2 + (1 - 2\Phi(\mathbf{x}))d\mathbf{x}^2 \quad (5.1.13)$$

This is exactly the condition Eq. 3.1.10. Interestingly, for a point particle $\Phi(\mathbf{x}) = -\frac{GM}{r}$ and the metric coincides with the leading expansion of Schwarzschild metric (the g_{00} term is exact).

§5.2 Gravitational waves

To study the propagation of gravitational waves in vacuum and in the absence of sources, one needs to solve the linearized wave equation:

$$\square \bar{h}_{\mu\nu} = 0 \quad (5.2.1)$$

A possible solution is the gravitational wave:

$$\bar{h}_{\mu\nu}(x) = \Re\{H_{\mu\nu}e^{ik_\rho x^\rho}\} \quad (5.2.2)$$

where $H_{\mu\nu} \in \mathbb{C}^{4 \times 4}$ is a symmetric polarization matrix and the wave-vector k^μ is a real 4-vector. For simplicity, the \Re is made implicit in the following calculations. This plane-wave ansatz solves Eq. 5.2.1 if the wave-vector is null: $k_\mu k^\mu = 0$: therefore, gravitational waves propagate at the speed of light. Writing $k^\mu = (\omega, \mathbf{k})$, this condition becomes $\omega = \pm |\mathbf{k}|$. Moreover, being the vacuum wave equation linear, the general solution is just a linear combination of plane waves. The polarization matrix has 10 components, but gauge conditions are yet to be imposed. The plane wave ansatz satisfies de Donder gauge condition $\partial^\mu \bar{h}_{\mu\nu} = 0$ only if:

$$k^\mu H_{\mu\nu} = 0 \quad (5.2.3)$$

The polarization is then transverse to the direction of propagation. Furthermore, de Donder gauge still allows for gauge transformations $h_{\mu\nu} \mapsto h_{\mu\nu} \partial_\mu \xi_\nu + \partial_\nu \xi_\mu$, i.e. $\bar{h}_{\mu\nu} \mapsto \bar{h}_{\mu\nu} + \partial_\mu \xi_\nu + \partial_\nu \xi_\mu - \eta_{\mu\nu} \partial^\rho \xi_\rho$, so the solution is still in de Donder gauge only if:

$$\square \xi_\mu = 0 \quad \Rightarrow \quad \xi_\mu(x) = \lambda_\mu e^{ik_\rho x^\rho}$$

This gauge transformation shifts the polarization matrix as:

$$H_{\mu\nu} \mapsto H_{\mu\nu} + i(k_\mu \lambda_\nu + k_\nu \lambda_\mu - \eta_{\mu\nu} k^\rho \lambda_\rho) \quad (5.2.4)$$

Polarization matrices which differ by this term thus describe the same gravitational wave. Hence, it is possible to choose λ_μ in order to have:

$$H_{0\mu} = 0 \quad \wedge \quad H^\mu{}_\mu = 0 \quad (5.2.5)$$

These condition, together with Eq. 5.2.3, are called **transverse traceless gauge**. Being $H_{\mu\nu}$ traceless, in this gauge $\bar{h}_{\mu\nu} = h_{\mu\nu}$, as it is traceless too. Of the 10 components of the polarization matrix, only 2 are independent: de Donder condition Eq. 5.2.3 poses 4 constraints and the gauge transformation Eq. 5.2.4 poses 4 more of them (since one of the $H_{0\mu} = 0$ conditions is made redundant by $k^\mu H_{\mu\nu} = 0$). Therefore, there are only two independent polarizations in $H_{\mu\nu}$.

Example 5.2.1 (Linear polarizations)

Consider a gravitational wave propagating in the z direction: $k^\mu = (\omega, 0, 0, \omega)$, thus $H_{0\nu} + H_{3\nu} = 0$ by Eq. 5.2.3. By Eq. 5.2.5, the polarization matrix is then restricted to be:

$$H_{\mu\nu} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & H_+ & H_\times & 0 \\ 0 & H_\times & -H_+ & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (5.2.6)$$

where in general $H_+, H_\times \in \mathbb{C}$. The two polarizations are seen explicitly.

§5.2.1 Polarizations

Point particles moving along geodesics are not affected by gravitational waves due to the equivalence principle: to measure their passage, one needs to study how the relative distance between two observers changes, which is done using the geodesic deviation (recall §2.3.3).

Consider a family of geodesics $x^\mu(\tau, s)$ with tangent vector field $u^\mu = \partial_\tau|_s x^\mu$ and deviation vector field $S^\mu = \partial_s|_\tau x^\mu$. The geodesic deviation equation Eq. 2.3.8 reads:

$$\frac{D^2 S^\mu}{D\tau^2} = R^\mu_{\rho\sigma\nu} u^\rho u^\sigma S^\nu$$

Suppose that, in the absence of gravitational waves, the geodesics are in a rest-frame such that $u^\mu = (1, 0, 0, 0)$: as the gravitational wave passes $u^\mu = (1, 0, 0, 0) + o(h)$, so the aim is to compute the geodesic deviation to leading order in h . The Riemann tensor is already $o(h)$, thus other correction terms can be neglected; similarly, proper time τ can be replaced with coordinate time t , so that the geodesic deviation equation becomes $\frac{d^2 S^\mu}{dt^2} = R^\mu_{00\nu} S^\nu$.

In the linearized regime, the Riemann tensor is given by Eq. 5.1.2, hence, using $h_{\mu 0} = 0$, the needed components are $R^\mu_{00\nu} = \frac{1}{2} \partial_0^2 h^\mu_\nu$ and:

$$\frac{d^2 S^\mu}{dt^2} = \frac{1}{2} \frac{d^2 h^\mu_\nu}{dt^2} S^\nu \quad (5.2.7)$$

For simplicity, consider a wave propagating in the z direction with polarization matrix Eq. 5.2.6 and solve the geodesic deviation equation in the $z = 0$ plane only, as S^0 and S^3 are not affected by the gravitational wave.

§5.2.1.1 + polarization

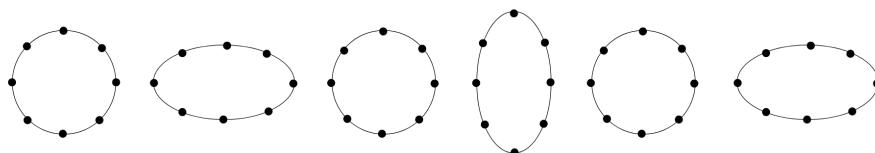
Setting $H_x = 0$, Eq. 5.2.7 becomes:

$$\frac{d^2 S^1}{dt^2} = -\frac{\omega^2}{2} H_+ e^{i\omega t} S^1 \quad \frac{d^2 S^2}{dt^2} = +\frac{\omega^2}{2} H_+ e^{i\omega t} S^2$$

Solving perturbatively in H_+ , at leading order:

$$S^1(t) = S^1(0) \left[1 + \frac{1}{2} H_+ e^{i\omega t} + \dots \right] \quad S^2(t) = S^2(0) \left[1 - \frac{1}{2} H_+ e^{i\omega t} + \dots \right] \quad (5.2.8)$$

where, again, the \Re is implicit (recall that in general $H_+ \in \mathbb{C}$). To visualize these solutions, consider a family of neighbouring geodesics which, at $t = 0$, are arranged around a circle of radius R : the initial conditions then satisfy $S^1(0)^2 + S^2(0)^2 = R^2$. The relative negative sign determines:



§5.2.1.2 \times polarization

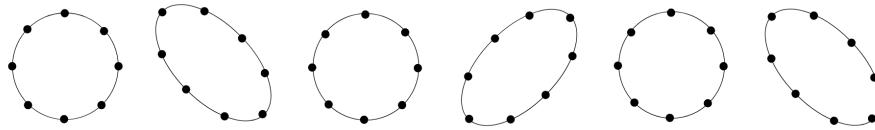
Setting $H_x = 0$, Eq. 5.2.7 becomes:

$$\frac{d^2 S^1}{dt^2} = -\frac{\omega^2}{2} H_x e^{i\omega t} S^2 \quad \frac{d^2 S^2}{dt^2} = -\frac{\omega^2}{2} H_x e^{i\omega t} S^1$$

Solving perturbatively in H_x , at leading order:

$$S^1(t) = S^1(0) + \frac{1}{2} S^2(0) H_x e^{i\omega t} + \dots \quad S^2(t) = S^2(0) + \frac{1}{2} S^1(0) H_x e^{i\omega t} + \dots \quad (5.2.9)$$

This is the same displacement as Eq. 5.2.8, but rotated by 45° : to see this, note that $S^1(t) \pm S^2(t)$ have the same functional expression as Eq. 5.2.8. Thus:



The general polarization is a linear combination of both: the result is yet an elliptic displacement whose axes rotate, analogously to the circular polarization of light. Interestingly, note that displacements due to gravitational waves are invariant under π rotations, while light polarizations, being described by vectors, are invariant under 2π rotations: this reflects the fact that the graviton has spin 2 and the photon has spin 1.

§5.2.1.3 Gravitational wave detection

Gravitational wave detectors are interferometers, which bounce light back and forth between two arms. If the gravitational wave propagates perpendicular to the plane of detection, it will shorten one arm and lengthen the other: assuming the arms are aligned with x and y axes, the maximum change in length by Eq. 5.2.8 is $L' = L(1 \pm H_+/2)$, i.e. $\delta L/L = H_+/2$. For a typical astrophysical source $H_+ \sim 10^{-21}$, while for LIGO $L \sim 3$ km, thus $\delta L \sim 10^{-18}$ m: this is smaller than the radius of the proton and extremely difficult to detect, however the first direct measurement of gravitational waves was performed in 2015 and now LIGO and VIRGO detectors have observed a large number of mergers involving black holes and neutron stars.

§5.2.2 Exact solutions

Given the found wave-like solution to the linearized field equations, the metric of a wave moving in the positive z direction takes the form:

$$ds^2 = -dt^2 + (\delta_{ab} + h_{ab}(z-t)) dx^a dx^b + dz^2 \quad (5.2.10)$$

with $a, b = 1, 2$. Being the wave equation linear, any function $h_{ab}(z-t)$ is a solution: Eq. 5.2.2 is simply the Fourier decomposition of the general solution.

The extreme weakness of gravitational waves makes the linearized metric Eq. 5.0.1 suitable to describe their properties. However, the wave solution has an extension to the general non-linear field equations. Consider a wave propagating in the positive z direction and introduce lightcone coordinates:

$$u = t - z \quad v = t + z$$

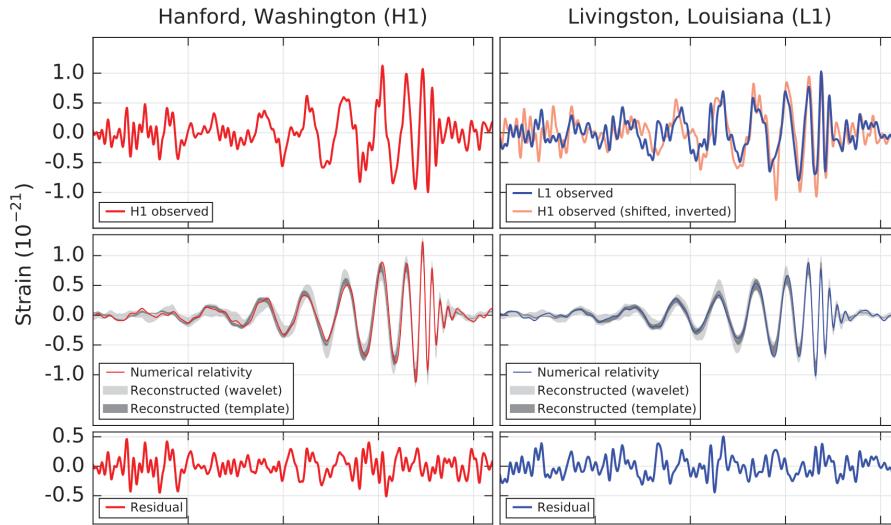


Figure 5.1: First detection of gravitational waves by LIGO.

Then, consider the plane wave ansatz, also called **Brinkmann metric**:

$$ds^2 = -du dv + dx^a dx^a + H_{ab}(u) x^a x^b du^2$$

Note that to obtain the linearized metric Eq. 5.2.10 from the Brinkmann metric, one needs some change of coordinates. It is possible to show that the Brinkmann metric is Ricci flat for any traceless matrix $H_{ab}(u)$, hence solving the vacuum Einstein equations:

$$R_{\mu\nu} = 0 \iff H^a_a(u) = 0 \iff H_{ab}(u) = \begin{bmatrix} H_{11}(u) & H_{12}(u) \\ H_{12}(u) & -H_{11}(u) \end{bmatrix}$$

The general metric has again two independent polarization states.

§5.3 Perturbing spacetime

The gravitational plane wave solution Eq. 5.2.2 is not realistic, as it comes from infinity and goes to infinity: in reality, gravitational waves are produced at some point and radiate out. The analogy with Electromagnetism persists: as electromagnetic waves are produced by oscillating charges, gravitational waves are produced by oscillating masses.

§5.3.1 Green's function

Recall the linearized Einstein field equations Eq. 5.1.11:

$$\square \bar{h}_{\mu\nu} = -16\pi G T_{\mu\nu}$$

which assumes that both $h_{\mu\nu}$ and $T_{\mu\nu}$ are small. These are a set of decoupled wave equations. Consider a matter field localized in a spatial region Σ , in which a time-dependent energy-momentum source $T_{\mu\nu}(\mathbf{x}, t)$ is present (e.g. two orbiting black holes): the localization condition then is simply $T_{\mu\nu}(\mathbf{x}, t) = 0 \forall \mathbf{x} \notin \Sigma$. The question is what the metric $h_{\mu\nu}$ looks like far away from Σ . The solution to Eq. 5.1.11 outside of Σ can be expressed using the retarded Green's function:

$$\bar{h}_{\mu\nu}(\mathbf{x}, t) = 4G \int_{\Sigma} d^3x' \frac{T_{\mu\nu}(\mathbf{x}', t_r)}{|\mathbf{x} - \mathbf{x}'|} \quad t_r \equiv t - |\mathbf{x} - \mathbf{x}'| \quad (5.3.1)$$

Retarded time expresses the causality of the wave equation. This solution satisfies de Donder condition $\partial^\mu \bar{h}_{\mu\nu} = 0$ only if $\partial^\mu T_{\mu\nu} = 0$, i.e. the energy-momentum tensor is conserved; however, it does not automatically satisfy conditions Eq. 5.2.5.

Denoting the characteristic size of Σ as d and $r \equiv |\mathbf{x}|$, then approximating:

$$|\mathbf{x} - \mathbf{x}'| \gg d \quad \forall \mathbf{x}' \in \Sigma \quad \Rightarrow \quad |\mathbf{x} - \mathbf{x}'| = r - \frac{\mathbf{x} \cdot \mathbf{x}'}{r} + \dots \quad \Rightarrow \quad \frac{1}{|\mathbf{x} - \mathbf{x}'|} = \frac{1}{r} + \frac{\mathbf{x} \cdot \mathbf{x}'}{r^3} + \dots$$

To approximate the energy-momentum tensor, assume that the motion of matter is non-relativistic, so that $T_{\mu\nu}$ does not change much over the time $\tau \sim d$ needed to light to cross Σ . For example, in the case of two objects orbiting each other with characteristic frequency ω , then $T_{\mu\nu} \sim e^{-i\omega t}$ and the non-relativistic condition reads $d \ll 1/\omega$. Taylor-expanding $T_{\mu\nu}$:

$$T_{\mu\nu}(\mathbf{x}', t_r) = T_{\mu\nu}(\mathbf{x}', t - r + \mathbf{x} \cdot \mathbf{x}'/r + \dots) = T_{\mu\nu}(\mathbf{x}', t - r) + \dot{T}_{\mu\nu}(\mathbf{x}', t - r) \frac{\mathbf{x} \cdot \mathbf{x}'}{r} + \dots$$

At leading order in d/r , then, the solution becomes:

$$\bar{h}_{\mu\nu}(\mathbf{x}, t) \approx \frac{4G}{r} \int_{\Sigma} d^3x' T_{\mu\nu}(\mathbf{x}', t - r)$$

The temporal component is:

$$\bar{h}_{00}(\mathbf{x}) \approx \frac{4G}{r} E \quad E \equiv \int_{\Sigma} d^3x' T_{00}(\mathbf{x}', t - r)$$

This is simply the Newtonian limit; note that there is no time-dependency, as $\partial^\mu T_{\mu\nu} = 0$ ensures that the energy E inside Σ is conserved. Similarly:

$$\bar{h}_{0i}(\mathbf{x}) \approx -\frac{4G}{r} P_i \quad P_i \equiv - \int_{\Sigma} d^3x' T_{0i}(\mathbf{x}', t - r)$$

Again, the total momentum of matter inside Σ is conserved, hence the absence of time-dependence. Note that it is always possible to choose a rest-frame where matter is stationary, i.e. $P_i = 0$ and $h_{0i} = 0$. The motion of matter inside Σ is instead described by \bar{h}_{ij} .

Proposition 5.3.1 (Quadrupole momentent)

Far away from the source, the spatial metric takes the form:

$$\bar{h}_{ij}(\mathbf{x}, t) \approx \frac{2G}{r} \ddot{I}_{ij}(t - r) \quad I_{ij}(t) := \int_{\Sigma} d^3x T^{00}(\mathbf{x}, t) x_i x_j \quad (5.3.2)$$

where $I_{ij}(t)$ is the **quadrupole moment** for the energy.

Proof. The thesis is equivalent to:

$$\int_{\Sigma} d^3x' T_{ij}(\mathbf{x}', t) = \frac{1}{2} \ddot{I}_{ij}(t)$$

Recalling current conservation $\partial_{\mu} T^{\mu\nu} = 0$:

$$T^{ij} = \partial_k(T^{ik}x^j) - (\partial_k T^{ik})x^j = \partial_k(T^{ik}x^j) + \partial_0 T^{0i}x^j$$

$$\begin{aligned} T^{0i} &= \partial_k(T^{0k}x^i) - (\partial_k T^{0k})x^i = \partial_k(T^{0k}x^i) + \partial_0 T^{00}x^i \\ &\implies T^{0(i}x^{j)} = \frac{1}{2}\partial_k(T^{0k}x^i x^j) + \frac{1}{2}\partial_0 T^{00}x^i x^j \end{aligned}$$

Integrating the first over Σ , recalling that $T^{ij} = T^{(ij)}$ and dropping the total spatial derivatives yields the result. \square

The physical meaning of Eq. 5.3.2 is that if a matter distribution is shaken, then, after the time needed for the signal to propagate, it will affect the metric. Given the linearity of these equations, if matter oscillates at a frequency ω , then spacetime will create waves at parametrically the same frequency.

The gauge condition $\partial^{\mu}\bar{h}_{\mu\nu} = 0$ means that $\partial_0\bar{h}_{0i} = \partial_j\bar{h}_{ji}$ and $\partial_0\bar{h}_{00} = \partial_i\bar{h}_{ii}$. The first equation gives:

$$\partial_0\bar{h}_{0i} = -\frac{2G\hat{x}_j}{r^2} \ddot{I}_{ij}(t - r) - \frac{2G\hat{x}_j}{r} \ddot{I}_{ij}(t - r)$$

where $\partial_j r = x_j/r \equiv \hat{x}_j$. Note that, although the first term is $\sim 1/r^2$ and the second $\sim 1/r$, the latter has an additional time derivative, i.e. an extra factor of the characteristic source frequency ω : therefore, the second term dominates for $r \gg 1/\omega$, i.e. $r \gg \lambda$ (emitted gravitational waves' wavelength). This is the so-called *far-field zone* or **radiation zone**, and in this regime:

$$\bar{h}_{0i} \approx -\frac{4G}{r} P_i - \frac{2G\hat{x}_j}{r} \ddot{I}_{ij}(t - r) \quad (5.3.3)$$

This expression, found integrating the preceding equation, contains an integration constant given by the aforementioned P_i , but it can be set to zero choosing coordinates in which the center of mass of the system is stationary. With analogous reasoning:

$$\bar{h}_{00} \approx \frac{4G}{r} E + \frac{2G\hat{x}_i\hat{x}_j}{r} \ddot{I}_{ij}(t - r) \quad (5.3.4)$$

Example 5.3.1 (Gravitational waves in binary system)

Consider two objects of mass M and separated by a distance R , orbiting in the (x, y) plane:

by Newtonian gravity, the orbit frequency is $\omega^2 = 2GM/R^3$. Treating these objects as point particles, their energy density is:

$$T^{00}(\mathbf{x}, t) = M\delta(z) \left[\delta(x - \frac{R}{2}\cos\omega t) \delta(y - \frac{R}{2}\sin\omega t) + \delta(x + \frac{R}{2}\cos\omega t) \delta(y + \frac{R}{2}\sin\omega t) \right]$$

The quadrupole moment then is:

$$I_{ij}(t) = \frac{MR^2}{2} \begin{bmatrix} \cos^2\omega t & \cos\omega t \sin\omega t & 0 \\ \cos\omega t \sin\omega t & \sin^2\omega t & 0 \\ 0 & 0 & 0 \end{bmatrix} = \frac{MR^2}{4} \begin{bmatrix} 1 + \cos 2\omega t & \sin 2\omega t & 0 \\ \sin 2\omega t & 1 - \cos 2\omega t & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

The resulting metric perturbation is:

$$\bar{h}_{ij}(\mathbf{x}, t) \approx -\frac{2GMR^2\omega^2}{r} \begin{bmatrix} \cos 2\omega t_r & \sin 2\omega t_r & 0 \\ \sin 2\omega t_r & -\cos 2\omega t_r & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

with $t_r = t - r$. This gravitational wave propagates approximately radially. To estimate its expected strength:

$$|h_{ij}| \sim \frac{2GMR^2\omega^2}{r} \sim \frac{G^2M^2}{Rr}$$

Thus, the signal is largest for large masses orbiting very tightly. The most dense objects are black holes, for which $R_s = 2GM$, so considering two black holes orbiting at $R \approx R_s$ then $|h_{ij}| \sim GM/r \sim R_s/r$: black holes of a few solar masses have $R_s \sim 10$ km, so if they are situated in a galaxy a billion light-years away from Earth $r \sim 10^{21}$ km, hence $|h| \sim 10^{-20}$.

§5.3.1.1 Multipole expansion

Comparing to Electromagnetism, recall the first multipoles of the charge distribution $\rho(\mathbf{x})$:

$$Q = \int_{\Sigma} d^3x \rho(\mathbf{x}) \quad \mathbf{p} = \int_{\Sigma} d^3x \rho(\mathbf{x}) \mathbf{x} \quad \mathcal{Q}_{ij} = \int_{\Sigma} d^3x \rho(\mathbf{x}) (3x_i x_j - \delta_{ij} x^2)$$

Charge conservation $\dot{Q} = 0$ nullifies the monopole contribution to electromagnetic waves. The leading order contribution is that of the dipole:

$$\mathbf{A}(\mathbf{x}, t) \approx \frac{\mu_0}{4\pi} \dot{\mathbf{p}}(t - r)$$

For gravity, the first multipoles of the energy distribution $T_{00}(\mathbf{x})$ are:

$$E = \int_{\Sigma} d^3x T_{00}(\mathbf{x}) \quad \mathbf{X} = \frac{1}{E} \int_{\Sigma} d^3x T_{00}(\mathbf{x}) \mathbf{x} \quad I_{ij} = \int_{\Sigma} d^3x T_{00}(\mathbf{x}, t) x_i x_j$$

i.e. the total energy, the center of mass (energy) of the distribution and the quadrupole. Energy conservation $\dot{E} = 0$ is responsible for the lack of monopole contribution to gravitational radiation. But in contrast to Electromagnetism, the dipole contribution also vanishes due to momentum conservation $\dot{\mathbf{P}} = \mathbf{0}$:

$$E \dot{X}_i = \int_{\Sigma} d^3x (\partial_0 T_{00}) x_i = \int_{\Sigma} d^3x (\partial_j T_{j0}) x_i = - \int_{\Sigma} d^3x T_{i0} = P_i \quad \Rightarrow \quad E \dot{\mathbf{X}} = \dot{\mathbf{P}} = \mathbf{0}$$

In Electromagnetism, another dipole contribution to the gauge potential is possible:

$$\mathbf{A}(\mathbf{x}, t) = -\frac{\mu_0}{4\pi r} \hat{\mathbf{x}} \times \dot{\mathbf{m}}(t-r) \quad \mathbf{m} := \frac{1}{2} \int_{\Sigma} d^3x \mathbf{x} \times \mathbf{J}(\mathbf{x})$$

For gravity, there's an equivalent contribution with the analogue of the magnetic dipole being:

$$J_i = \int_{\Sigma} d^3x \epsilon_{ijk} x_j T_{0k}$$

This is nothing other than the total angular momentum of the system, thus the dipole contribution to gravitational radiation is nullified by angular momentum conservation $\mathbf{J} = \mathbf{0}$ too. The leading order effect for gravitational waves is then the quadrupole.

§5.3.2 Radiated power

In the context of Electromagnetism, the power radiated as electromagnetic waves are emitted is easily found defining the Poynting vector $\mathbf{S} := \frac{1}{\mu_0} \mathbf{E} \times \mathbf{B}$ and, using the dipole approximation, finding the **Larmor formula**:

$$\mathcal{P} = \int_{\mathbb{S}^2} d^2\mathbf{r} \cdot \mathbf{S} = \frac{\mu_0}{6\pi c} |\ddot{\mathbf{p}}|^2$$

Analogously, a source emitting gravitational waves loses energy which is carried by them. The problem of finding an equivalent to Larmor formula is that there is no local energy-momentum tensor for gravitational fields, so there is no analogue to the Poynting vector. A possible way forward is to define an energy-momentum tensor $t_{\mu\nu}$ for gravitational waves which, in the linearized theory, obeys $\partial^\mu t_{\mu\nu} = 0$: Eq. 4.5.15 does not allow to do so in a diffeomorphism-invariant way, i.e. in the full non-linear theory $t_{\mu\nu}$ is not a tensor and in the linearized theory it is not invariant under gauge transformations Eq. 5.1.8. It turns out that there are a number of different ways to define such a tensor.

§5.3.2.1 Fierz–Pauli action

Recall the Fierz–Pauli action Eq. 5.1.7 for linearized gravity: viewed as an action describing a spin 2 field propagating through Minkowski spacetime, it can be treated as any classical field theory and used to compute an energy-momentum tensor. In the transverse traceless gauge, i.e. with $h = 0$ and $\partial^\mu h_{\mu\nu} = 0$, after integration by parts the action becomes:

$$\mathcal{S}_{\text{FP}} = -\frac{1}{8\pi G} \int d^4x \frac{1}{4} \partial_\rho h_{\mu\nu} \partial^\rho h^{\mu\nu}$$

This looks like the action for a set of massless scalar fields, hence the energy density takes the schematic form:

$$t^{00} \sim \frac{1}{G} \dot{h}_{\mu\nu} \dot{h}^{\mu\nu}$$

There are gradient terms too, but, due to the wave equation, they contribute as time derivatives. Note that t^{0i} scales in the same way. In the transverse traceless gauge, Eq. 5.3.2 becomes:

$$h_{ij}(\mathbf{x}, t) \sim \frac{G}{r} \ddot{\mathcal{Q}}(t-r) \quad \mathcal{Q}_{ij} \equiv I_{ij} - \frac{1}{3} I_{kk} \delta_{ij}$$

where \mathcal{Q}_{ij} is the traceless part of the quadrupole moment I_{ij} . The energy density carried by gravitational waves should then be:

$$t^{00} \sim \frac{G}{r^2} \ddot{\mathcal{Q}}_{ij}^2$$

Integrating over a sphere at large distance suggests that $\mathcal{P} \sim G \ddot{\mathcal{Q}}_{ij}^2$, and this is indeed correct, as it can be shown that:

$$\mathcal{P}(t) = \frac{G}{5} \ddot{\mathcal{Q}}_{ij}(t_r) \ddot{\mathcal{Q}}^{ij}(t_r) \quad (5.3.5)$$

This is the **quadrupole formula** and it is analogous to Larmor formula. Note that r is the distance at which the gravitational waves are observed.

§5.3.2.2 Conceptual issues

Although Eq. 5.3.5 is indeed correct, as the 1993 Nobel prize was awarded for data in agreement with it, there remains some conceptual issues in the definition of $t_{\mu\nu}$, which can be improved. First, note that the definition of $t_{\mu\nu}$ from the Fierz–Pauli action suffers a number of ambiguities: if one attempts to compute it as the Noether currents associated to spacetime translations, the result is neither symmetric nor gauge invariant, but this is also true for Maxwell theory. An idea could be to add a term: $t_{\mu\nu} \mapsto t_{\mu\nu} + \partial^\rho \Theta_{\rho\mu\nu}$, with $\Theta_{\rho\mu\nu} = -\Theta_{\mu\rho\nu}$ as to not violate current conservation; such a term would make $t_{\mu\nu}$ symmetric, but still not gauge invariant.

A different approach is to interpret the lack of energy conservation for matter fields as energy transferred to the gravitational field. Although covariant conservation is not actual conservation for $T_{\mu\nu}$, it can be rewritten as:

$$0 = \nabla_\mu T^\mu_\nu = \frac{1}{\sqrt{-g}} \partial_\mu (\sqrt{-g} T^\mu_\nu) - \Gamma^\rho_{\mu\nu} T^\mu_\rho = \frac{1}{\sqrt{-g}} \partial_\mu (\sqrt{-g} T^\mu_\nu) - \frac{1}{2} \partial_\nu g_{\mu\rho} T^{\mu\rho}$$

where the symmetry of $T_{\mu\nu}$ was used. Note that $\Gamma^\rho_{\mu\nu}$ reduces to $g_{\mu\rho,\nu}$ only when ν is down: this reflects that this equation is non-covariant. Using the field equations:

$$\partial_\mu (\sqrt{-g} T^\mu_\nu) = \frac{1}{16\pi G} \sqrt{-g} \partial_\nu g_{\mu\rho} \left(R^{\mu\rho} - \frac{1}{2} R g^{\mu\rho} \right) = \frac{1}{16\pi G} \sqrt{-g} \partial_\nu g_{\mu\rho} R^{\mu\rho}$$

The idea is to express this equation as $\partial_\mu (\sqrt{-g} T^\mu_\nu) = -\partial_\mu (\sqrt{-g} t^\mu_\nu)$, for some t^μ_ν referred to as the **Landau–Lifshitz pseudotensor**: this suggests that the sum of matter energy T^μ_ν and gravitational energy t^μ_ν is conserved, however it would be coordinate-dependent as t^μ_ν is not a real tensor.

The final approach assumes that $t_{\mu\nu}$ is quadratic in $h_{\mu\nu}$, so, keeping $g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$, expand the Einstein field equations to second order:

$$\left[R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \right]^{(1)} + \left[R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \right]^{(2)} = 8\pi G T_{\mu\nu} \iff \left[R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \right]^{(1)} = 8\pi G (T_{\mu\nu} + t_{\mu\nu})$$

where superscripts (n) denote a restriction to terms of order h^n . The gravitational energy-momentum non-tensor thus is:

$$t_{\mu\nu} = -\frac{1}{8\pi G} \left[R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \right]^{(2)} = -\frac{1}{8\pi G} \left[R_{\mu\nu}^{(2)} - \frac{1}{2} R^{(2)} \eta_{\mu\nu} - \frac{1}{2} R^{(1)} h_{\mu\nu} \right]$$

Far away from the source $R^{(1)}$ can be neglected, since it vanishes by the equations of motion (at linear order), so:

$$t_{\mu\nu} = -\frac{1}{8\pi G} \left[R_{\mu\nu}^{(2)} - \frac{1}{2} R^{(2)} \eta_{\mu\nu} \right] \quad (5.3.6)$$

The linearized Bianchi identity is $\partial^\mu \left[R_{\mu\nu} - \frac{1}{2} R g_{\mu\nu} \right]^{(1)} = 0$, hence, far away from sources, i.e. $T_{\mu\nu} = 0$, necessarily $\partial^\mu t_{\mu\nu} = 0$ as befits a conserved current. However, $t_{\mu\nu}$ is still not gauge invariant.

§5.3.2.3 Gauge invariance

Although none of the defined $t_{\mu\nu}$ are gauge invariant, it is still possible to extract a gauge-invariant quantity from it, which has physical meaning.

First, if the spacetime is asymptotically Minkowski, it could be possible to integrate t^{00} on an infinite spatial hypersurface, obtaining the so-called **ADM energy**, which is shown to be constant in time and gauge invariant. Alternatively, one could integrate t^{0i} on a sphere at \mathcal{I}^+ , yielding the so-called **Bondi energy**, which is gauge-invariant and time-dependent, and it can be defined in the full non-linear theory too.

Moreover, consider that, like any other waves, gravitational waves vary over some typical length λ . Thus, averaging over these oscillations is possible by:

$$\langle t_{\mu\nu} \rangle := \int_V d^4y W(x-y) t_{\mu\nu}(y)$$

where V is a spacetime region of typical size a and $W \in \mathcal{C}^\infty(V) : \int_V d^4x W(x) = 1 \wedge W|_{\partial V} = 0$. This integral averages total derivatives as $\langle \partial X \rangle \sim 1/a$, so $\langle X \partial Y \rangle = -\langle Y \partial X \rangle + o(1/a)$. Then:

$$\langle t_{\mu\nu} \rangle = \frac{1}{32\pi G} \langle \partial_\mu h_{\rho\sigma} \partial_\nu h^{\rho\sigma} \rangle \quad (5.3.7)$$

This is indeed a conserved quantity:

$$\partial^\mu \langle t_{\mu\nu} \rangle = \frac{1}{32\pi G} \langle (\square h_{\rho\sigma}) \partial_\nu h^{\rho\sigma} + \frac{1}{2} \partial_\nu (\partial_\mu h_{\rho\sigma} \partial^\mu h^{\gamma\sigma}) \rangle = 0$$

as the first term vanishes by equations of motion and the second yields a negligible total derivative. Under a gauge transformation like Eq. 5.1.8:

$$\delta \langle t_{\mu\nu} \rangle = \frac{1}{16\pi G} \langle \partial_\mu h_{\rho\sigma} \partial_\nu (\partial^\rho \xi^\sigma + \partial^\sigma \xi^\rho) \rangle = \frac{1}{16\pi G} \langle \partial_\mu \partial^\rho h_{\rho\sigma} \partial_\nu \xi^\sigma + \partial_\mu \partial^\sigma h_{\rho\sigma} \partial_\nu \xi^\rho \rangle + o(a^{-1}) = o(a^{-1})$$

where, after integration by parts, de Donder gauge condition $\partial^\rho h_{\rho\sigma} = 0$ was invoked. Hence, $t_{\mu\nu}$ is almost gauge-invariant, and it is properly gauge-invariant if $a \rightarrow \infty$, i.e. if the averaging is over all of spacetime. The power emitted by a gravitational wave at infinity is:

$$\mathcal{P} = \int_{\mathbb{S}_\infty^2} d^2x \hat{n}_i \langle t^{0i} \rangle$$

with \hat{n}_i a normal versor to \mathbb{S}_∞^2 . This indeed gives Eq. 5.3.5.

§5.3.2.4 Orders of magnitude

Black holes Consider two masses M , separated by a distance R and orbiting each other with frequency ω . Newtonian gravity approximation gives:

$$\omega^2 R \sim \frac{GM}{R^2}$$

The quadrupole is $\mathcal{Q} \sim MR^2$, hence $\ddot{\mathcal{Q}} \sim \omega^3 MR^2$, and the emitted power scales as:

$$\mathcal{P} \sim G \ddot{\mathcal{Q}}^2 \sim \frac{G^4 M^5}{R^5}$$

Returning to SI units, $[G] = M^{-1}L^3T^{-2}$ and, for a black hole, $R_s = 2GM/c^2$, thus:

$$\mathcal{P} = \left(\frac{R_s}{R}\right)^5 L_p$$

where the **Plank luminosity** is defined as:

$$L_p \equiv \frac{c^5}{G} \approx 3.6 \cdot 10^{52} \text{ Js}^{-1}$$

To get a sense of scale, the Sun emits $L_\odot \approx 10^{-26} L_p$ and the Milky Way, with 10^{11} stars, emits $L \approx 10^{-15} L_p$. Moreover, in the visible Universe there are roughly 10^{10} galaxies, so all the stars of the visible Universe shine with $L \approx 10^{-5} L_p$. Yet, when two black holes spiral towards each other and their separation becomes comparable to their Schwarzschild radius, they emit in gravitational waves 10^5 times the energy emitted by all the stars of the visible Universe.

Solar System If two objects with masses $M_1 \gg M_2$ orbit each other, then the gravitationally-radiated power is:

$$\mathcal{P} \sim \frac{G^4 M_1^3 M_2^2}{R^5}$$

Considering Jupiter, which has $M \sim 10^{-3} M_\odot$ and orbits at $R \approx 10^9$ km from the Sun, which has $R_s \approx 3$ km:

$$\mathcal{P} \approx 10^{-50} L_p \approx 10^{-24} L_\odot$$

This is completely negligible.

Human body Consider a human being shaking its arms around. The mass of a human arm is a few kg and moves a distance of around 1 m with a frequency $\omega \sim 1$ Hz, so $\dot{Q} \approx 1 \text{ kg m}^2$ and $\ddot{Q} \approx 1 \text{ kg m}^2 \text{ s}^{-3}$, hence the gravitationally-radiated power is:

$$\mathcal{P} \sim \frac{G \ddot{Q}^2}{c^5} \approx 10^{-52} \text{ Js}^{-1}$$

Suppose the existence of gravitons with $E = \hbar\omega$: to produce a single graviton with $\omega = 1$ Hz, i.e. $E \approx 10^{-34} \text{ J}$, a human needs to shake its arms around for $t \approx 10^{18} \text{ s}$: this is approximately the age of the Universe.

Chapter 6

Black Holes

§6.1 Schwarzschild black holes

The simplest black hole solution is the Schwarzschild metric:

$$ds^2 = - \left(1 - \frac{2GM}{r}\right) dt^2 + \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2 (d\vartheta^2 + \sin^2 \vartheta d\varphi^2) \quad (6.1.1)$$

This is a special case of the general metric in §4.2.2 with $f(r)^2 = 1 - 2GM/r$: it solves the vacuum Einstein equations $R_{\mu\nu} = 0$.

This solution depends on a single parameter M , which is identified as the mass of the black hole, as highlighted in §4.6. The same can be shown via Komar integrals: note that the Schwarzschild metric admits a timelike Killing vector field $K = \partial_t$, with dual 1-form $K^\flat = g_{00}dt$ and Komar charge:

$$F = dK^\flat = -\frac{2GM}{r^2} dr \wedge dt \implies M_{\text{Komar}} = -\frac{1}{8\pi G} \int_{\mathbb{S}_R^2} \star F = M$$

for all $R > 2GM$ (radius of the horizon). Although the 2-form F obeys the vacuum Maxwell equations $d \star F = 0$, thus one would expect the associated Komar charge to vanish as there is no current, $M_{\text{Komar}} = M \neq 0$: this is possible because the mass of the black hole is localized at the origin, where F diverges. Moreover, the Schwarzschild solution is mathematically valid for $M \in \mathbb{R}$: the solution $M = 0$ is just Minkowski spacetime, while $M < 0$ will be shown to be unphysical.

§6.1.1 Horizon

The Schwarzschild metric diverges at $r = 0$ and $r = 2GM \equiv R_s$. To distinguish between true singularities and coordinate singularities, the usual way is defining a coordinate-independent scalar quantity and studying it at divergence points.

Due to the vacuum field equations, the simplest scalars R and $R_{\mu\nu}R^{\mu\nu}$ both vanish: the next simplest curvature scalar is the **Krentschmann scalar** $R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma}$, which for the Schwarzschild metric is:

$$R^{\mu\nu\rho\sigma}R_{\mu\nu\rho\sigma} = \frac{48G^2M^2}{r^6} \quad (6.1.2)$$

At $r = 2GM$, the Krentschmann scalar is $\sim 1/(GM)^4$, suggesting that this is only a coordinate singularity: it turns out that is the case, and indeed the surface $r = R_s$ is known as the **event horizon** of the black hole. Note that, interestingly, heavier black holes have smaller curvature at the horizon.

On the contrary, $r = 0$ is a true singularity, simply known as the **singularity** of the (classical) black hole.

§6.1.1.1 Near horizon limit

To study the spacetime in the vicinity of the horizon, set $r = 2GM + \eta$, with $|\eta| \ll 2GM$. Moreover, consider $\eta \in \mathbb{R}^+$, thus restricting to the spacetime just outside the horizon. The metric then becomes, at first order in η :

$$ds^2 = -\frac{\eta}{2GM}dt^2 + \frac{2GM}{\eta}d\eta^2 + (2GM)^2d\Omega_2^2$$

Spacetime is thus decomposed in a direct product of \mathbb{S}_{2GM}^2 and a 2-dimensional Lorentzian manifold. Focusing on the latter, set $\rho^2 = 8GM\eta$, so that:

$$ds^2 = -\left(\frac{\rho}{4GM}\right)^2 dt^2 + d\rho^2 \quad (6.1.3)$$

This is called **Rindler metric** and it is, in fact, just Minkowski spacetime:

$$T = \rho \sinh\left(\frac{t}{4GM}\right) \quad X = \rho \cosh\left(\frac{t}{4GM}\right) \implies ds^2 = -dT^2 + dX^2$$

These are precisely the coordinates Eq. 3.2.5 experienced by an observer undergoing constant acceleration $a = 1/(4GM)$: this makes sense, as an observer sitting at constant ρ , i.e. constant r , must accelerate in order not to fall into the black hole.

The outside of the black hole is $(\rho, t) \in \mathbb{R}^+ \times \mathbb{R}$, corresponding to $X > |T|$. The horizon $\rho = 0$, instead, is mapped to the origin $T = X = 0$ of Minkowski space. However, note that the time coordinate is undefined at the origin since $g_{00} = 0$: scaling $t \rightarrow \infty$ and $\rho \rightarrow 0$ keeping $\rho e^{\pm t/4GM}$ fixed makes it clear that the horizon actually corresponds to the lines:

$$r = 2GM \implies X = \pm T$$

Therefore, the event horizon is not a timelike surface (like the surface of a star), but a null surface.

Although the outside of the event horizon is only described by $X > |T|$, it is nonetheless possible to extend the metric to the whole Minkowski space $X, T \in \mathbb{R}$: this makes it clear that the event horizon is not a real singularity, as nothing particular happens at $X = \pm T$ (see Fig. 6.1). Zooming out, however, the peculiar properties of the horizon start to emerge from a global perspective.

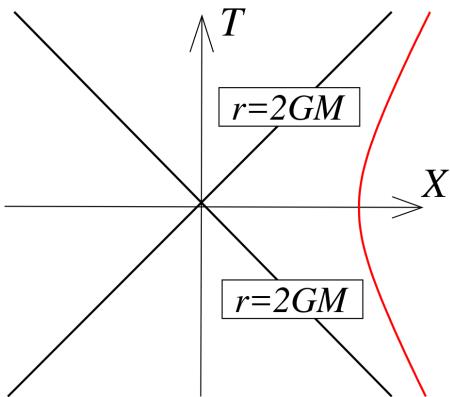


Figure 6.1: Rindler spacetime, with null lines in black and a line at constant $r > 2GM$ in red.

§6.1.2 Eddington–Finkelstein coordinates

Both the Schwarzschild coordinates and the Rindler coordinates do not cover the whole manifold. To find a set of coordinates that does, first introduce a new radial coordinate:

$$dr_*^2 = \left(1 - \frac{2GM}{r}\right)^{-2} dr^2 \implies r_* = r + 2GM \log\left(\frac{r - 2GM}{2GM}\right) \quad (6.1.4)$$

This is the **Regge–Wheeler radial coordinate** and it maps the region outside the horizon $r \in (2GM, \infty)$ to $r_* \in \mathbb{R}$: as an observer approaches the horizon, r changes increasingly slow varying r_* , since $\frac{dr}{dr_*} \rightarrow 0$ as $r \rightarrow 2GM$. This coordinate is suited to describe the path of lightrays travelling in the radial direction:

$$ds^2 = 0 \implies \frac{dr}{dt} = \pm \left(1 - \frac{2GM}{r}\right) \implies \frac{dr_*}{dt} = \pm 1 \implies t \pm r_* = \text{const.}$$

These null radial geodesics correspond to an ingoing lightray for the positive sign and an outgoing lightray for the negative sign (r_* must decrease/increase as t increases). Next, introduce a pair of null coordinates:

$$v = t + r_* \quad u = t - r_* \quad (6.1.5)$$

These allow to extend the Schwarzschild solution beyond the horizon.

§6.1.2.1 Ingoing coordinates

Replace t with $t = v - r_*(r)$:

$$dt = dv - \left(1 - \frac{2GM}{r}\right)^{-1} dr$$

The Schwarzschild metric then becomes, in coordinates (v, r) :

$$ds^2 = -\left(1 - \frac{2GM}{r}\right) dv^2 + 2dv dr + r^2 d\Omega_2^2 \quad (6.1.6)$$

This is the Schwarzschild black hole in **ingoing Eddington–Finkelstein coordinates**. Note that there is no more singularity at $r = 2GM$; however, the dv^2 term vanishes at the horizon and becomes negative for $r < 2GM$. Nonetheless, the metric is still non-degenerate thanks to the cross-term:

$$\det g = \det \begin{bmatrix} -(1 - \frac{2GM}{r}) & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2 \sin^2 \vartheta \end{bmatrix} = -r^4 \sin^2 \vartheta$$

Thus, the metric is non-degenerate and with Lorentzian signature for all values of $r \in \mathbb{R}_+$ (except for the known poles of \mathbb{S}^2): this means that the radial coordinate can be extended past the horizon. The globally-timelike Killing vector field $K = \partial_t$ of the Schwarzschild metric is retained in the ingoing Eddington–Finkelstein coordinates as $K = \partial_v$, however now it is no longer globally-timelike: it remains so outside the horizon, where $g_{vv} < 0$, but become spacelike inside it, where $g_{vv} > 0$. This means that the full black hole geometry is not time-independent. Recall that ingoing lightrays follow geodesics given by $v = \text{const.}$, while outgoing lightrays follow $u = t - r_* = \text{const.}$. As a function of (v, r_*) , the latter reads $v = 2r_* + \text{const.}$, so in (v, r) coordinates:

$$v = 2r + 4GM \log\left(\frac{r - 2GM}{2GM}\right) + \text{const.}$$

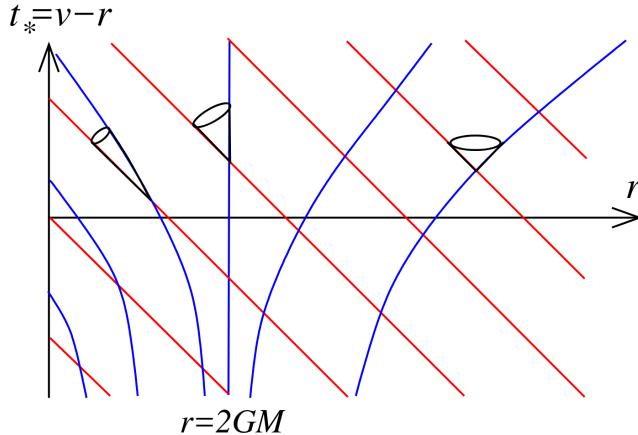


Figure 6.2: Finkelstein diagram in ingoing coordinates: ingoing null geodesics are shown in red, outgoing ones in blue.

This clearly has domain $r \in (2GM, \infty)$, therefore the Regge–Wheeler radial coordinate needs a redefinition:

$$r_* = r + 2GM \log \left| \frac{r - 2GM}{2GM} \right|$$

This means that now r_* is a multi-valued function: $r \in (2GM, \infty)$ is mapped to $r_* \in \mathbb{R}$, while $r \in (0, 2GM)$ is mapped to $r_* \in \mathbb{R}_-$, with the singularity at $r_* = 0$. Outgoing lightrays inside the horizon will then follow geodesics given by:

$$v = 2r + 4GM \log \left(\frac{2GM - r}{2GM} \right) + \text{const.}$$

Outgoing null geodesics at the horizon are easily found: the dv^2 term in the metric Eq. 6.1.6 vanishes and the surface $r = 2GM$ is itself a null geodesic (as event horizons are always null surfaces).

All of this information about lightrays can be pictured in a **Finkelstein diagram**, which shows null geodesics in a radial-temporal diagram. Since r_* is multi-valued, r is chosen as radial coordinate: as a consequence, one needs to define a new temporal coordinate $t_* : v = t + r_* = t_* + r$, i.e. $t_* = v - r$. The Finkelstein diagram in ingoing coordinates is shown in Fig. 6.2: ingoing null geodesics are $v = \text{const.}$, i.e. $t_* = \text{const.} - r$, hence travelling at 45° , while outgoing ones are given by the previously found expressions. Note that only outgoing null geodesics outside of the horizon effectively “go out”, since $r \rightarrow \infty$ as $t \rightarrow \infty$ (so $t_* \rightarrow \infty$): those inside the horizon move inexorably towards the curvature singularity at $r = 0$ as time passes, and each one of them reaches it in a finite time t_* . The boundary between these two regions are the null geodesics which run along the horizon at $r = 2GM$.

This analysis can be extended to massive particles: they must move along timelike geodesics inside their future lightcone, which is determined by a pair of ingoing and outgoing future-pointing null geodesics. These future lightcones are also shown in Fig. 6.2: the name “black hole” captures the causal structure of the spacetime, as neither massive particles nor lightrays can escape outside of the event horizon, thus an outside observer can know nothing of the inside of the black hole.

Consider two observers, one at constant $r > 2GM$ and the other moving towards the singularity: as the second approaches the event horizon, light signals emitted by it will take longer and longer to reach the first, and they will appear progressively more redshifted due to the increasingly

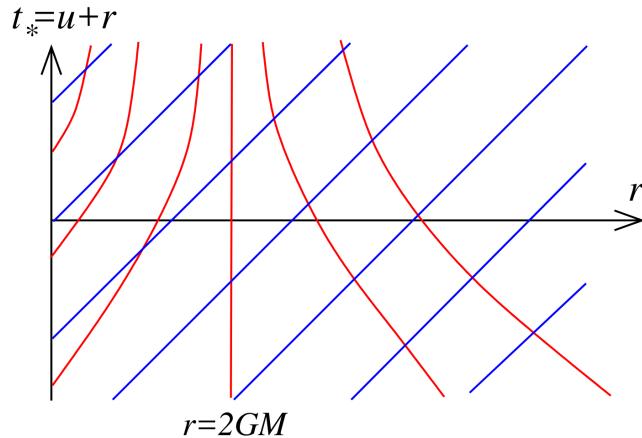


Figure 6.3: Finkelstein diagram in outgoing coordinates: ingoing null geodesics are shown in red, outgoing ones in blue.

strong gravitational well. Thus, the outside observer will see the ingoing observer forever, but progressively more slowed down and redshifted, and will never see it passing the event horizon.

§6.1.2.2 Outgoing coordinates

A different extension of the exterior of the Schwarzschild black hole is obtained by replacing t with $u = t - r_*$, i.e. $t = u + r_*(r)$:

$$ds^2 = - \left(1 - \frac{2GM}{r}\right) du^2 - 2du dr + r^2 d\Omega_2^2 \quad (6.1.7)$$

This is the Schwarzschild black hole in **outgoing Eddington–Finkelstein coordinates**. Once again, the metric is smooth and non-degenerate at the horizon, thus it is possible to extend $r \in \mathbb{R}_+$. However, now $r \in (0, 2GM)$ describe a different part of spacetime from the analogous region in ingoing coordinates. To see this, set coordinates $(r, t_*) : t_* = u + r$ and draw the Finkelstein diagram (Fig. 6.3): outgoing null geodesics hence travel at 45° , effectively “going out” regardless of their position with respect to the horizon. On the contrary, ingoing null geodesics now do not “go in”: those starting outside of the horizon pile up at the horizon itself, unable to cross it, while those starting inside the horizon move outwards and pile up at it, still unable to cross it.

Studying lightcones, it’s easy to see that massive particles inside the horizon are inexorably expelled outside of it in a finite amount of time t_* . This is clearly a very different physics from a black hole: the metric Eq. 6.1.7 is that of a **white hole**, an object which expels any matter inside. A white hole can be viewed as the time reversal of a black hole: this is to be traced to the negative sign in the cross-term and can be seen by flipping Fig. 6.2 upside-down, getting Fig. 6.2.

White holes are perfectly acceptable solutions to the Einstein field equations, and their existence is to be expected from the time-reversal invariance of the field equations. However, they’re not physically relevant, as they cannot be formed from collapsing matter.

§6.1.3 Kruskal spacetime

To better understand how the region parametrized by $r \in (0, 2GM]$ corresponds to two different parts of spacetime, one needs to introduce coordinates which cover the entire spacetime, including both black and white holes.

First, rewrite the Schwarzschild metric in (u, v) coordinates:

$$ds^2 = - \left(1 - \frac{2GM}{r}\right) du dv + r^2 d\Omega_2^2$$

where $r = r(u - v)$. There's a degeneracy at $r = 2GM$ which can be solved introducing the **Kruskal–Szekeres coordinates**:

$$U = -\exp\left(-\frac{u}{4GM}\right) \quad V = \exp\left(\frac{v}{4GM}\right) \quad (6.1.8)$$

Both U and V are null coordinates, and the exterior of the black hole is parametrized by $U < 0, V > 0$. The metric becomes:

$$ds^2 = -\frac{32(GM)^3}{r} e^{-\frac{r}{2GM}} dU dV + r^2 d\Omega_2^2 \quad (6.1.9)$$

This is known as the metric of *Kruskal spacetime*. The function $r = r(U, V)$ can be found inverting:

$$UV = -\exp\left(\frac{r_*}{2GM}\right) = \frac{2GM - r}{2GM} \exp\left(\frac{r}{2GM}\right) \quad (6.1.10)$$

On the other hand, the function $t = t(U, V)$ can be found inverting:

$$\frac{U}{V} = -\exp\left(-\frac{t}{2GM}\right) \quad (6.1.11)$$

The original Schwarzschild metric, which covers $U < 0, V > 0$, can now be extended to $(U, V) \in \mathbb{R}^2$, without any divergence at $r = 2GM$.

Proposition 6.1.1

Kruskal spacetime is the **maximal extension** of the Schwarzschild metric.

Proof. To check whether a given spacetime can be further extended, one needs to look at all possible geodesics: if they are defined for infinite affine parameter, then they can escape to infinity; however, if they alt at some finite affine parameter, that is either a coordinate singularity, hinting that the spacetime can in fact be extended, or a true singularity. A spacetime is maximally extended if any geodesics that alts at a finite affine parameter does so at a true singularity, and that is the case for Kruskal spacetime. \square

The extension process allows to solve the field equations in some particular region (open subset) of spacetime: being the metric components real analytic functions, this is sufficient to determine the metric in all of spacetime.

§6.1.3.1 Kruskal diagram

To see where the event horizon is mapped in Kruskal spacetime:

$$r = 2GM \quad \Rightarrow \quad U = 0 \quad \vee \quad V = 0$$

This means that the event horizon no longer is a null surface, but two null surfaces intersecting at $U = V = 0$, in agreement with the near horizon limit (Rindler space): the null surface $U = 0$ is the horizon of the black hole, known as the *future horizon*; the null surface $V = 0$ is

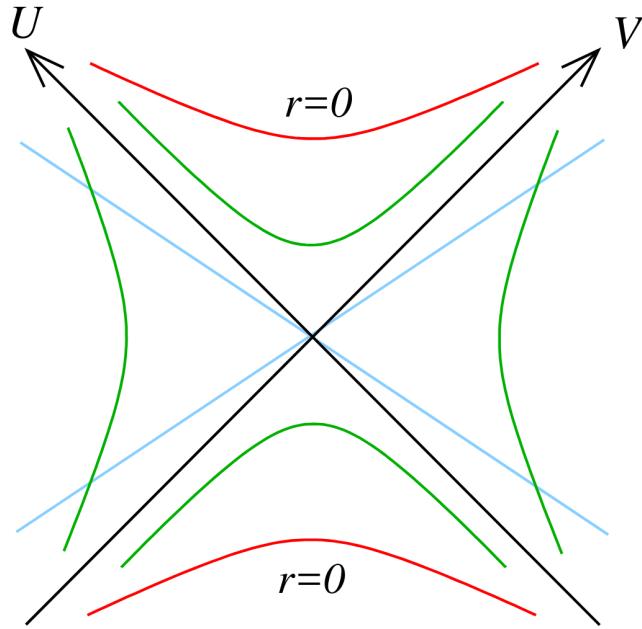


Figure 6.4: Kruskal diagram for the Schwarzschild black hole: green lines are $r = \text{const.}$ and blue lines $t = \text{const.}$, while singularities are in red and horizons in black.

the horizon of the white hole, known as the *past horizon*. On the other hand, the singularity becomes:

$$r = 0 \quad \Rightarrow \quad UV = 1$$

This hyperbola has two disconnected components: one ($U, V > 0$) corresponding to the singularity of the black hole and one ($U, V < 0$) to the singularity of the white hole.

All of this information can be represented in a **Kruskal diagram**, as shown in Fig. 6.4: the U and V axes are at 45° , reflecting that they are null lines, and these are the two horizons. In this diagram, the vertical direction can be viewed as a time coordinate $T = \frac{1}{2}(V + U)$ and the horizontal one as a spatial coordinate $X = \frac{1}{2}(V - U)$. This diagram makes it clear that the black hole and the white hole cohabit the same spacetime. Lines of constant r and t correspond respectively to lines of constant UV and U/V , as from Eq. 6.1.10-6.1.11. The Kruskal diagram makes it clear that, once an observer crosses the event horizon, the singularity is unavoidable: the $r = 0$ are not timelike worldlines, but *the singularity is spacelike*. Therefore, for the infalling observer, after crossing the horizon the singularity of the black hole does not lie in a spatial direction, but it lies in the future. Similarly, the singularity of the white hole lies in the past. This can be formally seen using the Killing vector field $K = \partial_t$ of the Schwarzschild solution. Outside of the horizon K is timelike, therefore it defines the conserved energy of geodesics; in Kruskal-Szekeres coordinates:

$$K = \partial_t = \frac{\partial V}{\partial t} \partial_V + \frac{\partial U}{\partial t} \partial_U = \frac{1}{4GM} \left(V \frac{\partial}{\partial V} - U \frac{\partial}{\partial U} \right) \quad \Rightarrow \quad g_{\mu\nu} K^\mu K^\nu = - \left(1 - \frac{2GM}{r} \right)$$

Outside of the horizon $K_\mu K^\mu < 0$ as expected. But inside, with $r < 2GM$, the Killing vector field is spacelike: therefore, by Def. 4.6.1 and Prop. 4.6.1, the full Schwarzschild spacetime is not time-independent, but it becomes clear only after crossing the event horizon.

Einstein–Rosen bridge With reference to the Kruskal diagram in Fig. 6.4, the right-hand quadrant is the exterior of the black hole, covered by the Schwarzschild coordinates, while

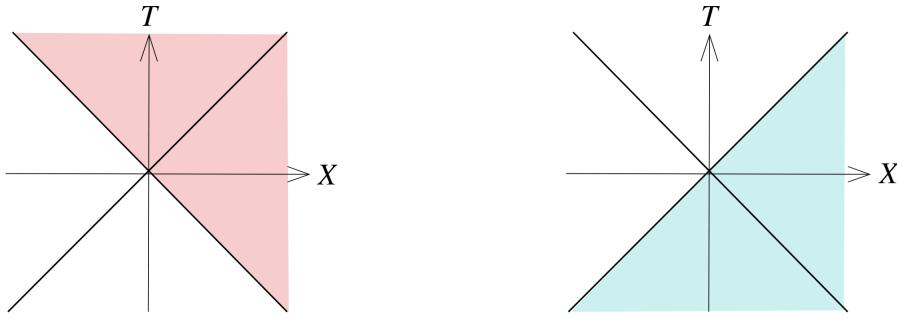


Figure 6.5: Regions of Kruskal spacetime covered by ingoing and outgoing Eddington–Finkelstein coordinates respectively.

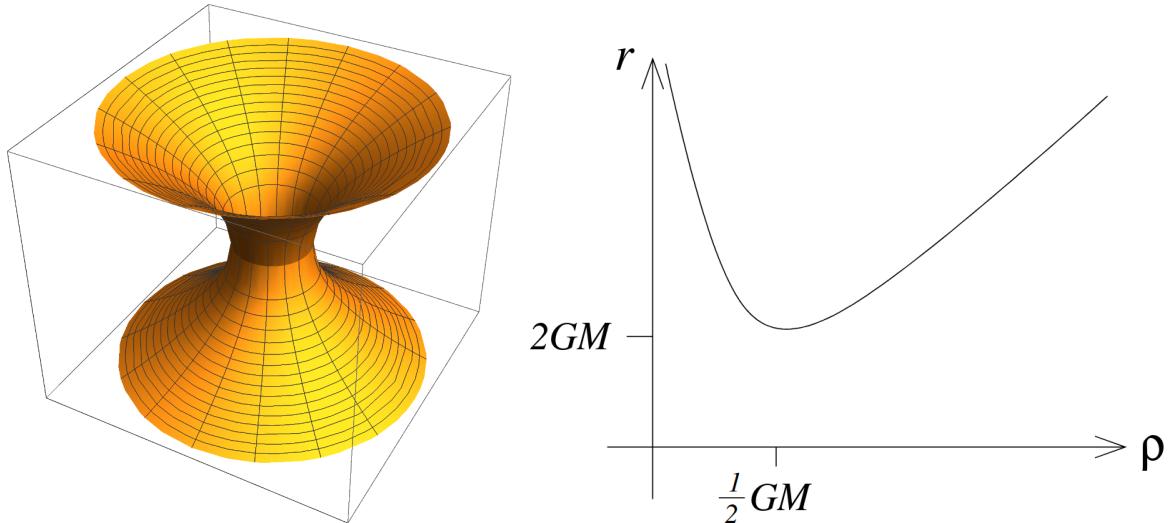


Figure 6.6: The Einstein–Rosen bridge and the ρ coordinate which parametrizes it.

the upper and lower quadrants are respectively the interior of the black hole and of the white hole, covered by ingoing and outgoing Eddington-Finkelstein coordinates (see Fig. 6.5). There remains the left-hand quadrant: this turns out to be just another copy of the exterior of the black hole, now covered by $U > 0, V < 0$.

The full spacetime hence contains two asymptotically flat regions, joined together by a black hole: note that it is not possible for an observer in one region to send a signal in the other, because the causal structure of the spacetime does not allow this.

To elucidate the spatial geometry connecting these two regions, one need to study the $t = 0$ hypersurface of Kruskal spacetime: this is a straight horizontal line passing through $U = V = 0$ in Fig. 6.4. In Schwarzschild coordinates, the geometry of the $t = 0$ hypersurface is given by:

$$ds^2 = \left(1 - \frac{2GM}{r}\right)^{-1} dr^2 + r^2(d\vartheta^2 + \sin^2 \vartheta d\varphi^2) \quad (6.1.12)$$

which is valid for $r > 2GM$. Kruskal spacetime presents two copies of this geometry, one in the right-hand quadrant and one in the left-hand one, which are glued together at $r = 2GM$ to give a wormhole-like geometry as in Fig. 6.6: this is called the **Einstein–Rosen bridge**. Note that it is not possible to travel through it, as the paths are spacelike, not timelike.

It is possible to write a metric which includes both sides. Define the radial coordinate ρ such

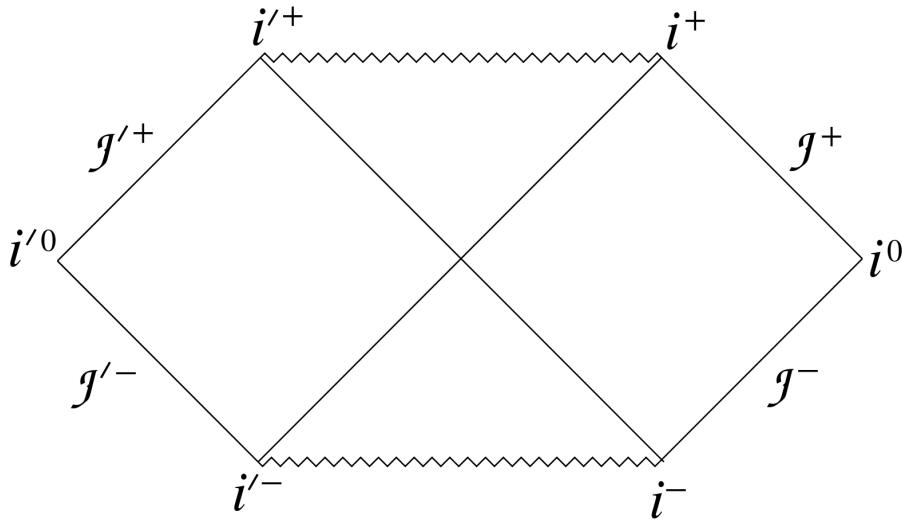


Figure 6.7: The Penrose diagram for Kruskal spacetime.

that:

$$r = \rho \left(1 + \frac{GM}{2\rho}\right)^2 = \rho + GM + \frac{G^2 M^2}{4\rho}$$

Plotting $r = r(\rho)$, as in Fig. 6.6, it's clear that there are two values of ρ for each value of $r > 2GM$, while $r = 2GM$ only corresponds to $\rho = GM/2$: thus, is possible to parametrize one side of the wormhole with $\rho \in (0, GM/2)$ and the other with $\rho \in (GM/2, \infty)$. The metric in Eq. 6.1.12 becomes:

$$ds^2 = \left(1 + \frac{GM}{2\rho}\right)^4 [d\rho^2 + \rho^2 (d\vartheta^2 + \sin^2 \vartheta d\varphi^2)] \quad (6.1.13)$$

It is clear that, as $\rho \rightarrow \infty$, the (spatial) metric is that of \mathbb{R}^3 . But the same is true for $\rho \rightarrow 0$: note that there's a symmetry $r \mapsto r$ for $\rho \mapsto G^2 M^2 / (4\rho)$, which swaps the two asymptotic spacetimes while leaving the midpoint at $\rho = GM/2$ invariant, i.e. it swaps the two sides of the wormhole.

Furthermore, the radius of the spatial \mathbb{S}^2 is minimum at $2GM$ in the midpoint $\rho = GM/2$ of the wormhole and increases in either direction: this middle point corresponds to the common point $U = V = 0$ of the two horizons and is called the **bifurcation sphere**.

Although there is no way that an observer in one quadrant can signal an observer in the other quadrant, there is one way for them to communicate: jump into the black hole and meet behind the respective horizons. However, as the white hole is thought to have no physical manifestation, the other universe in the left-hand quadrant of Kruskal space is thought to be a mathematical artifact too.

§6.1.3.2 Penrose diagram

From the Kruskal diagram in Fig. 6.4, it's easy to draw the Penrose diagram for Kruskal spacetime. First, introduce new coordinates which cover the entire spacetime in a finite range:

$$U = \tan \tilde{U} \quad V = \tan \tilde{V}$$

The new coordinates have range $(\tilde{U}, \tilde{V}) \in (-\frac{\pi}{2}, +\frac{\pi}{2})$ and the Kruskal metric becomes:

$$ds^2 = \frac{1}{\cos^2 \tilde{U} \cos^2 \tilde{V}} \left[-\frac{32(GM)^3}{r} e^{-\frac{r}{2GM}} d\tilde{U} d\tilde{V} + r^2 \cos^2 \tilde{U} \cos^2 \tilde{V} d\Omega_2^2 \right]$$

This metric is conformally equivalent to:

$$ds^2 = -\frac{32(GM)^3}{r} e^{-\frac{r}{2GM}} d\tilde{U} d\tilde{V} + r^2 \cos^2 \tilde{U} \cos^2 \tilde{V} d\Omega_2^2$$

However, the singularity at $r = 0$ must be retained: this is mapped to $UV = 1$, thus in the finite-range coordinates:

$$\begin{aligned} \tan \tilde{U} \tan \tilde{V} = 1 &\iff \sin \tilde{U} \sin \tilde{V} - \cos \tilde{U} \cos \tilde{V} = 0 \\ &\iff \cos(\tilde{U} + \tilde{V}) = 0 \iff \tilde{U} + \tilde{V} = \pm \frac{\pi}{2} \end{aligned}$$

These are straight horizontal lines in the Penrose diagram which chop the diamond-shaped diagram (as that in Fig. 4.5): the resulting diagram is shown in Fig. 6.7, with singularities represented by jagged lines.

The Penrose diagram contains the same information as the Kruskal diagram in Fig. 6.4, but it allows to state a more rigorous definition of black hole. Restricting to asymptotically flat spacetimes, the two asymptotic regions contain both null infinities \mathcal{I}^+ and \mathcal{I}^- : the black hole region is then defined as the set of points which cannot send signals to \mathcal{I}^+ . The boundary of the black hole region is the *future event horizon* \mathcal{H}^+ : equivalently, \mathcal{H}^+ is the boundary of the causal past of \mathcal{I}^+ . With reference to Fig. 6.7, the black hole associated to \mathcal{I}^+ is the upper quadrant and the left quadrant, while the black hole associated to \mathcal{I}'^+ is the upper quadrant and the right quadrant.

Importantly, note that to define a black hole one needs to know the whole spacetime, as all light rays must be run backwards from \mathcal{I}^+ to determine \mathcal{H}^+ as their boundary: there is no reference to any spacelike hypersurface at constant coordinate time, thus an observer cannot really know if it is inside a black hole unless by knowing the entire future evolution of the spacetime.

Equivalently, it is possible to define the white hole region and the *past event horizon* as the region that cannot receive signals from \mathcal{I}^- and its boundary.

§6.1.4 Weak cosmic censorship

Kruskal spacetime is unphysical in a number of ways. In fact, in reality, black holes do not emerge from white holes, but are formed by collapsing stars: the resulting Penrose diagram is rather different from that in Fig. 6.7. To understand the causal structure of such a black hole, consider the (unrealistic) situation of the spherically-symmetric collapse of a shell of matter: inside the shell, spacetime is flat, meanwhile on the outside it is described by Schwarzschild metric Eq. 6.1.1 (by Birkhoff's theorem, time dependence does not change this fact). Furthermore, assume (unrealistically) that the shell is travelling at the speed of light, so that the Penrose diagrams for Minkowski spacetime and Kruskal spacetime can be glued as in Fig. 6.8: this is the Penrose diagram of a collapsing black hole.

Despite the number of unphysical assumptions, the Penrose diagram thus derived correctly describes the causal structure of a realistic spherically-symmetric collapsing star: as in Eq. 6.1.9, its surface follows a timelike trajectory.

§6.1.4.1 Naked singularities

An important feature of the black hole is that the singularity remains shrouded behind the event horizon, so that an asymptotic observer cannot see it.

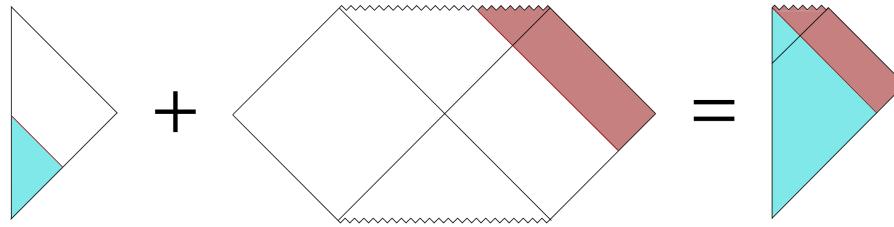


Figure 6.8: Joining two Penrose diagrams.

On the contrary, Einstein field equations also allow **naked singularities**, i.e. singularities without an horizon. For example, consider Schwarzschild metric with $M < 0$:

$$ds^2 = - \left(1 + \frac{2G|M|}{r}\right) dt^2 + \left(1 + \frac{2G|M|}{r}\right)^{-1} dr^2 + r^2 (d\vartheta^2 + \sin^2 \vartheta d\varphi^2)$$

This metric presents no coordinate singularity at $r = 2G|M|$, and so no event horizon. The Penrose diagram for such a spacetime is found analogously to that of Minkowski spacetime (see §4.4.2.1), setting null coordinates $u = t - r$ and $v = t + r$. The resulting diagram, shown in Fig. 6.10, is identical to that of Minkowski spacetime (Fig. 4.7), with the difference that now $r = 0$ presents a curvature singularity not shielded by an horizon: it is thus observed from \mathcal{I}^+ . An important conjecture in General Relativity, known as the **weak cosmic censorship conjecture**, states the following: given matter which obeys the dominant energy condition Eq. 4.5.21, generic smooth initial conditions on a spatial hypersurface for both the metric and matter fields in an asymptotically flat spacetime will not evolve to form naked singularities.

If this conjecture is true, then a dynamical evolution such as that in Fig. 6.10 must be ruled out: in this diagram, once the singularity forms, fields can no longer be evolved beyond the shown lightray; strictly speaking, this means that the dynamical evolution stops at the red line and cannot be extended further, thus abruptly ending \mathcal{I}^+ .

There is no proof of the weak cosmic censorship conjecture, but only circumstantial evidence. However, there is one naked singularity that seems to be physical: the Big Bang singularity. In fact, since it is in the far past, it does not violate cosmic censorship.

§6.1.5 de Sitter black holes

Schwarzschild metric Eq. 6.1.1 solves the Einstein field equations with $\Lambda = 0$, i.e. in asymptotically Minkowski spacetime. To generalize, consider $\Lambda \neq 0$, so that the field equations become

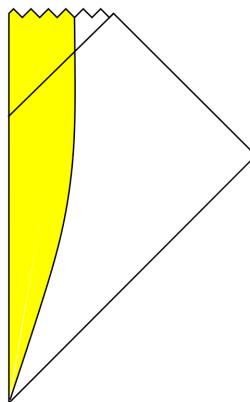


Figure 6.9: Penrose diagram for a collapsing star forming a black hole.



Figure 6.10: Penrose diagram for a negative mass black hole and unphysical collapse of a star.

$R_{\mu\nu} = \Lambda g_{\mu\nu}$, and use the following (already seen) ansatz:

$$ds^2 = -f(r)^2 dt^2 + f(r)^{-2} dr^2 + r^2 (d\vartheta^2 + \sin^2 \vartheta d\varphi^2)$$

The field equations imply that:

$$f'' + \frac{2f'}{r} + \frac{f'^2}{f} = -\frac{\Lambda}{f} \quad \wedge \quad 1 - 2ff'r - f^2 = \Lambda r^2$$

The most general solution is:

$$f(r)^2 = 1 - \frac{2GM}{r} \mp \frac{r^2}{R^2} \quad R^2 = \frac{3}{|\Lambda|} \quad (6.1.14)$$

where the negative sign is the solution with $\Lambda > 0$ and the positive sign with $\Lambda < 0$, hence describing black holes in de Sitter and anti-de Sitter spacetime respectively: in fact, if $2GM \ll R^2$, then for $r \ll R$ the metric is that of a Schwarzschild black hole in Minkowski spacetime.

Part III

Cosmology

Chapter 7

Cosmological Geometrodynamics

Of the few situations in which Einstein field equations sourced by matter Eq. 4.5.6 need to be solved directly, the one where $T_{\mu\nu}$ plays a crucial role is Cosmology, the study of the universe.

§7.1 Geometry

The key assumption of cosmology is that the universe is *spatially homogeneous and isotropic*, as inferred from empirical observations (at least on large scales): this determines a foliation of spacetime into three-dimensional spatial sections Σ_t at various times $t \in \mathbb{R}$, with each section being homogeneous and isotropic, i.e. maximally-symmetric.

Lemma 7.1.1 (Maximally-symmetric 3-spaces)

The three possible maximally-symmetric 3-spaces are:

- Euclidean space \mathbb{R}^3 , with vanishing curvature and metric:

$$d\ell^2 = dr^2 + r^2(d\vartheta^2 + \sin^2 \vartheta d\varphi^2)$$

- the 3-sphere \mathbb{S}^3 , with uniform positive curvature and metric (implicit unitary radius):

$$d\ell^2 = \frac{1}{1-r^2}dr^2 + r^2(d\vartheta^2 + \sin^2 \vartheta d\varphi^2)$$

- the 3-hyperboloid \mathbb{H}^3 , with uniform negative curvature and metric:

$$d\ell^2 = \frac{1}{1+r^2}dr^2 + r^2(d\vartheta^2 + \sin^2 \vartheta d\varphi^2)$$

These spatial metrics can be unified as:

$$d\ell^2 = \gamma_{ij}dx^i dx^j = \frac{dr^2}{1-kr^2} + r^2 d\Omega_2^2 \quad (7.1.1)$$

with $k = +1, 0, -1$ on $\mathbb{S}^3, \mathbb{R}^3, \mathbb{H}^3$ respectively.

Proof. First, show that these spaces are homogeneous and isotropic:

- Euclidean space \mathbb{R}^3 is clearly invariant under both translations and rotations;

- the 3-sphere \mathbb{S}^3 can be embedded in \mathbb{R}^4 as:

$$d\ell^2 = d\mathbf{x}^2 + du^2 \quad \mathbf{x}^2 + u^2 = a^2$$

where $a \in \mathbb{R}^+$ is the radius of the 3-sphere. Then, homogeneity and isotropy are inherited from the invariance under 4-rotations of the embedding space;

- the 3-hyperboloid \mathbb{H}^3 can be embedded in $\mathbb{R}^{1,3}$ as:

$$d\ell^2 = d\mathbf{x}^2 - du^2 \quad \mathbf{x}^2 - u^2 = -a^2$$

where $a \in \mathbb{R}^+$ is an arbitrary constant and u is a time coordinate. Then, homogeneity and isotropy are inherited from the invariance under pseudo-rotations, i.e. Lorentz transformations, of the embedding space.

In the last two cases, the coordinates can be rescaled as $\mathbf{x} \mapsto a\mathbf{x}$, $u \mapsto au$, so that the spaces can be written as:

$$d\ell^2 = a^2 [d\mathbf{x}^2 + kdu^2] \quad \mathbf{x}^2 + ku^2 = k$$

with $k = +1$ for \mathbb{S}^3 and $k = -1$ for \mathbb{H}^3 . The embedding condition gives $\mathbf{x} \cdot d\mathbf{x} = -ku du$, so that:

$$d\ell^2 = a^2 \left[d\mathbf{x}^2 + k \frac{(\mathbf{x} \cdot d\mathbf{x})^2}{1 - k\mathbf{x}^2} \right]$$

which also includes Euclidean space for $k = 0$. Switching to spherical coordinates and using $\mathbf{x} \cdot d\mathbf{x} = r dr$ the thesis is found. \square

The existence of these three maximally-symmetric 3-spaces is analogous to the existence of three maximally-symmetric spacetimes as solutions of the vacuum field equations: dS and AdS have constant spacetime curvature, supplied by the cosmological constant, while \mathbb{S}^3 and \mathbb{H}^3 have constant spatial curvature. Indeed, the metric on \mathbb{S}^3 corresponds to the spatial part of the de Sitter metric Eq. 4.2.1, while the metric on \mathbb{H}^3 corresponds to the spatial part of the anti-de Sitter metric Eq. 4.2.11.

§7.1.1 FLRW metric

Cosmology studies spacetimes in which space expands as the universe evolves, thus the metric takes the form:

$$ds^2 = dt^2 - a^2(t) \gamma_{ij} dx^i dx^j \quad (7.1.2)$$

This is the **Friedmann–Lemaître–Robertson–Walker metric** and the dimensionless factor $a(t)$ can be viewed as the size of the spatial dimensions, called the *scale factor*.

Example 7.1.1 (de Sitter metric)

de Sitter metric in global coordinates Eq. 4.2.10 is a FLRW metric with $k = +1$.

The FLRW metric has an important scale invariance:

$$a \mapsto \lambda a \quad r \mapsto \lambda^{-1}r \quad k \mapsto \lambda^2 k \quad (7.1.3)$$

This symmetry can be used to set $a_0 := a(t_0) \equiv 1$, where t_0 is the present coordinate time. Therefore, the scale factor is dimensionless, while both r and $k^{-1/2}$ have dimensions of a length.

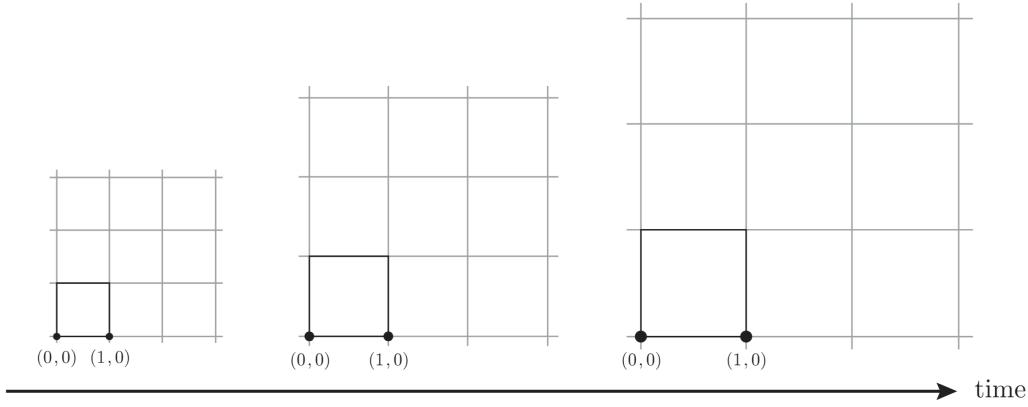


Figure 7.1: Comoving distance and physical distance between two fixed points: as the universe expands, the comoving distance remains constant, while the physical distance varies with $a(t)$.

An important clarification must be made on r : this is called a *comoving coordinate*, while the physical results depend only on the *physical coordinate* $r_{\text{phys}} := a(t)r$, as illustrated in Fig. 7.1. Then, the physical velocity of an object clearly is:

$$v_{\text{phys}} = v_{\text{pec}} + Hr_{\text{phys}} \quad (7.1.4)$$

where the peculiar velocity $v_{\text{pec}} = a(t)\dot{r}$ is the velocity measured by a comoving observer, i.e. an observer which follows the Hubble flow, and the **Hubble flow** Hr_{phys} is the contribution due to the expansion of the universe, with the *Hubble parameter* defined as:

$$H := \frac{\dot{a}}{a} \quad (7.1.5)$$

Note that the dot always represents a derivative with respect to coordinate time t .

It is convenient to express the metric in terms of redefined time η and radial coordinates, defined as:

$$d\chi \equiv \frac{dr}{\sqrt{1 - kr^2}} \quad d\eta \equiv \frac{dt}{a(t)} \quad (7.1.6)$$

where η is called **conformal time**. The new radial coordinate allows to simplify the expression for g_{rr} , while conformal time determines a factorization of the FLRW metric into a static metric and a time-dependent conformal factor:

$$ds^2 = a^2(\eta) [d\eta^2 - (d\chi^2 + S_k^2(\chi)d\Omega_2^2)] \quad (7.1.7)$$

with:

$$S_k(\chi) = \frac{1}{\sqrt{k}} \begin{cases} \sinh(\sqrt{k}\chi) & k < 0 \\ \sqrt{k}\chi & k = 0 \\ \sin(\sqrt{k}\chi) & k > 0 \end{cases} \quad (7.1.8)$$

§7.2 Kinematics

§7.2.1 Particle motion

In order to study the dynamics of particles in the FLRW metric from the geodesic equation, it is necessary to first compute the Christoffel symbols for this metric. It is straightforward to prove that the only non-vanishing Christoffel symbols are:

$$\Gamma_{ij}^0 = a\dot{a}\gamma_{ij} \quad \Gamma_{0j}^i = \frac{\dot{a}}{a}\delta_j^i \quad \Gamma_{jk}^i = \frac{1}{2}\gamma^{il}(\partial_j\gamma_{kl} + \partial_k\gamma_{jl} - \partial_l\gamma_{jk}) \quad (7.2.1)$$

Proposition 7.2.1 (Geodesics and 4-momentum)

A particle with 4-momentum p^μ moves along geodesics described by:

$$p^\alpha \partial_\alpha p^\mu = -\Gamma_{\alpha\beta}^\mu p^\alpha p^\beta \quad (7.2.2)$$

Proof. Recall Eq. 2.3.4 and express it in terms of the 4-velocity $u^\mu \equiv \frac{dx^\mu}{d\tau}$:

$$\frac{du^\mu}{d\tau} + \Gamma_{\nu\rho}^\mu u^\nu u^\rho = 0$$

But, using the chain rule:

$$\frac{du^\mu}{d\tau} = \frac{dx^\alpha}{d\tau} \frac{\partial u^\mu}{\partial x^\alpha} = u^\alpha \partial_\alpha u^\mu$$

Substituting $p^\mu = mu^\mu$ yields the thesis (which is valid in the massless case too). \square

Note that this is nothing but the condition for p^μ to be parallelly-transported along the geodesic, i.e. $p^\alpha \nabla_\alpha p^\mu = 0$. By the spatial homogeneity of the FLRW metric $\partial_i p^\mu = 0$, hence:

$$p^0 \dot{p}^\mu = -\Gamma_{\alpha\beta}^\mu p^\alpha p^\beta = - (2\Gamma_{0j}^\mu p^0 + \Gamma_{ij}^\mu p^i) p^j$$

First of all, a consequence of this equation is that massive particles which are at rest in the comoving frame, i.e. $p^j = 0$, remain at rest since $\dot{p}^\mu = 0$. Then, consider the $\mu = 0$ component of this equation and set $p^0 \equiv E$:

$$E \dot{E} = -a\dot{a}\gamma_{ij}p^i p^j = -\frac{\dot{a}}{a} \mathbf{p}^2$$

where $\mathbf{p}^2 \equiv -g_{ij}p^i p^j$ so that $p^2 = E^2 - \mathbf{p}^2 = m^2$. Thus, $E dE = \mathbf{p} \cdot d\mathbf{p} \equiv \mathbf{p} d\mathbf{p}$, with $\mathbf{p} \equiv \|\mathbf{p}\|$, and:

$$\frac{\dot{\mathbf{p}}}{\mathbf{p}} = -\frac{\dot{a}}{a} \implies \mathbf{p} \propto \frac{1}{a} \quad (7.2.3)$$

Thus, the physical 3-momentum of any particle decays with the expansion of the universe. In particular, for massless particles $E = p$ decays with the expansion, while for massive particles:

$$p = \frac{mv}{\sqrt{1 - \mathbf{v}^2}} \propto \frac{1}{a}$$

where $v^i = \dot{x}^i$ is the comoving peculiar velocity, i.e. the velocity with respect to the comoving frame. Hence, freely-falling particles converge onto the Hubble flow, as per Eq. 7.1.4.

§7.2.2 Redshift

Since the physical properties of the universe are inferred from light signals, it is crucial to quantify how the wavelength of light gets redshifted by the expansions of the universe.

From a quantum point of view, the wavelength of a light signal is linked to the photon's momentum by the de Broglie relation $\lambda = h/p$. From Eq. 7.2.3, then, $\lambda \propto a$, i.e.:

$$\lambda_0 = \frac{a_0}{a(t)} \lambda(t) \quad (7.2.4)$$

where $t < t_0$ is the time of emission of the photon. As $a(t) < a_0$, the light is redshifted.

The same result can be derived for classical EM waves. Consider a galaxy at fixed comoving distance d which, at a time η , emits a signal of short conformal duration $\Delta\eta$. The light signal arrives on Earth at $\eta_0 = \eta + d$ with the same conformal duration, but the physical time interval is different:

$$\Delta t = a(\eta)\Delta\eta \quad \Delta t_0 = a_0\Delta\eta$$

Assuming Δt to be the period of the EM wave, then $\lambda(t) = \Delta t$, hence Eq. 7.2.4 is recovered.

Definition 7.2.1 (Redshift parameter)

The **redshift parameter** of a light signal emitted at $t < t_0$ with a wavelength $\lambda(t)$ is

$$z := \frac{\lambda_0 - \lambda(t)}{\lambda(t)} \quad (7.2.5)$$

i.e. as the fractional shift in the photon's wavelength.

By Eq. 7.2.4, setting $a_0 \equiv 1$, the common definition of z is found:

$$1 + z = \frac{1}{a(t)} \quad (7.2.6)$$

For nearby sources, it is possible to expand $a(t)$ as a power series:

$$a(t) = a_0 [1 + H_0(t - t_0) + \dots]$$

where the **Hubble constant** is defined as:

$$H_0 \equiv \frac{\dot{a}(t_0)}{a(t_0)} = 100h \text{ km s}^{-1}\text{Mpc}^{-1} \quad h \approx 0.67 \pm 0.01 \quad (7.2.7)$$

The dimensionless parameter h is introduced to keep track of how the uncertainties in the measurement of H_0 propagate in other physical quantities, since H_0 normalizes everything. Inserting this expansion in the definition of z yields $z = H_0(t_0 - t) + \dots$, which, for close objects at distance $d = t_0 - t$, becomes:

$$z \simeq H_0 d \quad (7.2.8)$$

Hence, the rate of redshift of nearby objects measures the current expansion rate of the universe.

§7.2.3 Distances

For distant objects, it is important to precisely state which notion of distance is adopted.

Metric distance Rewrite the FLRW metric as:

$$ds^2 = dt^2 - a^2(t) [d\chi^2 + S_k^2(\chi)d\Omega_2^2] \quad S_k(\chi) = \begin{cases} R_0 \sinh(\chi/R_0) & k = -1 \\ \chi & k = 0 \\ R_0 \sin(\chi/R_0) & k = +1 \end{cases}$$

where the length scale R_0 has been introduced thanks to the scale invariance of the metric to compensate for the imposition of $a_0 \equiv 1$. The distance multiplying the solid angle element is the *metric distance*:

$$d_m = S_k(\chi) \quad (7.2.9)$$

In a flat universe, the metric distance is equal to the *comoving distance* χ , which can be computed as a function of redshift:

$$\chi(z) = \int_t^{t_0} \frac{dt'}{a(t')} = \int_0^z \frac{dz'}{H(z')} \quad (7.2.10)$$

which depends on the matter content of the universe through $H(z)$. It must be emphasized that neither the metric distance nor the comoving distance are physically observable.

Luminosity distance Type IA supernovae are called “standard candles”, since they are believed to be objects of known absolute luminosity L , i.e. energy emitted per second: therefore, the observed flux F , i.e. energy per second per receiving area, from a supernova explosion can be used to infer its *luminosity distance*. Considering a source at fixed comoving distance χ , in a static Euclidean space the relation between F and L would be:

$$F = \frac{L}{4\pi\chi^2}$$

However, in a generic FLRW spacetime, this relation needs to be modified for three reasons:

1. the proper area of a 3-sphere grows like $4\pi d_m^2$, not like $4\pi\chi^2$;
2. the rate of received photons is lower than the rate of emitted photons by a factor of $1/(1+z)$;
3. received photons have an energy that is lower than the energy of emitted photons by the same factor of $1/(1+z)$.

Hence, the correct relation is:

$$F = \frac{L}{4\pi d_L^2} \quad d_L = (1+z)d_m \quad (7.2.11)$$

where d_L is the *luminosity distance*.

Angular diameter distance Sometimes, it is possible to make use of “standard rulers”, i.e. objects of known physical size D (e.g. CMB fluctuations). Assuming that a standard ruler is at a fixed comoving distance χ , the *angular diameter distance* can be defined as:

$$d_A = \frac{D}{\delta\theta}$$

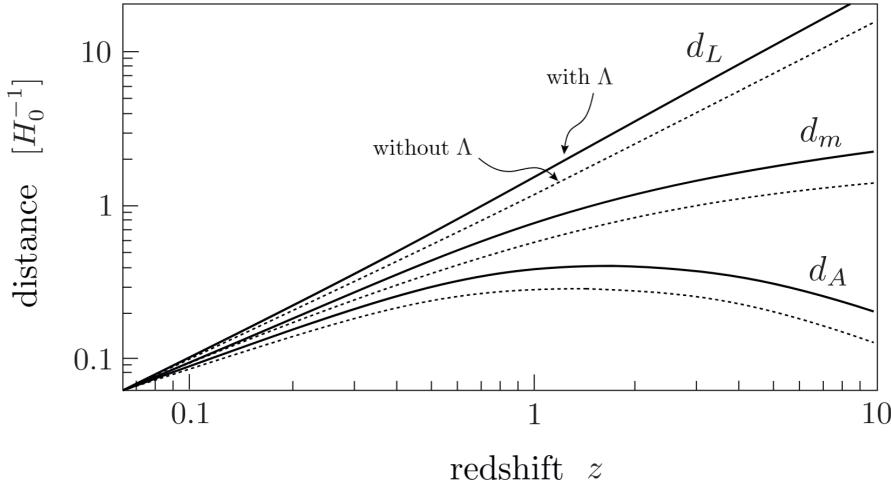


Figure 7.2: Distances measured in a flat universe, with matter only (dotted lines) and with 70% dark energy (solid lines).

where $\delta\theta \ll 1$ is the observed angular size of the object (small for all cosmological objects). In a generic FLRW spacetime, the relation between the physical transverse size of the object and its angular size in the sky instead is:

$$D = a(t)S_k(\chi)\delta\theta = \frac{d_m}{1+z}\delta\theta$$

Hence, the angular diameter distance is:

$$d_A = \frac{d_m}{1+z} \quad (7.2.12)$$

This is the measured distance between Earth and the object when the light was emitted, and in fact it is related to the luminosity distance by:

$$d_A = \frac{d_L}{(1+z)^2} \quad (7.2.13)$$

The redshift dependence of the three distances d_m , d_L and d_A is plotted in Fig. 7.2: note that the presence of a cosmological constant makes distances larger, a fact that has been crucial for the discovery of dark energy.

§7.3 Dynamics

§7.3.1 Matter sources

The requirements of spatial homogeneity and isotropy impose strict constraints on the energy-momentum tensor, so that it can only assume the form Eq. 4.5.12 for a perfect fluid.

§7.3.1.1 Number density

First, consider the number 4-current N^μ , whose components are the number density N^0 of “particles” (where here particles have a generic nature, e.g. a particle could be a galaxy) and the flux of particles N^i in the direction x^i .

Definition 7.3.1 (Isotropy and homogeneity)

A spacetime is **isotropic** if the mean value of any 3-vector vanishes, while it is **homogeneous** if the mean value of any 3-scalar^a is a function of the time coordinate only.

^aA 3-scalar is a scalar which is invariant under purely spatial coordinate transformations.

The isotropy and homogeneity conditions then impose that a comoving observer measures a number 4-current $N^\mu = (n(t), \mathbf{0})$, where $n(t)$ is the number of particles per proper volume as measured by the comoving observer. A general observer with a relative 4-velocity u^μ with respect to the rest frame of the particle distribution measures instead:

$$N^\mu = n u^\mu \quad (7.3.1)$$

Indeed, for a comoving observer $u^\mu = (1, \mathbf{0})$, while in general $u^\mu = \gamma(1, \mathbf{v})$, which is the correctly boosted result. Since the number of particles has to be conserved, in Minkowski spacetime $\dot{n} = -\partial_i N^i$, i.e. $\partial_\mu N^\mu = 0$, which generalizes to a curved spacetime as:

$$\nabla_\mu N^\mu = 0 \quad (7.3.2)$$

In the comoving frame, then, recalling Eq. 7.2.1:

$$\frac{dn}{dt} + \Gamma_{i0}^i n = 0 \implies \frac{\dot{n}}{n} = -3\frac{\dot{a}}{a} \implies n(t) \propto a^{-3}(t)$$

which means that the number density is proportional to the inverse of the proper volume, as expected.

§7.3.1.2 Energy-momentum tensor

In order to extend this analysis to the energy-momentum tensor $T_{\mu\nu}$, decompose it into a 3-scalar T_{00} , a 3-vector T_{i0} (recall that it is symmetric) and a 3-tensor T_{ij} . By isotropy $\pi_i \equiv T_{i0} = 0$ (energy-momentum flow in the x^i direction), while around a point $\mathbf{x} = \mathbf{0}$ it must be:

$$T_{ij}(\mathbf{x} = \mathbf{0}) \propto \delta_{ij} \propto g_{ij}(\mathbf{x} = \mathbf{0})$$

Since by homogeneity the proportionality constant can only depend on time, the energy-momentum tensor measured by a comoving observer takes the form:

$$T_{00} = \rho(t) \quad \pi_i \equiv T_{i0} = 0 \quad T_{ij} = -P(t)g_{ij}(t, \mathbf{x})$$

or, with mixed indices, $T_{\nu}^{\mu} = \text{diag}(\rho, -P, -P, -P)$, which is precisely the energy-momentum tensor for a perfect fluid with energy density $\rho(t)$ and pressure $P(t)$ in its comoving frame. For a general observer with a 4-velocity u^{μ} with respect to the rest frame of the fluid:

$$T_{\nu}^{\mu} = (\rho + P) u^{\mu} u_{\nu} - P \delta_{\nu}^{\mu} \quad (7.3.3)$$

To study the evolution of $\rho(t)$ and $P(t)$, recall that, in Minkowski spacetime, the fluid is subject to the continuity equation $\dot{\rho} = -\partial_i \pi^i$ and to the Euler equation $\dot{\pi}_i = \partial_i P$, which are combined into $\partial_{\mu} T_{\nu}^{\mu} = 0$. Generalizing to curved spacetime:

$$\nabla_{\mu} T_{\nu}^{\mu} = 0 \quad (7.3.4)$$

i.e. the energy momentum tensor is covariantly-conserved. Recalling Eq. 7.2.1, it is easy to see that the evolution of the energy density is given by this generalized **continuity equation**:

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + P) = 0 \quad (7.3.5)$$

which is equivalent to $dU = -PdV$, with $U \equiv \rho V$ and using $V \propto a^3$.

§7.3.1.3 Matter content

Since the universe is filled with a mixture of different matter components, it is useful to classify the various matter sources by their equation of state $\rho = \rho(P)$, which can be cast in the form:

$$\rho = wP \quad (7.3.6)$$

with $w \in \mathbb{R}$. With this relation, Eq. 7.3.5 implies the scaling relation:

$$\rho \propto a^{-3(1+w)} \quad (7.3.7)$$

Matter Matter is characterized by $w = 0$, i.e. the pressure is negligible when compared to the energy density (to be precise, $P = 0$ for pressureless dust only). Then, the dilution of $\rho \propto a^{-3}$ simply reflects the expansion of the volume $V \propto a^3$. This category has two elements:

- **baryons**, where in Cosmology this term is improperly used to denote both nuclei and electrons¹;
- **dark matter**, which is the dominant form of matter in the universe, although its nature is unknown.

Radiation Radiation is a characterized by $w = \frac{1}{3}$, which is the case, for instance, for a gas of relativistic particles, whose energy density is dominated by the kinetic energy (i.e. $p \gg m$). In this case, the dilution $\rho \propto a^{-4}$ includes an additional factor for the redshift of the energy $E \propto a^{-1}$. This category includes:

- **photons**, which are always relativistic (since they are massless particles) and dominated the early universe;
- **neutrinos**, which behaved like radiation for most of the history of the universe, and only recently started behaving like matter due to their small masses;
- **gravitons**, which are believed to have been formed in the early universe as a background of gravitational waves.

¹Since the mass of the lightest nucleus ${}^1\text{H} \equiv p^+$ is $m_p \simeq 1836m_e$, this convention is approximately justified, since the mass of baryonic matter is dominated by the mass of the nuclei.

Dark energy By experimental observations, the universe today seems to be dominated by a mysterious *negative* pressure $P = -\rho$, commonly referred to as dark energy. This mysterious form of energy has $w = -1$, hence its energy density does not dilute $\rho \propto a^0$, which means that energy has to be created as the universe expands².

A possible candidate as a form of dark energy is the vacuum energy predicted by Quantum Field Theory. Indeed, the ground state of the vacuum has the following energy-momentum tensor:

$$T_{\mu\nu}^{\text{vac}} = \rho_{\text{vac}} g_{\mu\nu}$$

which clearly implies $P_{\text{vac}} = -\rho_{\text{vac}}$. However, the predicted ρ_{vac} is off of a factor of $\sim 10^{120}$ with respect to the observed dark energy density, and this is called the “cosmological constant problem”.

§7.3.2 Curvature tensors

To compute the Einstein tensor for the FLRW metric, recall the Christoffel symbols Eq. 7.2.1.

Proposition 7.3.1 (Ricci tensor)

The Ricci tensor for a FLRW metric has non-vanishing components:

$$R_{00} = -3\frac{\ddot{a}}{a} \quad R_{ij} = -\left[\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 + 2\frac{k}{a^2}\right]g_{ij} \quad (7.3.8)$$

Proof. First, $R_{0i} = 0$ due to the isotropy of the FLRW metric. Then, contracting Eq. 2.2.13 and recalling the only non-vanishing Christoffel symbols:

$$R_{00} = -\partial_0\Gamma_{i0}^i - \Gamma_{i0}^j\Gamma_{j0}^i = -3\frac{d}{dt}\left(\frac{\dot{a}}{a}\right) - 3\left(\frac{\dot{a}}{a}\right)^2 = -3\frac{\ddot{a}}{a}$$

For the spatial components, consider the spatial metric in Cartesian coordinates (as derived in the proof of Lemma 7.1.1):

$$\gamma_{ij} = \delta_{ij} + \frac{kx_i x_j}{1 - k\mathbf{x} \cdot \mathbf{x}}$$

The Christoffel symbols depend on $\partial\gamma$ and the Ricci tensor on $\partial^2\gamma$, thus to evaluate the latter at $\mathbf{x} = 0$ one only needs the metric up to quadratic order:

$$\gamma_{ij} = \delta_{ij} + kx_i x_j + o(x^4) \Rightarrow \gamma^{ij} = \delta^{ij} - kx^i x^j + o(x^4) \Rightarrow \Gamma_{jk}^i = kx^i \delta_{jk} + o(x^3)$$

where i, j indices are raised or lowered by δ^{ij} . The Ricci tensor is then computed as:

$$\begin{aligned} R_{ij} &= \partial_\rho\Gamma_{ij}^\rho - \partial_j\Gamma_{\rho i}^\rho + \Gamma_{ij}^\lambda\Gamma_{\rho\lambda}^\rho - \Gamma_{\rho i}^\Gamma\Gamma_{j\lambda}^\rho \\ &= (\partial_0\Gamma_{ij}^0 + \partial_k\Gamma_{ij}^k) - \partial_j\Gamma_{ki}^k + (\Gamma_{ij}^0\Gamma_{k0}^k + \Gamma_{ij}^k\Gamma_{lk}^l) - (\Gamma_{ki}^0\Gamma_{j0}^k + \Gamma_{0i}^k\Gamma_{jk}^0 + \Gamma_{li}^k\Gamma_{jk}^l) \end{aligned}$$

Evaluating this expression at $\mathbf{x} = \mathbf{0}$ allows to drop the $\Gamma_{ij}^k\Gamma_{lk}^l$ term and to replace any undifferentiated γ_{ij} with δ_{ij} , so that:

$$\begin{aligned} R_{ij}(\mathbf{x} = \mathbf{0}) &= (\partial_0(a\dot{a}) + 3k - k + 3\dot{a}^2 - \dot{a}^2 - \dot{a}^2)\delta_{ij} + o(x^2) \\ &= (a\ddot{a} + 2\dot{a}^2 + 2k)\delta_{ij} + o(x^2) \end{aligned}$$

²Recall §4.5.4: since it is not possible to clearly state an energy conservation principle for a gravitational system, the only meaningful equation is Eq. 7.3.5, which is not violated by the creation of additional dark energy as the universe expands.

Spatial homogeneity and isotropy imply that $R_{ij} \propto \gamma_{ij}$, thus the general result is:

$$R_{ij} = (a\ddot{a} + 2\dot{a}^2 + 2k) \gamma_{ij} = -\frac{1}{a^2} (a\ddot{a} + 2\dot{a}^2 + 2k) g_{ij}$$

which is the thesis. \square

The Ricci scalar is then easily computed:

$$R = -6 \left[\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} \right] \quad (7.3.9)$$

Finally, the non-vanishing components of the Einstein tensor are:

$$G_{00} = 3 \left[\left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} \right] \quad G_{ij} = \left[2\frac{\ddot{a}}{a} + \left(\frac{\dot{a}}{a} \right)^2 + \frac{k}{a^2} \right] g_{ij} \quad (7.3.10)$$

§7.3.3 Friedmann equations

With Eq. 7.3.10, the Einstein field equations Eq. 4.5.6 with energy-momentum tensor Eq. 7.3.3 become (respectively the temporal and spatial components):

$$\left(\frac{\dot{a}}{a} \right)^2 = \frac{8\pi G}{3} \rho - \frac{k}{a^2} \quad \frac{\ddot{a}}{a} = -\frac{4\pi G}{3} (\rho + 3P) \quad (7.3.11)$$

where the cosmological constant has been suppressed since it is accounted for by dark energy as a form of matter content. Note that these equations are not independent, as the latter is the time derivative of the former: nonetheless, these are called the **Friedmann equations**. The energy density ρ and the pressure P are to be understood as the sums of the contributions from the various forms of matter: the contribution from radiation $\rho_r = \rho_\gamma + \rho_\nu$ (photons and neutrinos), the one from matter $\rho_m = \rho_b + \rho_c$ (baryons and cold dark matter) and the vacuum energy contribution ρ_Λ . Recalling the definition of the Hubble parameter Eq. 7.1.5, the first Friedmann equation becomes:

$$H^2 = \frac{8\pi G}{3} \rho - \frac{k}{a^2} \quad (7.3.12)$$

It is convenient to define dimensionless density parameters. To do so, observe that a flat universe corresponds to the **critical density** today:

$$\rho_{\text{crit},0} = \frac{3H_0^2}{8\pi G} \simeq 1.9 \cdot 10^{-29} h^2 \text{ g cm}^{-3}$$

where $H_0 \equiv 100h \text{ km s}^{-1} \text{Mpc}^{-1}$. Then, the dimensionless density parameters are defined as:

$$\Omega_a \equiv \frac{\rho_{a,0}}{\rho_{\text{crit},0}} \quad (7.3.13)$$

with $a = r, m, \Lambda, \dots$ running over all possible matter species. Eq. 7.3.12 can thus be rewritten as:

$$H^2(a) = H_0^2 \left[\Omega_r \left(\frac{a_0}{a} \right)^4 + \Omega_m \left(\frac{a_0}{a} \right)^3 + \Omega_k \left(\frac{a_0}{a} \right)^2 + \Omega_\Lambda \right] \quad (7.3.14)$$

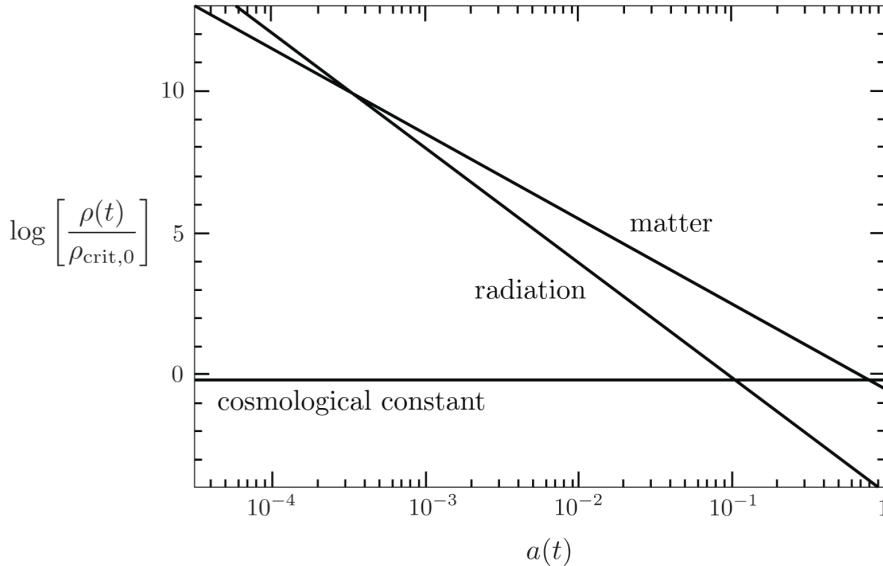


Figure 7.3: Evolution of energy densities in the universe.

where a curvature density parameter was defined as:

$$\Omega_k \equiv -\frac{k}{(a_0 H_0)^2} \quad (7.3.15)$$

Adopting the conventional normalization for the scale factor $a_0 \equiv 1$:

$$\frac{H^2}{H_0^2} = \Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda \quad (7.3.16)$$

§7.3.3.1 Λ CDM

Empirical observations show that the universe is filled with radiation, matter and dark energy so that:

$$|\Omega_k| \leq 0.01 \quad \Omega_r \simeq 9.4 \cdot 10^{-5} \quad \Omega_m \simeq 0.32 \quad \Omega_\Lambda \simeq 0.68$$

In particular, matter splits into 5% baryonic matter, i.e. with $\Omega_b \simeq 0.05$, and 27% cold dark matter, i.e. with $\Omega_c \simeq 0.27$. Moreover, note that even today curvature only contributes for less than 1% of cosmic energy, and, since $\Omega_k \propto a^{-2}$ while $\Omega_m \propto a^{-3}$ and $\Omega_r \propto a^{-4}$, in the past it was completely negligible: hence, setting $\Omega_k \equiv 0$ is justified, i.e. $k \equiv 0$.

§7.3.3.2 Single-component universe

Given the clearly distinct scalings of $\Omega_r \propto a^{-4}$, $\Omega_m \propto a^{-3}$ and $\Omega_\Lambda \propto a^0$, for most of its history the universe was dominated by a single component: first radiation, then matter and finally vacuum energy. This is illustrated in Fig. 7.3.

For a flat single-component universe, the first Friedmann equation becomes:

$$\frac{\dot{a}}{a} = H_0 \sqrt{\Omega_a} a^{-\frac{3}{2}(1+w_a)} \quad (7.3.17)$$

where w_a characterizes the component. Note that $\Omega_a \equiv 1$, since this single-component universe is flat. This equation can be integrated, obtaining the time-dependence of the scale factor:

$$a(t) \propto \begin{cases} t^{\frac{2}{3(1+w_a)}} & w_a \neq -1 \\ e^{Ht} & w_a = -1 \end{cases} \quad (7.3.18)$$

Specializing for a radiation-dominated (RD), matter-dominated (MD) and Λ -dominated (Λ D) universe:

$$a(t) \propto \begin{cases} t^{1/2} & \text{RD} \\ t^{2/3} & \text{MD} \\ e^{Ht} & \Lambda\text{D} \end{cases} \implies a(\eta) \propto \begin{cases} \eta & \text{RD} \\ \eta^2 & \text{MD} \\ -\eta^{-1} & \Lambda\text{D} \end{cases} \quad (7.3.19)$$

Chapter 8

Inflation

§8.1 Horizon problem

§8.1.1 Particle horizon

The propagation of lightrays is best studied using conformal time. Given the spatial isotropy of spacetime, it is always possible to define a coordinate system in which the lightrays move radially, so that their motion is determined by a 2-dimensional line element:

$$ds^2 = a^2(\eta) [d\eta^2 - d\chi^2]$$

where, for a flat spacetime, $\chi \equiv r$ as of Eq. 7.1.6. Since photons are massless, they move along null geodesics with $ds^2 = 0$, hence:

$$\Delta\chi(\eta) = \pm\Delta\eta$$

where the + describes outgoing photons and the - incoming photons. This means that the distance that time travels between η_1 and $\eta_2 > \eta_1$ is simply $\Delta\eta = \eta_2 - \eta_1$. Hence, assuming that the Big Bang started with a singularity¹ at $t_i \equiv 0$, then the greatest distance from which an observer at time t is able to receive signals is:

$$\chi_p(\eta) = \eta - \eta_i = \int_{t_i}^t \frac{dt}{a(t)} \quad (8.1.1)$$

which is called the **comoving particle horizon**. The particle horizon is obtained intersecting the past light-cone of the observer with the spacelike hypersurface $\eta = \eta_i$, as shown in Fig. 8.1: causal influences on the observer can only come from within this region.

The comoving particle horizon χ_p can be expressed in terms of the *comoving Hubble radius* $(aH)^{-1}$ as follows:

$$\chi_p(\eta) = \int_{a_i}^a \frac{da}{a\dot{a}} = \int_{\ln a_i}^{\ln a} \frac{d\ln a}{aH}$$

where a_i corresponds to the Big Bang singularity. For a universe dominated by a single matter component with parameter w , Eq. 7.3.17 determines the evolution of the comoving Hubble radius:

$$(aH)^{-1} = H_0^{-1} a^{\frac{1}{2}(1+3w)} \quad (8.1.2)$$

Note that for ordinary matter sources the strong energy condition (SEC) $1+3w > 0$ is satisfied, in which case the integral expression of χ_p is dominated by the upper limit, while receiving

¹Note that the Big Bang singularity is a moment in time, but *not* a point in space: indeed, in general it is pictured as an extended (possibly infinite) spacelike hypersurface.

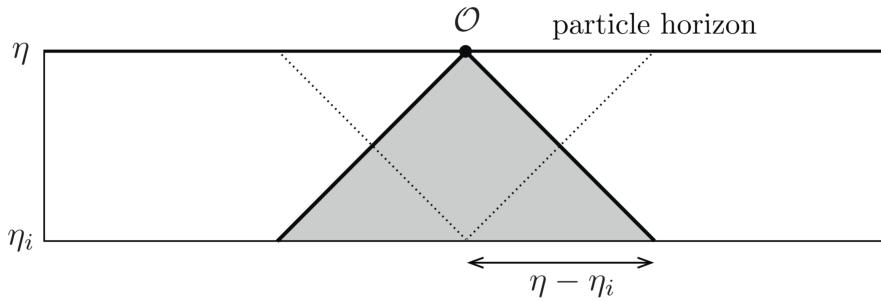


Figure 8.1: Spacetime diagram in the $\chi - \eta$ coordinate system illustrating the particle horizon.

vanishing contributions from earlier times. Performing the explicit integration:

$$\chi_p(a) = \frac{2H_0^{-1}}{1+3w} \left[a^{\frac{1}{2}(1+3w)} - a_i^{\frac{1}{2}(1+3w)} \right] \equiv \eta - \eta_i \quad (8.1.3)$$

Clearly, in presence of the SEC, $\eta_i \rightarrow 0$ as $a_i \rightarrow 0$, hence setting $a_i \equiv 0$ yields a finite comoving particle horizon:

$$\chi_p(t) = \frac{2H_0^{-1}}{1+3w} a^{\frac{1}{2}(1+3w)} = \frac{2}{1+3w} (aH)^{-1} \quad (8.1.4)$$

In the standard cosmology, then, $\chi_p \sim (aH)^{-1}$, since the Big Bang model postulates $a_i \rightarrow 0$.

§8.1.2 CMB uniformity

About 380'000 years after the Big Bang, the universe had cooled enough to allow the formation of the first hydrogen atoms: in this process, photons decoupled from the primordial plasma, a phenomenon known as Recombination (see §9.3.4) and observed today in the form of the cosmic microwave background (CMB).

The CMB is almost perfectly isotropic, with anisotropies in the order of 10^{-5} , and this is a serious problem for the Big Bang model. Indeed, $a_{\text{CMB}} \simeq 1100^{-1}$ and, since $a(\eta) \sim \eta^2$ for a matter-dominated universe, $\eta_{\text{CMB}} \sim \sqrt{a_{\text{CMB}}}$, which determines a comoving particle horizon of $\chi_{\text{CMB}} = \eta_{\text{CMB}} \sim 1100^{-1/2} \sim 1^\circ$: this means that any two points in the CMB that are separated by more than 1° in the sky are causally-disconnected.

The fact that the $\sim 10^4$ causally-disconnected patches of the CMB are remarkably isotropic is known as the *horizon problem*, since the comoving particle horizon in the Big Bang model is not sufficient to explain the isotropy of the CMB: there is not enough conformal time between the Big Bang and the CMB in the standard cosmology, as pictured in Fig. 8.2.

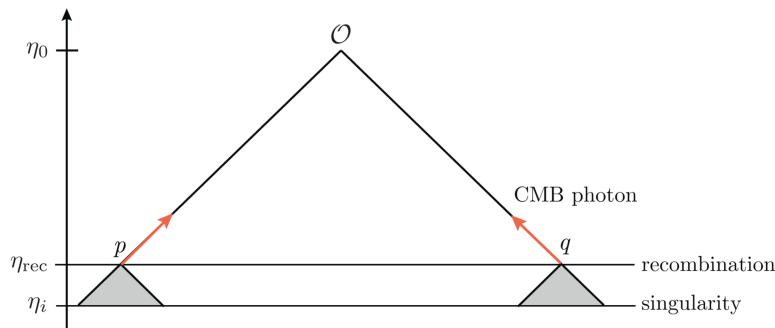


Figure 8.2: The horizon problem in the standard Big Bang model.

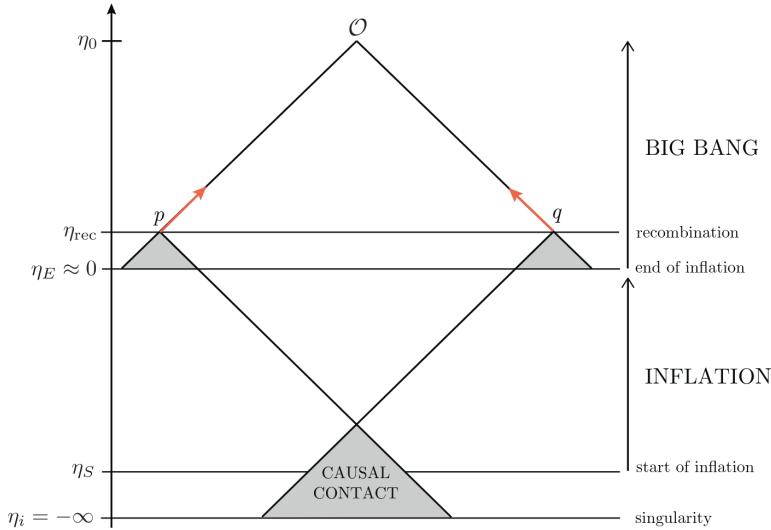


Figure 8.3: Inflationary solution to the horizon problem.

§8.1.3 SEC violation and Inflation

To explain the uniformity of the CMB, more conformal time between the Big Bang and Recombination is needed. In order to do so, the SEC must be violated: if e.g. $w = -\frac{1}{3}$, then from a certain point in time $(aH)^{-1}$ remains constant going backwards, resulting in an infinite available conformal time for the various patches of the CMB to come into causal contact with each other.

However, the preferred model is one with $w = -1$ (constrained by other observations): in this case, the comoving Hubble radius expands going backwards in the past, as shown in Fig. 8.3. Indeed:

$$\frac{d}{dt}(aH)^{-1} = \frac{1+3w}{2} \frac{1}{a} < 0$$

which confirms the shrinking of the comoving Hubble radius, i.e. its expansion going backwards. Moreover:

$$\frac{d}{dt}(aH)^{-1} = \frac{d}{dt}\dot{a}^{-1} = -\frac{\ddot{a}}{\dot{a}^2}$$

which means that a shrinking comoving Hubble radius implies an accelerated expansion, as $\ddot{a} > 0$. For this reason, this is called an *inflationary model*: in this model, the $\eta = 0$ surface is no longer the initial singularity of the Big Bang, but merely the end of the inflationary period, i.e. the transition between inflationary cosmology and standard cosmology, while the Big Bang singularity is pushed to negative conformal time:

$$\eta_i = \frac{2H_0^{-1}}{1+3w} a_i^{\frac{1}{2}(1+3w)} \xrightarrow{a_i \rightarrow 0, w < -\frac{1}{3}} -\infty$$

Moreover, since $w = -1$ in the inflationary period, the energy density of the dominant form of matter ρ_I was (approximately) constant, resulting in a constant Hubble parameter H_I and an exponential evolution of the scale parameter: $a(t) = e^{H_I t}$, from Eq. 7.3.17. The inflationary metric is then (approximately) expressed as:

$$ds^2 = dt^2 - e^{2H_I t} d\mathbf{x}^2$$

which is a de Sitter metric. In fact, Inflation is considered a period of quasi-de Sitter expansion.

§8.1.3.1 Estimates

Measurements on the CMB allow to give some estimates on the inflationary period. In particular, in the power spectrum of the CMB only wavelengths smaller than today's comoving Hubble horizon are observed, meaning that, at the start of inflation, the comoving Hubble radius had to be at least the same as today, if not larger, i.e. $(a_I H_I)^{-1} \gtrsim (a_0 H_0)^{-1}$.

On the other hand, the end of Inflation is conventionally taken to be around the GUT phase transition, at $T_{\text{GUT}} \sim 10^{15} \text{ GeV}$. Assuming that, after Inflation, the universe has been dominated by radiation only:

$$\frac{a_0 H_0}{a_e H_e} = \frac{a_0 H_0}{a_e H_0 a_e^{-2}} = a_e = \frac{T_{\text{CMB},0}}{T_e} \sim 10^{-28}$$

where $T_{\text{CMB},0} \simeq 0.2 \text{ meV}$ is the CMB temperature measured today. Then, since during Inflation $(aH)^{-1} \sim a^{-1}$, while after Inflation $(aH)^{-1} \sim a$, logarithmically $(a_I H_I)^{-1} \sim (a_0 H_0)^{-1}$ means that $(\ln a_0 - \ln a_e) \sim 2 (\ln a_0 - \ln a_I)$, i.e. $a_I \sim 10^{-56}$.

§8.2 Inflationary model

§8.2.1 Scalar field dynamics

As a simple model for Inflation, consider a real scalar field $\phi(t, \mathbf{x})$, the **inflaton**, subject to a potential $V(\phi)$: if the energy-momentum carried by the inflaton field dominates the universe, then it determines the evolution of the FLRW background. From Eq. 4.5.10 (recall the change of signature):

$$T_{\mu\nu} = \partial_\mu \phi \partial_\nu \phi - g_{\mu\nu} \left[\frac{1}{2} \partial^\rho \phi \partial_\rho \phi - V(\phi) \right]$$

Consistency with the symmetries of the FLRW spacetime requires that $\phi = \phi(t)$, hence, from $T_0^0 = \rho_\phi$, the energy density of the inflaton field is found to be:

$$\rho_\phi = \frac{1}{2} \dot{\phi}^2 + V(\phi) \quad (8.2.1)$$

which is simply the sum of the kinetic energy density and the potential energy density. Their difference is instead the pressure of the inflaton field, as inferred from $T_j^i = -P_\phi \delta_j^i$:

$$P_\phi = \frac{1}{2} \dot{\phi}^2 - V(\phi) \quad (8.2.2)$$

Since inflation requires $P_\phi < -\frac{1}{3}\rho_\phi$, for the inflaton the potential energy must dominate over the kinetic energy. Inserting ρ_ϕ into Eq. 7.3.12 (rewritten in terms of the reduced Planck mass²):

$$H^2 = \frac{1}{3M_p^2} \left[\frac{1}{2} \dot{\phi}^2 + V(\phi) \right] \quad (8.2.3)$$

Proposition 8.2.1 (Klein–Gordon Equation)

The equation of motion for the inflaton field is:

$$\ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = 0 \quad (8.2.4)$$

Proof. First, from Eq. 7.3.5, 7.3.12:

$$2H\dot{H} = \frac{\dot{\rho}}{3M_p^2} = -\frac{H}{M_p^2} (\rho + P) \implies \dot{H} = -\frac{\rho + P}{2M_p^2}$$

Since $\rho_\phi + P_\phi = \dot{\phi}$, inserting this equation in the time derivative of Eq. 8.2.3:

$$2H\dot{H} = \frac{1}{3M_p^2} \left[\dot{\phi}\ddot{\phi} + V_{,\phi}\dot{\phi} \right] \implies \ddot{\phi} + 3H\dot{\phi} + V_{,\phi} = 0$$

which is the thesis. □

Note that the evolution of the inflaton field is sourced by the potential, which acts like a force $V_{,\phi}$, while the expansion of the universe determines a *friction* term $3H\dot{\phi}$.

²In place of Newton's constant G , it is useful to use the reduced Planck mass:

$$M_p \equiv \sqrt{\frac{\hbar c}{8\pi G}} \simeq 2.4 \cdot 10^{18} \text{ GeV}$$

so that Eq. 7.3.12 for a flat spacetime can be written as $H^2 = \rho/(3M_p^2)$.

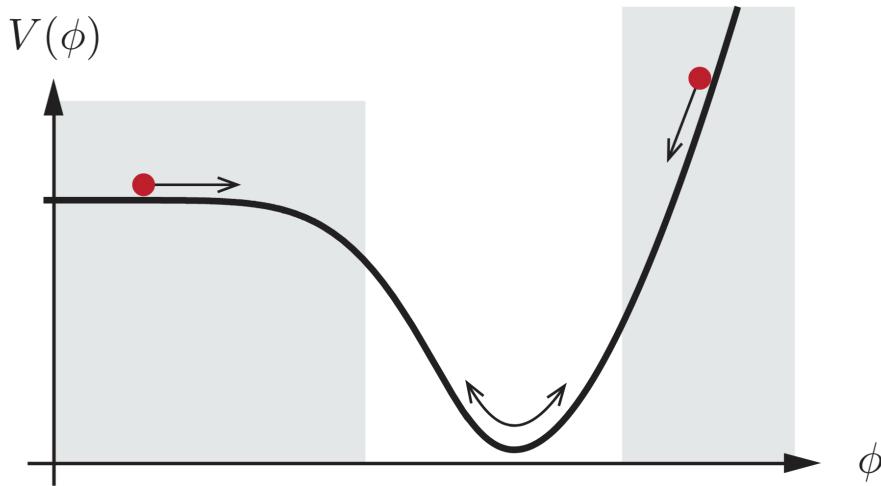


Figure 8.4: Example of slow-roll potential: inflation occurs in the shaded parts of the potential.

§8.2.1.1 Slow-roll inflation

Near a minimum ϕ_0 , the potential can be approximated by a harmonic potential:

$$V(\phi) \approx V_0 + \frac{1}{2}m^2(\phi - \phi_0)^2$$

with $V_0 \equiv V(\phi_0)$ and $m^2 \equiv V_{,\phi\phi}(\phi_0)$. Since near the minimum $\dot{\phi} \approx 0$, the energy-momentum tensor can be written as:

$$T^\mu_\nu = \begin{bmatrix} \rho_\phi & 0 & 0 & 0 \\ 0 & -P_\phi & 0 & 0 \\ 0 & 0 & -P_\phi & 0 \\ 0 & 0 & 0 & -P_\phi \end{bmatrix} \approx \begin{bmatrix} V_0 & 0 & 0 & 0 \\ 0 & V_0 & 0 & 0 \\ 0 & 0 & V_0 & 0 \\ 0 & 0 & 0 & V_0 \end{bmatrix}$$

Comparing to Eq. 4.5.7, it is clear that V_0 acts like a cosmological constant Λ . Moreover, near the minimum $w \approx -1$ as sought.

However, with a harmonic potential the inflaton field does not significantly evolve over time, while in the inflationary cosmology Inflation must only occur between a_I and a_e . To achieve this, a so-called “slow-roll model” is introduced: the potential $V(\phi)$ starts nearly constant, with a slightly negative slope, which imposes the condition $V \gg T$ and results into $w \approx -1$. Then, the potential smoothly changes from linear to parabolic, reaching finally a minimum. An example of potential is drawn in Fig. 8.4: the formulation of a precise model is futile, since it is not possible to empirically probe Inflation.

§8.2.2 Reheating

During Inflation, the energy density of the universe is dominated by the inflaton potential $V(\phi)$. Then, as Inflation ends and the inflaton field gains kinetic energy, the energy of the inflaton sector has to be transferred to the Standard Model sector: this process is called **Reheating**.

Setting WLOG $V_0 \equiv 0$, the homogeneous evolution of the inflaton field $\phi = \phi(t)$ reads:

$$\ddot{\phi} = -m^2\phi - 3H\dot{\phi}$$

The expansion timescale soon becomes much larger than the oscillation period, i.e. $H^{-1} \gg m^{-1}$, hence the friction term can be neglected. Then, the continuity equation can be written as:

$$\dot{\rho}_\phi + 3H\rho_\phi = -3HP_\phi = -\frac{3}{2}H \left[m^2\phi^2 - \dot{\phi}^2 \right]$$

Over a period of oscillation the term in parentheses averages to zero, hence the oscillating inflaton field behaves like pressureless matter with $\rho_\phi \propto a^{-3}$, while the decrease of its energy density is reflected in a decrease of its oscillation amplitude.

After Reheating, the universe is dominated by radiation and its evolution is described by the standard cosmology.

Chapter 9

Thermal History

At early times, the thermodynamical properties of the universe were determined by local equilibrium. However, non-equilibrium dynamics is crucial for massive particles to acquire cosmological abundances and to understand the CMB.

§9.1 Hot Big Bang

To understand the thermal history of the universe, the key is to compare the **rate of interactions** Γ with the **rate of expansion** H (Hubble parameter). In particular, when $\Gamma \gg H$, the timescale of particle interactions is much smaller than the expansion timescale:

$$t_\Gamma \equiv \frac{1}{\Gamma} \ll \frac{1}{H} \equiv t_H \quad (9.1.1)$$

Therefore, local thermal equilibrium is reached before the effect of the expansion becomes relevant. As the universe cools, however, the rate of interactions typically decreases faster than the expansion rate, and as $t_\Gamma \sim t_H$ the particles decouple from the thermal bath. Note that different species have different interaction rates and decouple at different times.

Consider a generic $1 + 2 \leftrightarrow 3 + 4$ scattering process. In principle, each species has its own rate of interactions, e.g. $\Gamma_1 = n_2 \sigma v_{1,2}$, where n_2 is the number density of the target species 2 and $v_{1,2}$ is the average relative velocity between species 1 and 2, but at high energies it is expected that $n_1 \sim n_2 \equiv n$, hence it is possible to define a single rate of interactions:

$$\Gamma \equiv n \sigma v \quad (9.1.2)$$

where n is the number density of particles, v is their average velocity and σ is the interaction cross-section. For $T \gtrsim 100 \text{ GeV}$, all known particles are ultra-relativistic, so $v \sim 1$, and their masses can be ignored, leaving the temperature T as the only dimensionful scale: by dimensional analysis $n \sim T^3$. Moreover, interactions are mediated by gauge bosons, which are massless above the scale of electroweak symmetry breaking, leaving a similar dependence for the cross-section of strong and electroweak processes:

$$\sigma \sim \left| \begin{array}{c} \diagdown \\ \bullet \\ \diagup \end{array} \right. \cdot \left| \begin{array}{c} \text{---} \\ \bullet \\ \text{---} \end{array} \right. \cdot \left| \begin{array}{c} \diagup \\ \bullet \\ \diagdown \end{array} \right|^2 \sim \frac{\alpha^2}{T^2}$$

where $\alpha \equiv g^2/4\pi$ is the generalized structure constant of the considered interaction. Then, dimensional analysis results into:

$$\Gamma = n \sigma v \sim T^3 \cdot \frac{\alpha^2}{T^2} \cdot 1 \sim \alpha^2 T$$

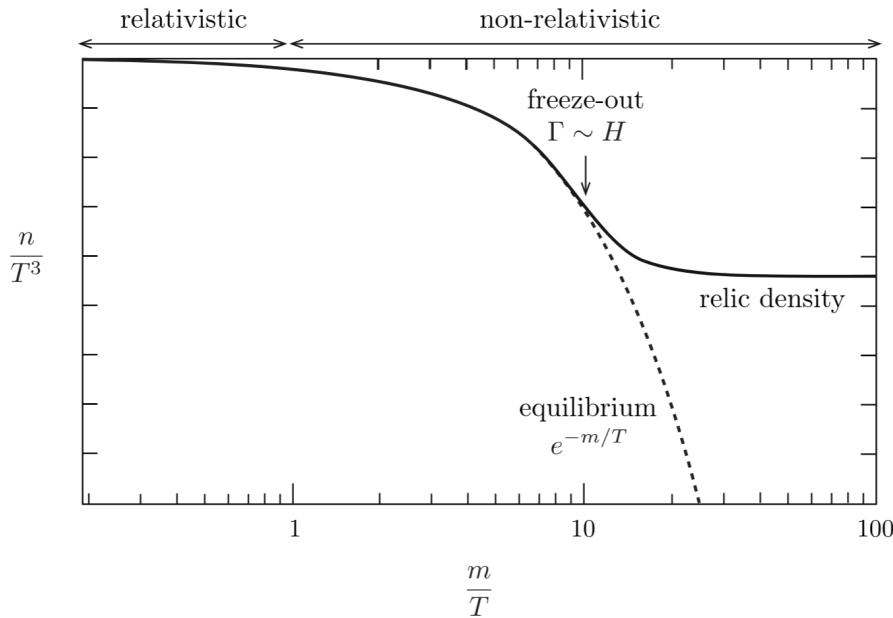


Figure 9.1: Schematic illustration of freeze-out: at high temperature ($T \gg m$) the particle abundance tracks its equilibrium value, while at low temperature ($T \ll m$) particles freeze-out and maintain a density that is much larger than the Boltzmann-suppressed equilibrium abundance.

Now, recall Eq. 7.3.12, and in particular express it in terms of the reduced Planck mass:

$$M_p \equiv \sqrt{\frac{\hbar c}{8\pi G}} \simeq 2.4 \cdot 10^{18} \text{ GeV} \quad \Rightarrow \quad H = \frac{\sqrt{\rho}}{3M_p}$$

The same argument as above gives $\rho \sim T^4$, thus $H \sim T^2/M_p$ and:

$$\frac{\Gamma}{H} \sim \frac{\alpha^2 M_p}{T} \sim \frac{10^{16} \text{ GeV}}{T} \tag{9.1.3}$$

where $\alpha \sim 0.01$ has been used as numerical estimate. Therefore, the condition Eq. 9.1.1 is satisfied for $100 \text{ GeV} \lesssim T \lesssim 10^{16} \text{ GeV}$: in this range, all SM particles are in thermal equilibrium. This means that they exchange energy and momentum efficiently, thus reaching a state of maximum entropy: in this state, the number of particles per unit volume in phase space, i.e. the *distribution function*, takes the form of the Bose–Einstein or the Fermi–Dirac distribution, depending on the statistics of the considered species.

On the contrary, when temperature drops below the mass of a particle species, that species becomes sub-relativistic and its distribution function acquires an exponential suppression of the form $\sim e^{-m/T}$: this means that relativistic particles dominate the density and pressure of the primordial plasma. Furthermore, another consequence is that, had equilibrium persisted until today, the universe would be mostly photons, since any massive particle would be exponentially suppressed: this is not the case, due to the **freeze-out** of massive particles caused by deviations from equilibrium (see Fig. 9.1 and Ex. 9.3.1).

Below the scale of electroweak symmetry breaking, i.e. $T \lesssim 100 \text{ GeV}$, the gauge bosons of weak interactions receive masses through the Higgs mechanism ($M_W \approx 80 \text{ GeV}$ and $M_Z \approx 90 \text{ GeV}$),

and the cross-section of weak processes becomes:

$$\sigma \sim \left| \begin{array}{c} \diagup \\ \bullet \\ \diagdown \end{array} \right|^2 \sim G_F^2 T^2 \quad G_F \sim \frac{\alpha}{M_W^2} \simeq 1.17 \cdot 10^{-5} \text{ GeV}^{-2}$$

The strength of weak interactions now decreases as the temperature of the universe drops, and the ratio Γ/H becomes:

$$\frac{\Gamma}{H} \sim \frac{\alpha^2 M_p T^3}{M_W^4} \sim \left(\frac{T}{1 \text{ MeV}} \right)^3 \quad (9.1.4)$$

The **decoupling** of particles which interact with the primordial plasma only through weak interactions then happens at $T_{\text{dec}} \sim 1 \text{ MeV}$.

§9.2 Equilibrium

Observational evidence (mainly the perfect blackbody spectrum of the CMB) suggests that the early universe was in *local thermal equilibrium*¹, a fact supported by the above analysis of the Standard Model at $T \gtrsim 100 \text{ GeV}$.

§9.2.1 Equilibrium thermodynamics

Consider a gas of weakly interacting particles. Quantitatively, the momentum eigenstates of a particle in a volume V have a discrete spectrum: in particular, the density of states in momentum space $\{\mathbf{p}\}$ is V/h^3 , and the density of states in phase space $\{\mathbf{x}, \mathbf{p}\}$ is $1/h^3$.

Considering a particle with g internal degrees of freedom (e.g. spin), in natural unit the density of states (which from now on is implicitly assumed to be in phase space) is $g/(2\pi)^3$. In order to obtain the number density, it is necessary to know how the particles are distributed amongst the momentum eigenstates: this information is encoded in the (*phase space*) *distribution function* $f(\mathbf{x}, \mathbf{p}, t)$. Homogeneity restricts f to be independent of \mathbf{x} , while isotropy implies that the momentum-dependence is only through the magnitude $p \equiv \|\mathbf{p}\|$. The time dependence is typically left implicit, as it manifests itself in terms of the temperature dependence of the distribution function. Putting everything together, the number density of particles in real space is found to be:

$$n(T) = \frac{g}{(2\pi)^3} \int d^3 p f(p, T) \quad (9.2.1)$$

The energy density is instead found by weighting the integration over momentum eigenstates with their energy. In particular, the assumption of the particles being weakly-interacting allows to approximate the energy as $E(p) = \sqrt{m^2 + p^2}$, so that:

$$\rho(T) = \frac{g}{(2\pi)^3} \int d^3 p f(p, T) E(p) \quad (9.2.2)$$

Pressure, on the other hand, requires more care.

Proposition 9.2.1 (Pressure of a weakly-interacting gas)

$$P = \frac{g}{(2\pi)^3} \int d^3 p f(p, T) \frac{p^2}{3E} \quad (9.2.3)$$

Proof. Consider a small area element dA with normal versor $\hat{\mathbf{n}}$. All particles with velocity $v \equiv \|\mathbf{v}\|$, striking this area element between t and $t+dt$, were located at $t=0$ on a spherical shell of radius $R = vt$ and width $v dt$, hence a solid angle $d\Omega$ of this shell defines a volume $dV = R^2 v dt d\Omega$. Given Eq. 9.2.1, the number of particles in this volume is:

$$dN = \frac{g}{(2\pi)^3} f(E, T) R^2 v dt d\Omega$$

Not all particles in dV reach the target, only those with velocities directed to the area element do. Taking into account the isotropy of the velocity distribution, the total number

¹Note that the universe can never be in perfect equilibrium, strictly speaking, since the FLRW metric does not have a time-like Killing vector. However, if the expansion is slow enough, particles have enough time to settle close to local equilibrium, and, since the universe is homogeneous, local values of thermodynamic quantities are also global values.

of these striking particles is:

$$dN_A = \frac{|\hat{\mathbf{v}} \cdot \hat{\mathbf{n}}| dA}{4\pi R^2} dN = \frac{g}{(2\pi)^3} f(E, T) v \frac{|\hat{\mathbf{v}} \cdot \hat{\mathbf{n}}|}{4\pi} dAdt d\Omega$$

where $\hat{\mathbf{v}} \cdot \hat{\mathbf{n}} < 0$ since the striking particles are directed towards the area element. Assuming elastic reflections, each particles transfers a momentum $2|\mathbf{p} \cdot \hat{\mathbf{n}}|$ to the target, hence the contribution of particles with velocity v to the pressure is:

$$P(v) = \int dN_A \frac{2|\mathbf{p} \cdot \hat{\mathbf{n}}|}{dAdt} = \frac{g}{(2\pi)^3} f(E, T) \frac{p^2}{2\pi E} \int_0^{2\pi} d\varphi \int_{-1}^0 d\cos\theta \cos^2\theta = \frac{g}{(2\pi)^3} f(E, T) \frac{p^2}{3E}$$

where $v = p/E$ was used, and θ was defined as $\hat{\mathbf{v}} \cdot \hat{\mathbf{n}} = -\cos\theta < 0$. Integrating over energy or momentum yields the thesis. \square

The form of the distribution function depends on the thermodynamical state of the system. Assuming that the gas is in *kinetic equilibrium*, i.e. if the particles can exchange energy and momentum efficiently, then the system is in a state of maximal entropy and the distribution function is either the Fermi–Dirac distribution (fermions) or the Bose–Einstein distribution (bosons):

$$f(p, T) = \left[e^{\frac{E(p)-\mu(T)}{T}} \pm 1 \right]^{-1} \quad (9.2.4)$$

with the + sign for fermions and the – sign for bosons. At low temperatures, i.e. $T < E - \mu$, both distributions reduce to the Maxwell–Boltzmann distribution:

$$f(p, T) \approx e^{-\frac{E(p)-\mu(T)}{T}} \quad (9.2.5)$$

The chemical potential $\mu(T)$ characterizes the response of a system to a change in particle number: the second law of thermodynamics implies that, given a reaction, particles flow towards the side of the reaction with the lower total chemical potential, until *chemical equilibrium* is achieved:

$$\sum_{\text{reactants}} \mu_i = \sum_{\text{products}} \mu_j$$

Note that, since the number of photons is not conserved (e.g. double Compton scattering $e^- \gamma \leftrightarrow e^- \gamma \gamma$), then $\mu_\gamma = 0$ and $\mu_{\bar{X}} = -\mu_X$, due to particle-antiparticle annihilation $X\bar{X} \leftrightarrow \gamma\gamma$. A system is in **thermal equilibrium** if its species are both in kinetic and chemical equilibrium: these species then share a common temperature $T_i = T$, which is often identified with the photon temperature T_γ (the “temperature of the universe”).

§9.2.2 Density and pressure

At early times, the chemical potentials of all particles are so small that they can be neglected, thus Eq. 9.2.1–9.2.2 can be written as:

$$n(T) = \frac{g}{2\pi^2} T^3 I_\pm(x) \quad \rho(T) = \frac{g}{2\pi^2} T^4 J_\pm(x) \quad (9.2.6)$$

with $x \equiv m/T$, $\xi \equiv p/T$ and:

$$I_\pm(x) \equiv \int_0^\infty d\xi \frac{\xi^2}{\exp \sqrt{\xi^2 + x^2} \pm 1} \quad J_\pm(x) \equiv \int_0^\infty d\xi \frac{\xi^2 \sqrt{\xi^2 + x^2}}{\exp \sqrt{\xi^2 + x^2} \pm 1} \quad (9.2.7)$$

These integrals need, in general, to be evaluated numerically, but analytic expressions are possible in the ultra-relativistic and sub-relativistic limits.

Lemma 9.2.1 (Ultra-relativistic limit)

In the limit of $T \gg m$, i.e. $x \rightarrow 0$:

$$n(T) = \frac{\zeta(3)}{\pi^2} g T^3 \begin{cases} 1 & \text{bosons} \\ \frac{3}{4} & \text{fermions} \end{cases} \quad \rho(T) = \frac{\pi^2}{30} g T^4 \begin{cases} 1 & \text{bosons} \\ \frac{7}{8} & \text{fermions} \end{cases} \quad (9.2.8)$$

Proof. For $x \rightarrow 0$, the integrals in Eq. 9.2.7 reduce to:

$$I_{\pm}(0) = \int_0^\infty d\xi \frac{\xi^2}{e^\xi \pm 1} \quad J_{\pm}(0) = \int_0^\infty d\xi \frac{\xi^3}{e^\xi \pm 1}$$

These integrals can be solved using the standard integral:

$$\int_0^\infty d\xi \frac{\xi^n}{e^\xi - 1} = \zeta(n+1)\Gamma(n+1) \quad (9.2.9)$$

In the bosonic case they are straightforward: $I_-(0) = 2\zeta(3)$ and $J_-(0) = 6\zeta(4) = \frac{\pi^4}{15}$. In the fermionic case, use instead:

$$\frac{1}{e^\xi + 1} = \frac{1}{e^\xi - 1} - \frac{2}{e^{2\xi} - 1}$$

so that:

$$I_+(0) = I_-(0) - 2 \left(\frac{1}{2}\right)^3 I_-(0) = \frac{3}{2} I_-(0)$$

$$J_+(0) = J_-(0) - 2 \left(\frac{1}{2}\right)^4 J_-(0) = \frac{7}{8} J_-(0)$$

which concludes the proof. \square

From Eq. 9.2.3 with $E \simeq p$ in the relativistic limit, the equation of state for a relativistic gas $P = \frac{1}{3}\rho$ is recovered. Moreover, recalling Eq. 7.3.13, using $T_0 = 2.73$ K it is possible to compute the density parameter of radiation:

$$\rho_{\gamma,0} = \frac{\pi^2}{15} T_0^4 \simeq 4.6 \cdot 10^{-34} \text{ g cm}^{-3} \implies \Omega_\gamma h^2 \simeq 2.5 \cdot 10^{-5}$$

Lemma 9.2.2 (Sub-relativistic limit)

In the limit $T \ll m$, i.e. $x \gg 1$:

$$n(T) = g \left(\frac{mT}{2\pi}\right)^{3/2} e^{-m/T} \quad \rho(T) = mn(T) + \frac{3}{2}n(T)T \quad (9.2.10)$$

Proof. For $x \gg 1$:

$$I_{\pm}(x) \approx \int_0^\infty d\xi \xi^2 e^{-\sqrt{\xi^2+x^2}}$$

Most of the contribution comes from the $\xi \ll x$ regime, hence, at lowest order in ξ :

$$I_{\pm}(x) \approx \int_0^{\infty} d\xi \xi^2 \exp\left(-x - \frac{\xi^2}{2x}\right) = (2x)^{3/2} e^{-x} \int_0^{\infty} d\xi \xi^2 e^{-\xi^2}$$

Recalling the standard integral:

$$\int_0^{\infty} d\xi \xi^n e^{-\xi^2} = \frac{1}{2} \Gamma\left(\frac{n+1}{2}\right) \quad (9.2.11)$$

with $\Gamma(\frac{3}{2}) = \frac{\sqrt{\pi}}{2}$:

$$I_{\pm}(x) = \sqrt{\frac{\pi}{2}} x^{3/2} e^{-x} \implies n(T) = g \left(\frac{mT}{2\pi}\right)^{3/2} e^{-m/T}$$

For the energy density, use the sub-relativistic expansion at lowest order in ξ in Eq. 9.2.7:

$$\begin{aligned} J_{\pm}(x) &\approx \int_0^{\infty} d\xi \xi^2 \left(x + \frac{\xi^2}{2x}\right) \exp\left(-x - \frac{\xi^2}{2x}\right) = (2x)^{3/2} e^{-x} \int_0^{\infty} d\xi \xi^2 (x + \xi^2) e^{-\xi^2} \\ &= (2x)^{3/2} e^{-x} \frac{1}{2} \left[x \Gamma\left(\frac{3}{2}\right) + \Gamma\left(\frac{5}{2}\right)\right] = \sqrt{\frac{\pi}{2}} x^{3/2} e^{-x} \left(x + \frac{3}{2}\right) \end{aligned}$$

Inserting $x = m/T$ yields the thesis. □

In a similar way, from Eq. 9.2.3 it can be shown that $P = nT$, which means that a non-relativistic gas behaves like pressureless dust since $P = nT \ll nm = \rho$.

As expected, in the sub-relativistic limit $T \ll m$ the number density, the energy density and the pressure of massive particle species fall exponentially (Boltzmann suppression). This is interpreted as the annihilation of particles with antiparticles: at high temperature (i.e. energy), matter-antimatter annihilation is balanced by pair production, but as T decreases below the mass of the particle the thermal energy is no longer sufficient for pair production and the number density decreases exponentially.

§9.2.3 Effective number of relativistic species

From now on, the temperature of the universe is identified with the temperature of the photon gas: $T \equiv T_{\gamma}$.

Definition 9.2.1 (Effective number of relativistic degrees of freedom)

The total radiation density in the universe is defined as:

$$\rho_r := \sum_i \rho_i \equiv \frac{\pi^2}{30} g_{\star}(T) T^4 \quad (9.2.12)$$

where the sum runs over all relativistic particle species. The **effective number of relativistic degrees of freedom** is denoted by $g_{\star}(T)$.

Note that $g_{\star}(T)$ receives different contributions from relativistic species in thermal equilibrium with the photons, i.e. with $T = T_i \gg m_i$, and from decoupled relativistic species, i.e. with

$T \neq T_i \gg m_i$. In particular, from Eq. 9.2.8:

$$g_{\star}^{\text{eq}}(T) = \sum_{\text{bosons}} g_i + \frac{7}{8} \sum_{\text{fermions}} g_j \quad g_{\star}^{\text{dec}}(T) = \sum_{\text{bosons}} g_i \left(\frac{T_i}{T}\right)^4 + \frac{7}{8} \sum_{\text{fermions}} g_j \left(\frac{T_j}{T}\right)^4 \quad (9.2.13)$$

When the temperature drops below the mass m_i of a particle species, it becomes non-relativistic and no longer contributes to $g_{\star}(T)$.

Example 9.2.1 ($g_{\star}(T)$ above 100 GeV)

For $T \gtrsim 100$ GeV, all SM particles are relativistic (as noted in §9.1), hence all species contribute to $g_{\star}^{\text{eq}}(T)$. The bosons have:

$$g_b = g_{\gamma} + g_g + g_{W,Z} + g_H = 2_{\text{spin}} + 8_{\text{colour}} \cdot 2_{\text{spin}} + 3_{\text{SU}(2)} \cdot 3_{\text{spin}} + 1_{\text{spin}} = 28$$

since massless vector bosons only have transverse polarization. The fermions, on the other hand, have:

$$\begin{aligned} g_f &= g_q + g_{e,\mu,\tau} + g_{\nu} \\ &= 6_{\text{species}} \cdot 2_{\text{spin}} \cdot 2_{\text{anti}} \cdot 3_{\text{colour}} + 3_{\text{species}} \cdot 2_{\text{spin}} \cdot 2_{\text{anti}} + 3_{\text{species}} \cdot 1_{\text{spin}} \cdot 2_{\text{anti}} = 90 \end{aligned}$$

since neutrinos only have one helicity state (left-handed neutrinos and right-handed antineutrinos). The effective number of relativistic degrees of freedom then is:

$$g_{\star} = g_b + \frac{7}{8} g_f = 106.75 \quad (9.2.14)$$

As the temperature drops, various particle species become non-relativistic and annihilate: the evolution of $g_{\star}(T)$ is plotted in Fig. 9.2. At $T \sim 150$ MeV, before the strange quarks have time to annihilate, the QCD phase transition takes place: quarks combine into baryons and mesons, which are all non-relativistic under 150 MeV, except for the pions (π^{\pm}, π^0). The only relativistic species left are then pions, electrons, muons, neutrinos and photons: since the pions are spin-0 particles, they contribute to g_b with $g_{\pi} = 3_{\text{species}} \cdot 1_{\text{spin}}$, resulting in $g_{\star} = 2 + 3 + \frac{7}{8} \cdot (4 + 4 + 6) = 17.25$.

Moreover, it is important to note that the transition from relativistic to non-relativistic behaviour is not instantaneous: about 80% of the particle-antiparticle annihilations takes place in the interval $T = m \rightarrow T = \frac{1}{6}m$.

§9.2.3.1 Entropy

According to the second law of thermodynamics, the total entropy of the universe can only increase or stay constant. Since the fraction of baryons to photons is measured to be:

$$\eta_b \equiv \frac{n_b}{n_{\gamma}} \simeq 5.5 \cdot 10^{-10} \frac{\Omega_b h^2}{0.020} \sim 10^{-9} \quad (9.2.15)$$

the entropy of the universe is dominated by the entropy of the photon bath (at least as long as the universe is sufficiently uniform): any entropy production from non-equilibrium processes is then insignificant relative to the total entropy, and therefore the expansion of the universe can be treated as *adiabatic*, so that the total entropy stays constant, even beyond equilibrium.

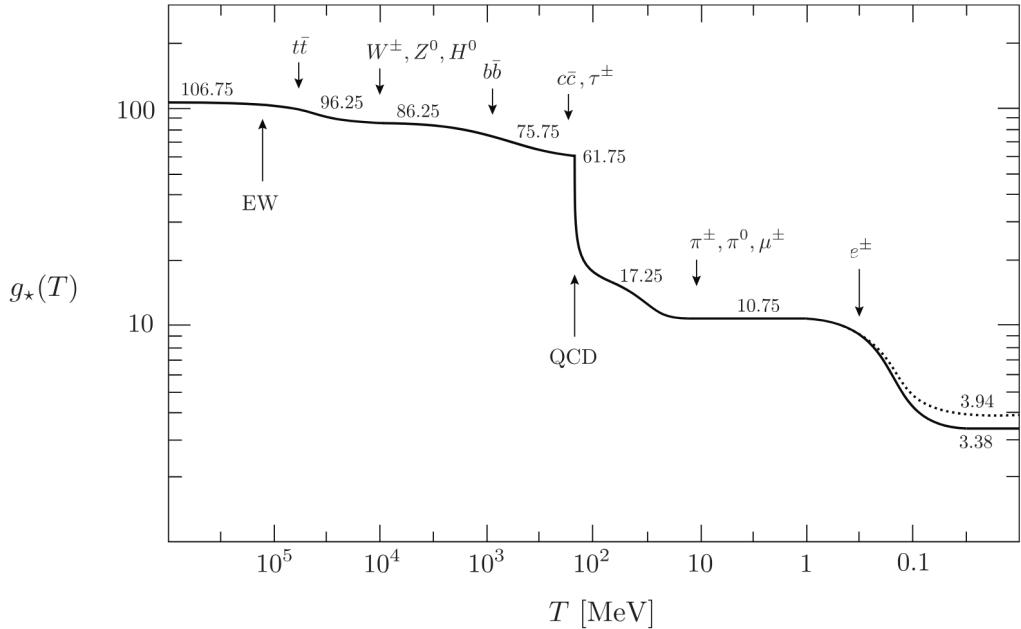


Figure 9.2: Evolution of $g_*(T)$ with temperature, assuming Standard Model particles. The dotted line is the effective number of degrees of freedom in entropy $g_{*S}(T)$.

Lemma 9.2.3 (Entropy at equilibrium)

The entropy of a system at equilibrium stays constant.

Proof. Consider the second law of thermodynamics: $TdS = dU + PdV$. Using $U = \rho V$:

$$dS = \frac{d[(\rho + P)V]}{T} - \frac{VdP}{T} = \frac{V}{T} \frac{\partial \rho}{\partial T} dT + \frac{\rho + P}{T} dV$$

Since the entropy function must exist, this differential form must be exact. By Th. 1.4.1, an exact form is also closed, hence:

$$\frac{\partial}{\partial V} \left[\frac{V}{T} \frac{\partial \rho}{\partial T} \right] = \frac{\partial}{\partial T} \left[\frac{\rho + P}{T} \right] \implies \frac{\partial P}{\partial T} = \frac{\rho + P}{T} \quad (9.2.16)$$

Rearranging the second law of thermodynamics, then:

$$dS = \frac{d[(\rho + P)V]}{T} - \frac{(\rho + P)V}{T^2} dT = d \left[\frac{\rho + P}{T} V \right]$$

To show that entropy is conserved at equilibrium, consider:

$$\frac{dS}{dt} = \frac{d}{dt} \left[\frac{\rho + P}{T} V \right] = \frac{V}{T} \left[\dot{\rho} + \frac{\dot{V}}{V} (\rho + P) \right] + \frac{V}{T} \left[\dot{P} - \frac{\rho + P}{T} \dot{T} \right]$$

Since $V \propto a^3$, then $\dot{V}/V = 3H$ and the first term vanishes by Eq. 7.3.5, while the second vanishes by Eq. 9.2.16. Therefore, $\dot{S} = 0$ at equilibrium. \square

It is convenient to define the **entropy density**, which ignoring an integration constant is (by

the above proof):

$$s = \frac{\rho + P}{T} \quad (9.2.17)$$

which, by adiabaticity (i.e. $dQ = 0 \implies dS = 0$), scales like $s \propto a^{-3}$. Since for a relativistic gas $P = \frac{1}{3}\rho$, recalling Eq. 9.2.8 it is possible to express the entropy density in the ultra-relativistic limit as:

$$s = \frac{2\pi^2}{45} g(T) T^3 \quad (9.2.18)$$

Definition 9.2.2 (Effective number of relativistic degrees of freedom in entropy)

The total entropy density in the universe is defined as:

$$s := \sum_i s_i \equiv \frac{2\pi^2}{45} g_{*S}(T) T^3 \quad (9.2.19)$$

where the sum runs over all relativistic particle species. The **effective number of degrees of freedom in entropy** is denoted by $g_{*S}(T)$.

As for $g_*(T)$, there are two different contributions to $g_{*S}(T)$ too: one from species in thermal equilibrium with the photon bath and one from decoupled species. Note that $g_{*S}^{\text{eq}}(T) = g_*^{\text{eq}}(T)$, while, since $s_i \propto T_i^3$, for decoupled species:

$$g_{*S}^{\text{dec}}(T) = \sum_{\text{bosons}} g_i \left(\frac{T_i}{T} \right)^3 + \frac{7}{8} \sum_{\text{fermions}} g_j \left(\frac{T_j}{T} \right)^3 \quad (9.2.20)$$

which is $g_{*S}^{\text{dec}}(T) \neq g_*^{\text{dec}}(T)$. It follows that $g_*(T) = g_{*S}(T)$ only when *all* relativistic species are in thermal equilibrium with the photon bath: in the real universe, this is the case until $t \approx 1 \text{ sec}$ (see Fig. 9.2). The conservation of entropy has two important consequences:

- since $s \propto a^{-3}$, if the number of particles in a comoving volume N_i is conserved (i.e. $n_i \propto a^{-3}$), then $n_i \propto s$, with $N_i = n_i/s$ since it is a scale-invariant quantity. This is the case, for example, for the net baryon number after baryogenesis, which is conserved, hence $N_{\text{net}} \equiv (n_B - n_{\bar{B}})/s$ is constant;
- from the definition of s , it follows that $T \propto g_{*S}^{-1/3} a^{-1}$. Away from mass thresholds, g_{*S} is approximately constant and $T \propto a^{-1}$ as expected, but when a particle species becomes non-relativistic the factor g_{*S} accounts for the transfer of its entropy to other relativistic species, making T change in the scaling as a consequence.

§9.2.3.2 Neutrino decoupling

Neutrinos are coupled to the thermal bath only through weak interaction processes (e.g. $\nu_e \bar{\nu}_e \leftrightarrow e^+ e^-$), whose cross-section is estimated as $\sigma \sim G_F^2 T^2$. By Eq. 9.1.2, then, $\Gamma \sim G_F^2 T^5$, which allows to estimate the temperature for neutrino decoupling (with $H \sim T^2/M_p$):

$$\frac{\Gamma}{H} \sim \left(\frac{T}{1 \text{ MeV}} \right)^3 \implies T_{\text{dec}} \sim 1 \text{ MeV}$$

A more accurate computation gives $T_{\text{dec}} \sim 0.8 \text{ MeV}$. After decoupling, neutrinos move freely along geodesics. Since $p \propto a^{-1}$ (by Eq. 7.2.3), it is convenient to define the time-independent

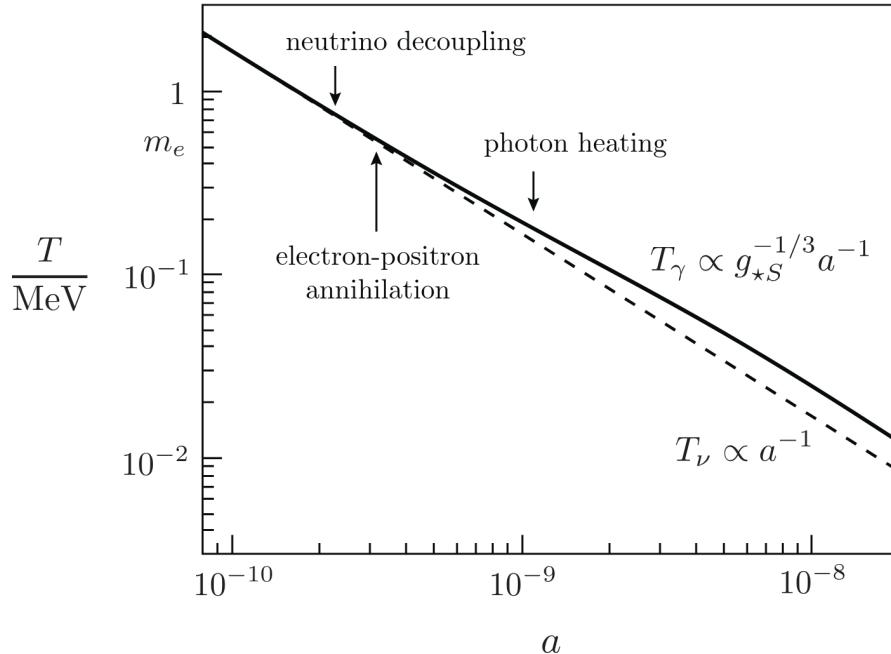


Figure 9.3: Difference in redshift between neutrino temperature $T_\nu \propto a^{-1}$ and photon temperature $T_\gamma \propto g_{*S}^{-1/3} a^{-1}$.

combination $q \equiv ap$, so that the neutrino number density after decoupling can be written as:

$$n_\nu \propto a^{-3} \int d^3q \left[\exp \frac{q}{aT_\nu} + 1 \right]^{-1}$$

But, after decoupling, particle number conservation requires $n_\nu \propto a^{-3}$, hence the neutrino temperature must scale as $T_\nu \propto a^{-1}$. The neutrinos then follow a relativistic Fermi–Dirac distribution with decreasing temperature, even after they become non-relativistic at later times (when $T_\nu \ll m_\mu$), due to the Liouville theorem².

§9.2.3.3 Electron-positron annihilation

Shortly after the neutrinos decouple, the temperature drops below the electron mass ($m_e \simeq 0.511 \text{ MeV}$), so that electron-positron annihilation is no longer balanced by the $\gamma\gamma \rightarrow e^-e^+$ pair production. The energy density and entropy of electrons and positrons are then transferred to the photons, but not to the decoupled neutrinos: the photons are thus heated relative to the neutrinos (see Fig. 9.3).

To quantify this effect, consider the change in the effective number of degrees of freedom in entropy. Neglecting neutrinos and other decoupled species³:

$$g_{*S}^{\text{eq}} = \begin{cases} 2_\gamma + \frac{7}{8} \cdot 4_e = \frac{11}{2} & T \gtrsim m_e \\ 2_\gamma = 2 & T < m_e \end{cases}$$

²The Liouville theorem states that, in a Hamiltonian system, the phase-space probability density (i.e. distribution function) is constant along the trajectories defined by the Hamilton equations of motion.

³Entropy is conserved separately for the thermal bath and the decoupled species, since they are non-interacting systems.

Since, in equilibrium, $g_{\star S}^{\text{eq}} T^3 a^3$ remains constant, the photon temperature $aT_\gamma \equiv aT$ must increase by a factor of $(11/4)^{1/3}$ after electron-positron annihilation, while aT_ν remains the same. This means that, for $T < m_e$:

$$T_\nu = \sqrt[3]{\frac{4}{11}} T_\gamma \quad (9.2.21)$$

For $T \ll m_e$, the only remaining relativistic species are photons and neutrinos, and the above equation implies that $g_\star \neq g_{\star S}$:

$$g_\star = 2_\gamma + \frac{7}{8} \cdot 2_{\text{anti}} \cdot N_{\text{eff}} \left(\frac{4}{11} \right)^{4/3} = 3.36 \quad g_{\star S} = 2_\gamma + \frac{7}{8} \cdot 2_{\text{anti}} \cdot N_{\text{eff}} \left(\frac{4}{11} \right) = 3.94$$

where N_{eff} is the **effective number of neutrino species**. If neutrino decoupling was instantaneous, then $N_{\text{eff}} = 3$; however, neutrino decoupling was not complete when e^+e^- annihilation began, so a fraction of the released energy and entropy did leak to the neutrinos, resulting into $N_{\text{eff}} = 3.046$ (the current bound from the Planck satellite is $N_{\text{eff}} = 3.04 \pm 0.18$).

Note that Eq. 9.2.21 holds until the present, thus allowing to estimate the temperature of the **cosmic neutrino background** (CνB): given the present CMB temperature $T_{\text{CMB}} = 2.73 \text{ K} = 0.24 \text{ meV}$, it is $T_{\text{C}\nu\text{B}} = 1.95 \text{ K} = 0.17 \text{ meV}$. From Eq. 9.2.8, the number density of neutrinos is:

$$n_\nu = \frac{3}{4} N_{\text{eff}} \cdot \frac{4}{11} n_\gamma \quad (9.2.22)$$

which today corresponds to $n_\nu/N_{\text{eff}} \simeq 112 \text{ cm}^{-3}$ per flavour, as $n_\gamma(2.73 \text{ K}) \simeq 5 \text{ K}^3 \simeq 415 \text{ cm}^{-3}$. On the other hand, the present energy density of neutrinos depends on whether they are relativistic or non-relativistic today:

- massless neutrinos: massless neutrinos are always relativistic, hence, by Eq. 9.2.8:

$$\rho_\nu = \frac{7}{8} N_{\text{eff}} \left(\frac{4}{11} \right)^{4/3} \rho_\gamma \implies \Omega_\nu h^2 \simeq 1.7 \cdot 10^{-5}$$

were $\Omega_\gamma h^2 \simeq 2.5 \cdot 10^{-5}$ was used;

- massive neutrinos: experiments on neutrino oscillations impose constraints on the sum of neutrino masses: $60 \text{ meV} < \sum_i m_{\nu_i} < 1 \text{ eV}$. This means that neutrinos behave as radiation-like particles in the early universe ($T \gg m_\nu$) and as matter-like particles in the late universe ($T \ll m_\nu$), which means that $\rho_{\nu,0} \approx \sum_i m_{\nu_i} n_{\nu_i,0}$ (by Eq. 9.2.10). Using the estimate $n_{\nu_i} \simeq 112 \text{ cm}^{-3}$:

$$\Omega_\nu h^2 \approx \frac{\sum_i m_{\nu_i}}{95 \text{ eV}} \quad (9.2.23)$$

Hence neutrinos are a subdominant component, with $0.001 \lesssim \Omega_\nu \lesssim 0.02$ (using $h = 0.7$).

§9.2.4 Thermal history

In radiation era $a \propto \sqrt{t}$, hence $H = \frac{1}{2t}$ and, using Eq. 9.2.12:

$$\frac{1}{2t} = H \simeq \sqrt{\frac{\rho_r}{3M_p^2}} \simeq \frac{\pi}{3} \sqrt{\frac{g_\star}{10}} \frac{T^2}{M_p} \implies \frac{T}{1 \text{ MeV}} \simeq 1.5 g_\star^{-1/4} \sqrt{\frac{1 \text{ sec}}{t}} \quad (9.2.24)$$

Therefore, 1 second after the Big Bang the universe was about 1 MeV. It is also possible to state an approximated thermal history of the universe (recalling Eq. 7.2.6):

Event	time t	redshift z	temperature T
Singularity	0	∞	∞
Inflation	$\sim 10^{-35}$ s	$\sim 10^{28}$	$\sim 10^{15}$ GeV
Baryogenesis	$\lesssim 20$ ps	$> 10^{15}$	> 100 GeV
EW phase transition	20 ps	10^{15}	100 GeV
QCD phase transition	$20 \mu\text{s}$	10^{12}	150 MeV
Dark matter freeze-out	?	?	?
Neutrino decoupling	1 s	$6 \cdot 10^9$	0.8 MeV
e^-e^+ annihilation	6 s	$2 \cdot 10^9$	500 keV
Big Bang nucleosynthesis	3 min	$4 \cdot 10^8$	100 keV
Matter-radiation equality	60 kyr	3400	0.75 eV
Recombination	$260 - 380$ kyr	$1100 - 1400$	$0.26 - 0.33$ eV
Photon decoupling (CMB)	380 kyr	1100	0.26 eV
Reionization	$100 - 400$ Myr	$10 - 30$	$2.6 - 7.0$ meV
Λ -matter equality	9 Gyr	0.4	0.33 meV
Today	13.8 Gyr	0	0.24 meV

Note that the expansion of the universe during Inflation had a very different timescale with respect to today, although in both cases it is exponential ($a \propto \exp Ht$):

$$H_I \sim \frac{T_I^2}{M_p} \sim 10^{11} \text{ GeV} \quad \Rightarrow \quad t_{H_I} \equiv \frac{1}{H_I} \sim 10^{-35} \text{ sec}$$

$$H_0 \approx \frac{70 \text{ km s}^{-1}}{1 \text{ Mpc}} \quad \Rightarrow \quad t_{H_0} \equiv \frac{1}{H_0} \sim 4.5 \cdot 10^{17} \text{ sec}$$

The expansion during Inflation was $\sim 10^{52}$ times faster than it is today.

§9.3 Beyond equilibrium

§9.3.1 Boltzmann equation

In the absence of interactions, the number of particles in a fixed physical volume $V \propto a^3$ is conserved, thus $n_i \propto a^{-3}$ and:

$$\frac{\dot{n}_i}{n_i} = -3\frac{\dot{a}}{a} \iff \frac{1}{a^3} \frac{d(n_i a^3)}{dt} = 0$$

To include the effect of interactions, a collision term $C_i(\{n_j\})$ is added, which depends on the number density of all particle species:

$$\frac{1}{a^3} \frac{d(n_i a^3)}{dt} = C_i(\{n_j\}) \quad (9.3.1)$$

This is the **Boltzmann equation**, which describes the behaviour of the system beyond equilibrium. The form of the collision term depends on the specific interactions under consideration: interactions between three or more particles are very unlikely, so it is reasonable to assume only single-particle decays and two-particle scatterings/annihilations.

Example 9.3.1 (2 \leftrightarrow 2 scattering)

Consider a generic process $1 + 2 \leftrightarrow 3 + 4$, and suppose to track the number density n_1 of species 1. The rate of change in the abundance of species 1 is given by the difference between the rates of annihilation (direct process) and production (inverse process) of the species, hence:

$$\frac{1}{a^3} \frac{d(n_1 a^3)}{dt} = -\alpha n_1 n_2 + \beta n_3 n_4$$

The first term describes the direct process, and the parameter $\alpha = \langle \sigma v \rangle$ is the *thermally-averaged cross-section*^a. To determine the parameter β , note that the collision term must vanish at (chemical) equilibrium, so that:

$$\beta = \left(\frac{n_1 n_2}{n_3 n_4} \right)_{\text{eq}} \alpha$$

where the equilibrium number densities are given by Eq. 9.2.1. The Boltzmann equation then becomes:

$$\frac{1}{a^3} \frac{d(n_1 a^3)}{dt} = -\langle \sigma v \rangle \left[n_1 n_2 - \left(\frac{n_1 n_2}{n_3 n_4} \right)_{\text{eq}} n_3 n_4 \right] \quad (9.3.2)$$

Rewriting this equation in terms of the number of particles in a comoving volume $N_i \equiv n_i / s$:

$$\frac{d \ln N_1}{d \ln a} = -\frac{\Gamma_1}{H} \left[1 - \left(\frac{N_1 N_2}{N_3 N_4} \right)_{\text{eq}} \frac{N_3 N_4}{N_1 N_2} \right] \quad (9.3.3)$$

with the interaction rate $\Gamma_1 \equiv n_2 \langle \sigma v \rangle$. The r.h.s. contains two different factors: the Γ_1/H factor, describing the **interaction efficiency**, and the $[1 - \dots]$ factor, describing the **deviation from equilibrium**. There are two main regimes for the reaction:

- $\Gamma_1 \gg H$: the natural state of the system is chemical equilibrium since, given the large interaction rate, if $N_1 \gg N_1^{\text{eq}}$ the r.h.s. is negative and N_1 reduces towards N_1^{eq} , while

if $N_1 \ll N_1^{\text{eq}}$ the r.h.s. is positive and N_1 increases towards N_1^{eq} . At equilibrium, then, the r.h.s. vanishes and particles assume their equilibrium abundances;

- $\Gamma_1 < H$: when the reaction rate drops below the Hubble scale, the r.h.s. gets suppressed and the comoving density of particles approaches a constant *relic density*, a phenomenon known as **freeze-out**.

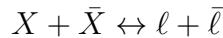
^aThe angle brackets denote an average over the relative velocity v between particles 1 and 2.

§9.3.2 Dark matter relics

In this section, the hypothesis that dark matter is composed of weakly interacting massive particles (WIMPs) is assumed. WIMPs were in close contact with the rest of the cosmic plasma at high temperatures, but then experiences freeze-out at a critical temperature T_f , which can be estimated using the Boltzmann equation.

§9.3.2.1 Dark matter freeze-out

Some assumptions on the WIMP interactions have to be made. In particular, assume that a heavy dark matter particle X and its antiparticle \bar{X} can annihilate to produce two light (essentially massless) particle:



Moreover, assume the light particles to be tightly coupled to the cosmic plasma (e.g. if they are charged), so that they maintain their equilibrium densities $n_\ell = n_\ell^{\text{eq}}$, and also that there are no initial asymmetries between X and \bar{X} , i.e. $n_X = n_{\bar{X}}$. Eq. 9.3.2 for the number of WIMPs in a comoving volume $N_X \equiv n_X/s$ then is:

$$\frac{dN_X}{dt} = -s \langle \sigma v \rangle [N_X^2 - (N_X^{\text{eq}})^2]$$

Since the interesting dynamics takes place at $T \sim M_X$, define $x \equiv M_X/T$, so that:

$$\frac{dx}{dt} = \frac{d}{dt} \frac{M_X}{T} = -\frac{1}{T} \frac{dT}{dt} x \simeq Hx$$

where $T \propto a^{-1}$ was used (assuming $g_{*S} \simeq \text{const.} \equiv g_{*S}(M_X)$) for the times relevant to freeze-out. Assuming radiation domination⁴, i.e. $H = H(M_X)/x^2$, the Boltzmann equation reduces to the so-called **Riccati equation**:

$$\frac{dN_X}{dx} = -\frac{\lambda}{x^2} [N_X^2 - (N_X^{\text{eq}})^2] \quad (9.3.4)$$

with:

$$\lambda \equiv \frac{2\pi^2}{45} g_{*S} \frac{M_X^3 \langle \sigma v \rangle}{H(M_X)} \quad (9.3.5)$$

This parameter can be treated as a constant in most WIMP theories, but, despite this, the Riccati equation has no analytic solutions. Numerical solutions for two different values of λ are pictured in Fig. 9.4: at high temperatures ($x < 1$) $N_X \approx N_X^{\text{eq}} \simeq 1$, while at low temperatures ($x \gg 1$) the equilibrium abundance is exponentially suppressed as $N_X^{\text{eq}} \sim e^{-x}$. Ultimately,

⁴From Eq. 9.2.24 (with $g_* \simeq \text{const.}$) $H/T^2 = \text{const.}$, hence $H/T^2 = H(M_X)/M_X^2$, i.e. $H = H(M_X)/x^2$.

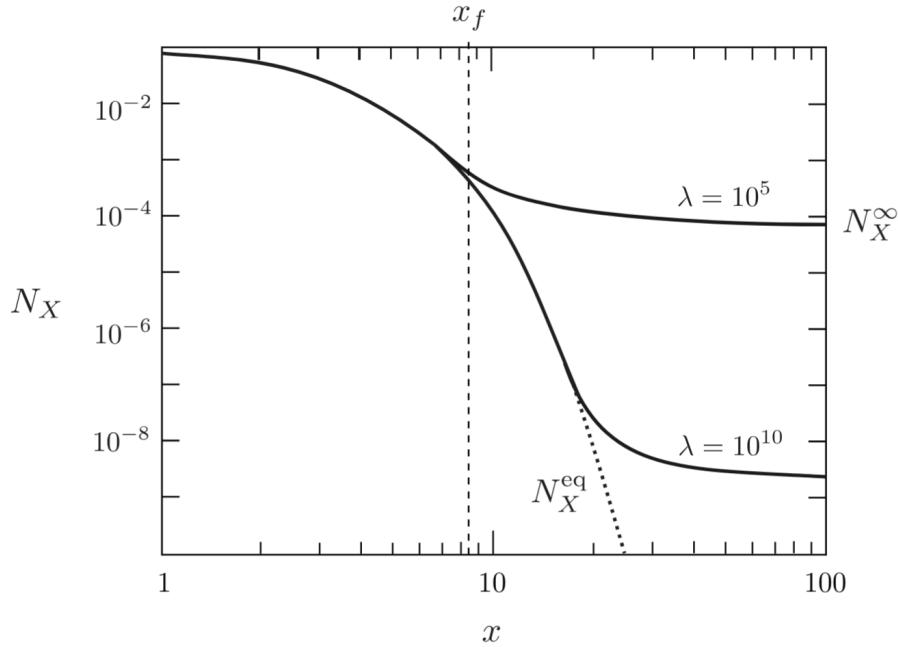


Figure 9.4: Abundance of WIMPs as the temperature drops below their mass.

WIMPs will become so rare that they will not be able to find each other fast enough to maintain the equilibrium abundance: numerically, this freeze-out happens at $x_f \sim 10$, where the solution N_X to the Riccati equation starts to deviate significantly from the Boltzmann-suppressed equilibrium abundance N_X^{eq} .

The final relic abundance $N_X^\infty \equiv N_X(x = \infty)$ determines the freeze-out density of dark matter, and it can be estimated as a function of λ . For $x \gg x_f$ (after freeze-out), N_X will be much larger than N_X^{eq} , hence the Riccati equation reduces to:

$$\frac{dN_X}{dx} \simeq -\frac{\lambda}{x^2} N_X^2$$

Integrating over the domain (x_f, ∞) , the following solution is found:

$$\frac{1}{N_X^\infty} - \frac{1}{N_X^f} = \frac{\lambda}{x_f}$$

with $N_X^f \equiv N_X(x_f)$. Typically $N_X^f \gg N_X^\infty$ (see Fig. 9.4), hence an estimate for the relic abundance is:

$$N_X^\infty \simeq \frac{x_f}{\lambda} \quad (9.3.6)$$

This relation predicts that the relic abundance decreases as the interaction rate λ increases: this makes sense, since larger interaction rates maintain equilibrium longer, i.e. deeper in the Boltzmann-suppressed regime.

Note that the value of $x_f \sim 10$ is not terribly sensitive to the precise value of λ : indeed, $x_f \propto |\log \lambda|$.

§9.3.2.2 WIMP miracle

The freeze-out abundance of dark matter relics can be related to the dark matter density today:

$$\Omega_X \equiv \frac{\rho_{X,0}}{\rho_{crit,0}} = \frac{M_X n_{X,0}}{3M_p^2 H_0^2} = \frac{M_X N_{X,0} s_0}{3M_p^2 H_0^2} = M_X N_X^\infty \frac{s_0}{3M_p^2 H_0^2}$$

as the number of WIMPs is conserved after freeze-out: $N_{X,0} = N_X^\infty$. Inserting Eq. 9.2.19, 9.3.6:

$$\Omega_X \simeq \frac{M_X x_f}{\lambda} \frac{2\pi^2}{45} \frac{g_{*S}(T_0) T_0^3}{3M_p^2 H_0^2} = \frac{H(M_X)}{M_X^2} \frac{x_f}{\langle \sigma v \rangle} \frac{g_{*S}(T_0)}{g_{*S}(M_X)} \frac{T_0^3}{3M_p^2 H_0^2}$$

Using Eq. 9.2.24 for $H(M_X)$ gives:

$$\Omega_X \simeq \frac{\pi}{9} \frac{x_f}{\langle \sigma v \rangle} \sqrt{\frac{g_*(M_X)}{10}} \frac{g_{*S}(T_0)}{g_{*S}(M_X)} \frac{T_0^3}{M_p^3 H_0^2} \quad (9.3.7)$$

Substituting the measured value $T_0 = 2.73$ K and using $g_{*S}(T_0) = 3.91$ and $g_{*S}(M_X) \simeq g_*(M_X)$:

$$\Omega_X h^2 \simeq \frac{x_f}{100} \sqrt{\frac{10}{g_*(M_X)}} \frac{10^{-8} \text{ GeV}^{-2}}{\langle \sigma v \rangle}$$

This reproduces the observed dark matter density if $\sqrt{\langle \sigma v \rangle} \sim 10^{-4} \text{ GeV}^{-1} \sim 0.1 \sqrt{G_F}$: the fact that a thermal relic with a cross-section typical of the weak interaction gives the right dark matter abundance is known as the *WIMP miracle*.

§9.3.3 Big Bang nucleosynthesis

At $T \sim 1$ MeV, photons, electrons and positrons are in equilibrium, neutrinos are about to decouple and baryons are non-relativistic, hence much fewer in number than relativistic particles. However, due to baryon number conservation, the total number of nucleons remains constant: weak nuclear reactions may convert neutrons and protons into each other, while strong nuclear reactions may build nuclei from them. This is known as **Big Bang nucleosynthesis** (BBN). In principle, BBN is a very complicated process involving many coupled Boltzmann equations. In practice, it is convenient to make two simplifications:

1. only consider hydrogen and helium: no heavier element is produced at appreciable levels in BBN, hence one only needs to track ${}^1\text{H} \equiv \text{H}$, ${}^2\text{H} \equiv \text{D}$, ${}^3\text{H} \equiv \text{T}$, ${}^3\text{He}$ and ${}^4\text{He}$;
2. only consider protons and neutrons for $T \gtrsim 0.1$ MeV: at these temperatures, only free protons and neutrons exist, while other light nuclei have yet to form.

§9.3.3.1 Equilibrium abundances

To show the validity of these assumptions, recall that $T \gtrsim 0.8$ MeV protons and neutrons are coupled via weak interactions (β -decay and inverse β -decay):



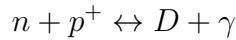
Assume $\mu_\nu, \mu_e \simeq 0$, so that $\mu_p = \mu_n$. Then, generalizing Eq. 9.2.10 to include the chemical potential:

$$n_i^{\text{eq}} = g_i \left(\frac{m_i T}{2\pi} \right) e^{\frac{\mu_i - m_i}{T}} \implies \left(\frac{n_n}{n_p} \right)_{\text{eq}} = \left(\frac{m_n}{m_p} \right)^{3/2} e^{-(m_n - m_p)/T} \simeq e^{-Q/T} \quad (9.3.8)$$

where $Q \equiv m_n - m_p \simeq 1.30$ MeV $\ll m_n, m_p$. This shows that for $T \gg 1$ MeV there are as many neutrons as protons, while for $T < 1$ MeV the neutron fraction gets smaller⁵.

⁵Were the up quark heavier than the down quark, then $m_n < m_p$ and, with the decrease of temperature, protons would weakly decay into neutrons, producing a much different universe.

Next, consider deuterium, which is produced by the following relation:



Since $\mu_\gamma = 0$, then $\mu_D = \mu_n + \mu_p$, recalling that $g_D = 3$ and $g_n = g_p = 2$:

$$\left(\frac{n_D}{n_n n_p}\right)_{\text{eq}} = \left(\frac{m_D}{m_n m_p} \frac{2\pi}{T}\right)^{3/2} e^{-(m_D - m_n - m_p)/T} \quad (9.3.9)$$

In the prefactor $m_D \approx 2m_n \approx 2m_p \simeq 1.9 \text{ GeV}$, while the exponent can be expressed in terms of the binding energy of deuterium $B_D \equiv m_n + m_p - m_D \simeq 2.22 \text{ MeV}$. Hence, as long as chemical equilibrium holds:

$$\left(\frac{n_D}{n_p}\right)_{\text{eq}} = n_n^{\text{eq}} \left(\frac{4\pi}{m_p T}\right)^{3/2} e^{B_D/T}$$

For an order-of-magnitude estimate, approximate the neutron density with the baryon density, so that:

$$n_n \sim n_b = \eta_b n_\gamma = \eta_b \frac{2\zeta(3)}{\pi^2} T^3 \implies \left(\frac{n_D}{n_p}\right)_{\text{eq}} \sim \eta_b \left(\frac{T}{m_p}\right)^{3/2} e^{B_D/T}$$

The smallness of $\eta_b \sim 10^{-9}$ inhibits deuterium production until the temperature drops well below the binding energy B_D , and the same applies to all other nuclei: at $T \gtrsim 0.1 \text{ MeV}$, then, virtually all baryons are in the form of neutrons and protons. At this time, deuterium and helium are produced, but the reaction rates are too slow to produce any heavier nucleus.

§9.3.3.2 Neutron freeze-out and decay

The primordial ration of neutrons is particularly important for BBN, since essentially all neutrons became incorporated into ${}^4\text{He}$. Defining the neutron fraction $X_n \equiv n_n / (n_n + n_p)$, with Eq. 9.3.8:

$$X_n^{\text{eq}}(T) = \frac{e^{-Q/T}}{1 + e^{-Q/T}} \quad (9.3.10)$$

Neutrons follow this equilibrium abundance until neutrinos decouple⁶ at $T_{\text{dec}} \simeq 0.8 \text{ MeV}$, when weak processes like the β -decay effectively shut off. At this moment $X_n^{\text{eq}}(T_{\text{dec}}) \simeq 0.17 \simeq \frac{1}{6}$, which can be taken as a rough estimate of the final freeze-out abundance, so that $X_n^\infty \sim \frac{1}{6}$ (a more precise estimate with QFT gives $X_n^\infty = 0.15$).

At $T \lesssim 0.2 \text{ MeV}$, i.e. $t \lesssim 100 \text{ sec}$, the finite lifetime of the neutron becomes important, and indeed neutron decay must be accounted for:

$$X_n(t) = X_n^\infty e^{-t/\tau_n} \sim \frac{1}{6} e^{-t/\tau_n} \quad (9.3.11)$$

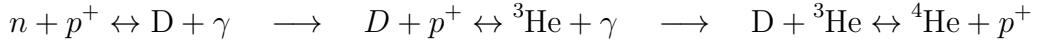
where $\tau_n = 886.7 \pm 0.8 \text{ sec}$ is the neutron's lifetime.

§9.3.3.3 Helium fusion

At this point, the universe is mostly made of protons and neutrons. Helium cannot form directly, since the density is too low and the time available is too short for reactions involving

⁶Note that $Q \sim T_{\text{dec}}$ is just a coincidence, since the former is determined by the strong and electromagnetic interactions, while the latter by the weak interaction.

three or more nuclei to occur (at least at an appreciable rate). Heavier nuclei are instead formed sequentially, according to the following chain of reactions:



Since deuterium is formed directly from neutrons and protons, it can follow its equilibrium abundance as long as there are enough free neutrons available. However, since B_D is rather small, the deuterium abundance becomes large rather late (at $T < 100$ keV), so, although heavier nuclei have larger binding energy and hence would have larger equilibrium abundances, they cannot be formed until sufficient deuterium has become available, a phenomenon known as **deuterium bottleneck**. Only when there is enough deuterium can helium be produced.

To get a rough estimate for the time of nucleosynthesis, first estimate the temperature T_{nuc} when the deuterium fraction in equilibrium would be of order 1, i.e. $(n_D/n_p)_{\text{eq}} \sim 1$:

$$\eta_b \left(\frac{T_{\text{nuc}}}{m_p} \right)^{3/2} e^{B_D/T} \sim 1 \quad \implies \quad T_{\text{nuc}} \sim 0.06 \text{ MeV}$$

From Eq. 9.2.24 with $g_* = 3.38$, the time of nucleosynthesis is found:

$$t_{\text{nuc}} \simeq 120 \text{ sec} \left(\frac{0.1 \text{ MeV}}{T_{\text{nuc}}} \right)^2 \sim 330 \text{ sec}$$

A better estimate would be imposing $(n_D/n_p)_{\text{eq}} \sim 10^{-3}$, as suggested by Fig. 9.6, resulting in $t_{\text{nuc}} \sim 250$ sec: in both cases $t_{\text{nuc}} \ll \tau_n$, so Eq. 9.3.11 is not very sensitive to a precise estimate of t_{nuc} . In both cases:

$$X_n(t_{\text{nuc}}) \sim \frac{1}{8} \quad (9.3.12)$$

Since $B_{\text{He}} > B_D$, the Boltzmann factor $e^{B/T}$ favours helium over deuterium, so helium is produced almost immediately after deuterium: virtually all remaining neutrons at $t \sim t_{\text{nuc}}$ are then processed into ${}^4\text{He}$. Since two neutrons go into one nucleus of ${}^4\text{He}$, its final abundance is equal to half the neutron abundance at t_{nuc} , hence:

$$\frac{n_{\text{He}}}{n_{\text{H}}} = \frac{n_{\text{He}}}{n_p} \simeq \frac{\frac{1}{2} X_n(t_{\text{nuc}})}{1 - X_n(t_{\text{nuc}})} \sim \frac{1}{16} \quad (9.3.13)$$

This prediction is consistent with the observed helium in the universe, as shown in Fig. 9.5. The abundances of other light elements are determined numerically solving the system of coupled Boltzmann equation, and the result is shown in Fig. 9.6 and Fig. 9.7: all these theoretical abundances have quantitative agreement with observational data, with the exception of lithium (the so-called *lithium problem*).

§9.3.3.4 Constraints on BSM physics

The result in Eq. 9.3.13 depends on several input parameters:

- g_* : during radiation era $H \propto g_*^{1/2} T^2$, hence variations in g_* affect the neutron freeze-out temperature $T_f \sim T_{\text{dec}}$, since the neutrino decoupling temperature is determined by $G_F T_{\text{dec}}^5 \sim \sqrt{G_N g_*} T_{\text{dec}}^2$, i.e. $T_f \sim T_{\text{dec}} \propto g_*^{1/6}$. An increase of g_* causes an increase of T_f , which in turn increases the n_n/n_p ratio at freeze-out, hence increasing the final helium abundance;

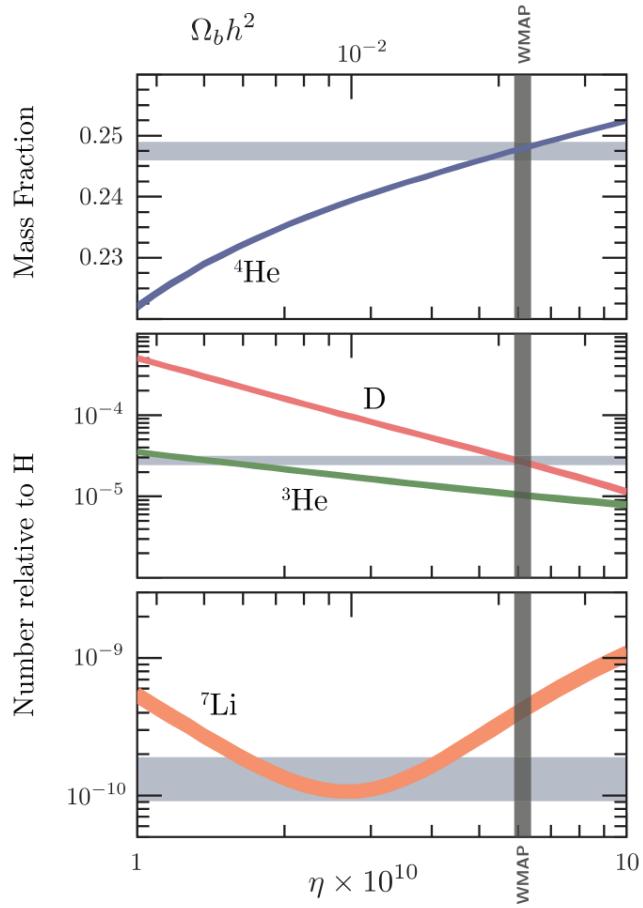


Figure 9.5: Theoretical predictions (coloured) and observational constraints (grey) on today abundances of light nuclei.

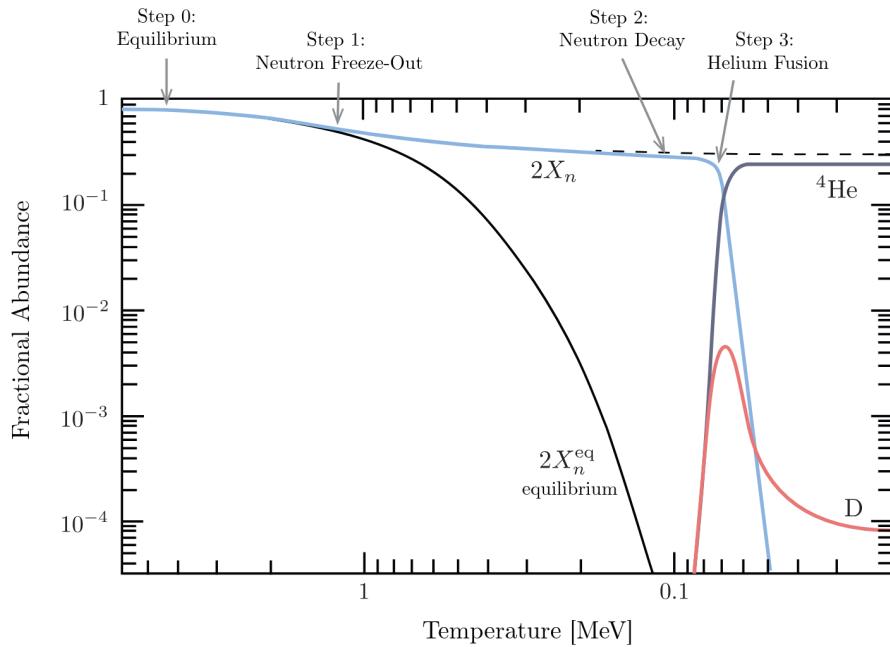


Figure 9.6: Numerical results for helium production in the early universe.

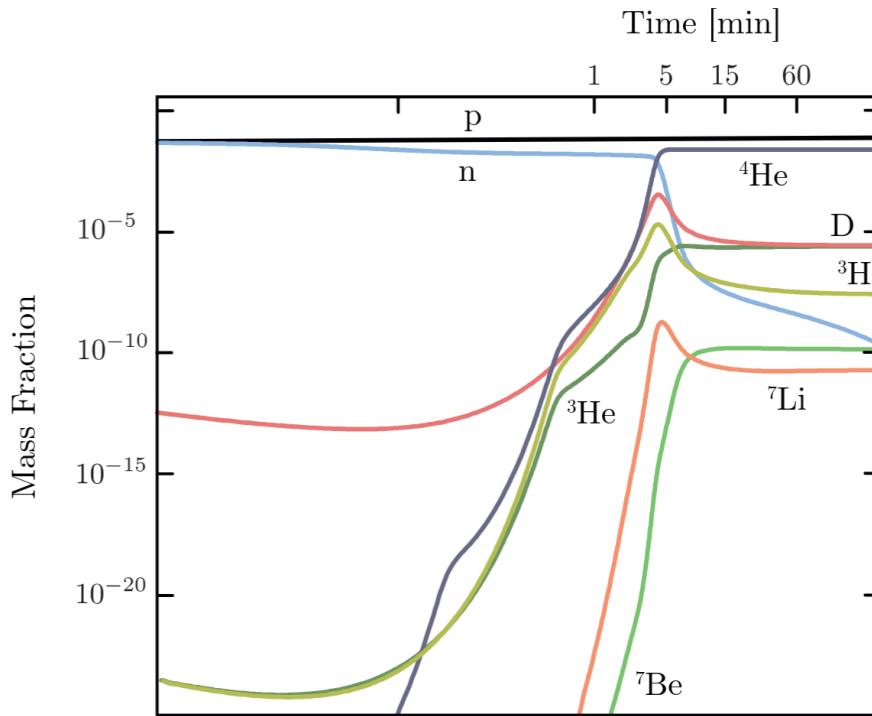


Figure 9.7: Numerical results for the evolution of light element abundances.

- τ_n : a larger neutron lifetime would reduce the amount of neutron decay after freeze-out, thus increasing the final helium abundance;
- Q : a larger mass difference between neutrons and protons would decrease the n_n/n_p ratio at freeze-out, so decreasing the final helium abundance;
- η_b : the final helium abundance increases with an increase of η_b , as nucleosynthesis starts earlier for larger baryon density (i.e. larger deuterium density);
- G_N : increasing the strength of gravity would increase the neutron freeze-out temperature $T_f \sim T_{dec} \propto G_N^{1/6}$, therefore increasing the final helium abundance;
- G_F : increasing the strength of the weak nuclear force would decrease the neutron freeze-out temperature $T_f \sim T_{dec} \propto G_F^{-2/3}$, thus decreasing the final helium abundance.

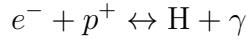
The dependence of BBN on SM parameters makes it a probe (and a constraint) of BSM physics.

§9.3.4 Recombination

At $T \gtrsim 1\text{ eV}$, the universe still consisted of a plasma of free electrons and nuclei: photons were tightly coupled to electrons through Compton scattering, while electrons were strongly coupled to protons via Coulomb scattering. When temperature became low enough, electrons and nuclei combined to form neutral atoms: this phenomenon is known as **Recombination**. As a consequence, the electron density fell sharply, thus causing a growth of the photon mean free path, which became larger than the horizon distance: photons decoupled from matter and the universe became transparent. Today, these photons are the cosmic microwave background.

§9.3.4.1 Saha equilibrium

At $T > 1 \text{ eV}$, baryons and photons are still in equilibrium through electromagnetic reactions, e.g.:



Since $T < m_e < m_p < m_{\text{H}}$, from Eq. 9.2.10 the equilibrium abundances of the various species are:

$$n_i^{\text{eq}} = g_i \left(\frac{m_i T}{2\pi} \right)^{3/2} \exp \frac{\mu_i - m_i}{T} \quad (9.3.14)$$

with $\mu_p + \mu_e = \mu_{\text{H}}$ (since $\mu_{\gamma} = 0$). Then:

$$\left(\frac{n_{\text{H}}}{n_e n_p} \right)_{\text{eq}} = \frac{g_{\text{H}}}{g_e g_p} \left(\frac{m_{\text{H}}}{m_e m_p} \frac{2\pi}{T} \right)^{3/2} \exp \frac{m_p + m_e - m_{\text{H}}}{T}$$

Introducing the binding energy of hydrogen $B_{\text{H}} \equiv m_p + m_e - m_{\text{H}} \simeq 13.6 \text{ eV}$ and the spin factors $g_{\text{H}} = 4$, $g_p = g_e = 2$, and noting that $n_e = n_p$ since the universe is not electrically charged (at least according to current observations), this equation becomes:

$$\left(\frac{n_{\text{H}}}{n_e^2} \right)_{\text{eq}} = \left(\frac{2\pi}{m_e T} \right)^{3/2} e^{B_{\text{H}}/T}$$

Now, define the *free electron-to-baryon fraction* as $X_e \equiv n_e/n_b$, where the baryon number can be written as $n_b = \eta_b(2\zeta(3)/\pi^2)T^3$, and note that protons are over 90% (in number) of all nuclei, so that $n_b \approx n_p + n_{\text{H}} = n_e + n_{\text{H}}$, so that:

$$\frac{1 - X_e}{X_e^2} = \frac{n_{\text{H}}}{n_e^2} n_b \quad (9.3.15)$$

Putting everything together, the **Saha equation** is found:

$$\left(\frac{1 - X_e}{X_e^2} \right)_{\text{eq}} = \frac{2\zeta(3)}{\pi^2} \eta_b \left(\frac{2\pi T}{m_e} \right)^{3/2} e^{B_{\text{H}}/T} \quad (9.3.16)$$

The solution of this equation as a function of redshift is plotted in Fig. 9.8: although the Saha approximation correctly identifies the onset of Recombination, it is clearly insufficient to determine the relic densities of electrons after freeze-out.

§9.3.4.2 Hydrogen recombination

To study Recombination, define the recombination temperature T_{rec} as the temperature at which 90% of electrons have combined with protons to form hydrogen, i.e.:

$$X_e(T_{\text{rec}}) = 0.1 \implies T_{\text{rec}} \approx 0.3 \text{ eV} \simeq 3600 \text{ K} \quad (9.3.17)$$

Note that $T_{\text{rec}} \ll B_{\text{H}}$: the reason is that there are many photons for each hydrogen atom ($\eta_b \sim 10^{-9}$), hence, even when $T < B_{\text{H}}$, the high-energy tail of the photon distribution contains a sufficient number of photons with energy $E > B_{\text{H}}$ which can ionize the hydrogen atoms.

This estimate puts $z_{\text{rec}} \simeq 1320$, since $T_{\text{rec}} = T_0(1 + z_{\text{rec}})$, which puts Recombination in the matter-dominated era which started at $z_{\text{eq}} \simeq 3500$. Then, using $a(t) = (t/t_0)^{2/3}$, the time of Recombination is found:

$$t_{\text{rec}} = \frac{t_0}{(1 + z_{\text{rec}})^{3/2}} \simeq 290'000 \text{ yrs} \quad (9.3.18)$$

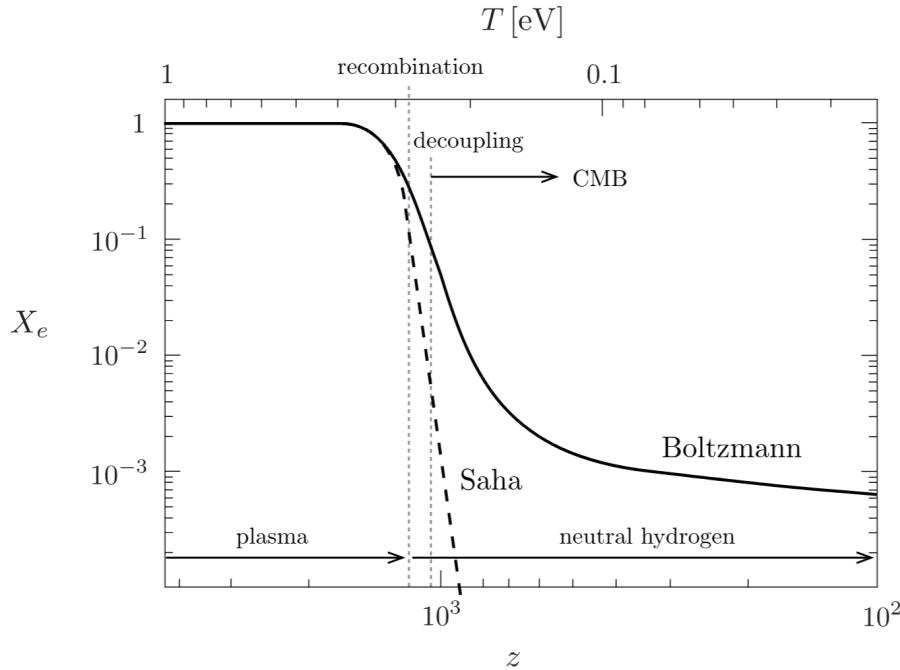
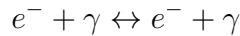


Figure 9.8: Free electron-to-baryon ratio as a function of redshift.

§9.3.4.3 Photon decoupling

At $T = T_{\text{rec}}$ there are still free electrons which are coupled to the photon bath. Indeed, photons are coupled to matter mostly through their interactions with electrons:



with an interaction rate $\Gamma_\gamma \approx n_e \sigma_T$ (as $v \approx c = 1$), where $\sigma_T \simeq 2 \cdot 10^{-3} \text{ MeV}^{-2}$ is the *Thomson cross-section*. The interaction rate therefore decreases as the density of free electrons drops, and the decoupling of photons happens roughly when it becomes smaller than the expansion rate, i.e.:

$$\Gamma_\gamma(T_{\text{CMB}}) \sim H(T_{\text{CMB}}) \quad (9.3.19)$$

Writing $n_e = n_b X_e$, the interaction rate becomes:

$$\Gamma_\gamma(T) = n_b X_e(T) \sigma_T = \frac{2\zeta(3)}{\pi^2} \eta_b \sigma_T X_e(T) T^3$$

On the other hand, in matter era the Ω_m term in Eq. 7.3.16 dominates, so that:

$$H(T) \simeq H_0 \sqrt{\Omega_m} \left(\frac{T}{T_0} \right)^{3/2}$$

Combining this equation, T_{CMB} is found numerically as:

$$X_e(T_{\text{CMB}}) T_{\text{CMB}}^{3/2} \sim \frac{\pi^2}{2\zeta(3)} \frac{H_0 \sqrt{\Omega_m}}{\eta_b \sigma_T T_0^{3/2}} \implies T_{\text{CMB}} \sim 0.27 \text{ eV} \quad (9.3.20)$$

which corresponds to $z_{\text{CMB}} \sim 1100$ and $t_{\text{CMB}} \sim 380 \text{ kyr}$. Note that, although T_{CMB} is not far from T_{rec} , the ionization fraction decreases significantly between recombination and photon decoupling: $X_e(T_{\text{rec}}) \simeq 0.1 \rightarrow X_e(T_{\text{CMB}}) \simeq 0.01$. This shows that a large degree of neutrality is necessary for the universe to become transparent to photon propagation: after decoupling, photons stream freely through the universe, and are observable today as the CMB.

Chapter 10

Structure Formation

Observations of the CMB reveal that, at least until z_{CMB} , the universe was extremely homogeneous, with temperature variations of the order $\delta T/T \sim 10^{-5}$. However, to understand the formation and evolution of large-scale structures, it is necessary to introduce inhomogeneities: today these perturbations of the homogeneous background are huge (e.g. galaxies and voids), but up to z_{CMB} they can be treated with perturbation theory.

Fig. 10.1 shows the evolution of the comoving Hubble radius with the key events of the history of the universe highlighted. Cosmic perturbations (assuming a wake-like ansatz) are pictured: these perturbations evolve in a linear regime up to a_{CMB} , which can be treated analytically, but then enter a non-linear regime which lasts to these days and has to be treated computationally.

§10.1 Newtonian perturbation theory

Since the perturbations are assumed small, the metric perturbation with respect to the background metric can be treated in the Newtonian approximation.

§10.1.1 Flat background

First, consider the evolution of cosmological perturbations in a flat universe. In the Newtonian approximation, the gravitational potential is the:

$$\Phi(t, \mathbf{x}) = \delta\Phi(t, \mathbf{x}) \quad (10.1.1)$$

since a flat background has $\bar{\Phi}(t) \equiv 0$. The matter source too must be perturbed: assuming the background to be at rest, i.e. $\bar{\mathbf{u}}(t) \equiv 0$:

$$\rho(t, \mathbf{x}) = \bar{\rho}(t) + \delta\rho(t, \mathbf{x}) \quad P(t, \mathbf{x}) = \bar{P}(t) + \delta P(t, \mathbf{x}) \quad \mathbf{u}(t, \mathbf{x}) = \delta\mathbf{u}(t, \mathbf{x}) \quad (10.1.2)$$

The equations that dictate the physical evolution of these perturbations are the continuity equation, the Euler equation and the Poisson equation:

$$\dot{\rho} + \nabla \cdot (\rho \mathbf{u}) = 0 \quad (10.1.3)$$

$$(\partial_t + \mathbf{u} \cdot \nabla) \mathbf{u} = -\frac{1}{\rho} \nabla P - \nabla \Phi \quad (10.1.4)$$

$$\Delta\Phi = 4\pi G\rho \quad (10.1.5)$$

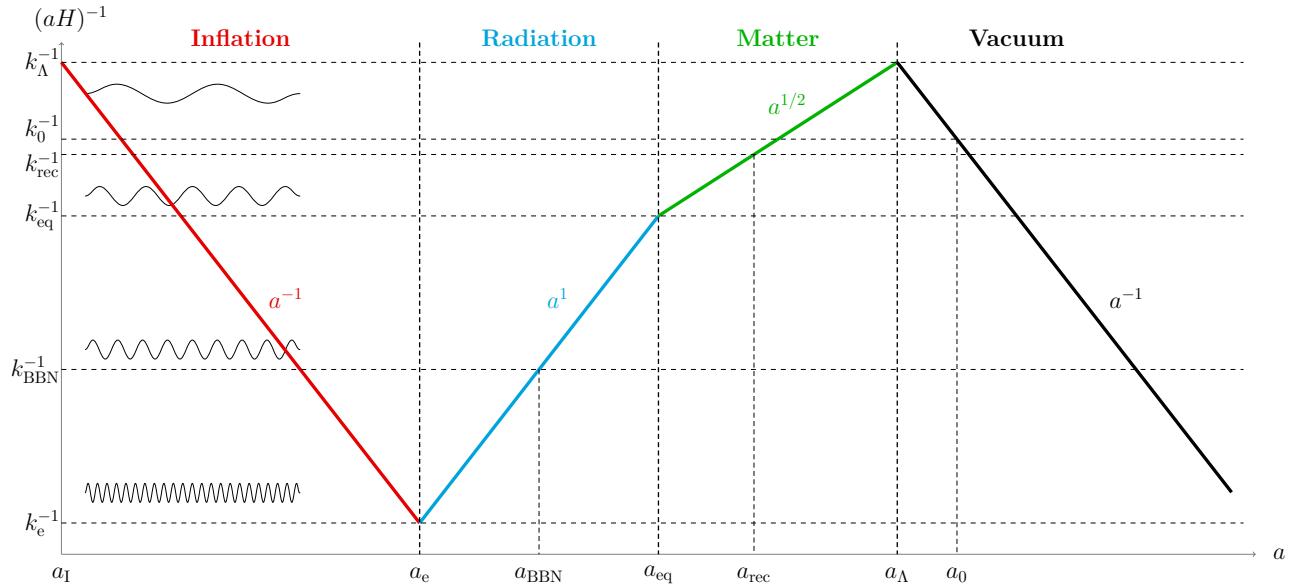


Figure 10.1: Evolution of the comoving Hubble radius. Four examples of cosmological perturbations with different wavelengths are drawn.

§10.1.1.1 Matter perturbations

To study how matter perturbations evolve, consider the complete absence of gravity, i.e. $\Phi(t, \mathbf{x}) \equiv 0$, and a static background $\bar{\rho}, \bar{P} = \text{const.}$, so that the (linear) continuity and Euler equations read:

$$\delta\dot{\rho} + \bar{\rho}\nabla \cdot \delta\mathbf{u} = 0 \quad \bar{\rho}\delta\dot{\mathbf{u}} = -\nabla\delta P$$

Introducing the density contrast $\delta \equiv \delta\rho/\bar{\rho}$, these equations are rewritten as:

$$\dot{\delta} = -\nabla \cdot \delta\mathbf{u} \quad \delta\dot{\mathbf{u}} = -c_s^2\nabla\delta$$

where the **sound velocity** is defined as:

$$c_s^2 \equiv \frac{\delta p}{\delta\rho} \tag{10.1.6}$$

Combining the above equation, a wave equation for the density contrast is found:

$$\ddot{\delta} - c_s^2\Delta\delta = 0 \tag{10.1.7}$$

Therefore, perturbations evolve as sound waves with sound velocity c_s . Since for cosmological matter sources $P = w\rho$, the sound velocity is $c_s^2 = w$: in matter ($w = 0$) no sound waves propagate since there is no pressure. The resulting dispersion relation is:

$$\omega^2(k) = c_s^2k^2 \tag{10.1.8}$$

§10.1.1.2 Gravitational perturbations

Introducing gravitational perturbations in the universe, i.e. $\delta\Phi(t, \mathbf{x}) \neq 0$ (which evolves as $\Delta\delta\Phi = 4\pi G\delta\rho$), the wave equation for the density contrast gains a source term:

$$\ddot{\delta} - c_s^2\Delta\delta = 4\pi G\bar{\rho}\delta \tag{10.1.9}$$

The dispersion relation is then modified as:

$$\omega^2(k) = c_s^2 k^2 - 4\pi G \bar{\rho} \quad (10.1.10)$$

The source term then acts as a *tachyonic mass term* (i.e. with $m^2 < 0$), which introduces a characteristic scale of perturbations, the Jeans scale:

$$k_J^2 \equiv \frac{4\pi G \bar{\rho}}{c_s^2} \quad (10.1.11)$$

For $k > k_J$, i.e. for small-scale perturbations, $\omega \in \mathbb{R}$ and the behavior of the perturbations is oscillating (sound waves), while for $k < k_J$, i.e. large-scale perturbations, $\omega \in \mathbb{C}$ and the perturbations experience an exponential evolution, which is known as **Jeans' instability**: this is the expected collapse of large structures.

After the decoupling of photons (i.e. the emission of the CMB), matter starts behaving with $w = 0$, i.e. $c_s = 0$: since there are no waves, $\omega \in \mathbb{C} \forall k \in \mathbb{R}_0^+$ (from Eq. 10.1.10), i.e. there is a collapse and consequent formation of structures at all scales (stars, galaxies, galaxy clusters, ...).

§10.1.2 FLRW background

Now, consider the evolution of cosmological perturbations in a FLRW background, which introduces the distinction between comoving coordinates \mathbf{x} and physical coordinates $\mathbf{r} := a(t)\mathbf{x}$. As a consequence, the physical velocity is $\mathbf{u} \equiv \dot{\mathbf{r}} = H\mathbf{r} + \mathbf{v}$, where $\mathbf{v} \equiv a(t)\dot{\mathbf{x}}$ is the physical peculiar velocity.

Lemma 10.1.1 (Derivatives in a FLRW background)

$$\nabla_r = a^{-1}(t)\nabla_x \quad \partial_t|_r = \partial_t|_x - H\mathbf{x} \cdot \nabla_x \quad (10.1.12)$$

Proof. Consider a function $f(t, \mathbf{r}) = f(t, a(t)\mathbf{x})$. Then, $\nabla_r = a^{-1}(t)\nabla_x$ by definition, and:

$$\left(\frac{\partial f}{\partial t} \right)_x = \frac{\partial}{\partial t} \Big|_x f(t, a(t)\mathbf{x}) = \frac{\partial}{\partial t} \Big|_r f(t, \mathbf{r}) + \frac{\partial(a(t)\mathbf{x})}{\partial t} \cdot \nabla_r f(t, \mathbf{r}) = \left[\frac{\partial}{\partial t} \Big|_r + H\mathbf{x} \cdot \nabla_x \right] f$$

which is the thesis. \square

This shows that, while in a static spacetime ∂_t and ∇ are independent, in an expanding spacetime this is no longer true. Then, assuming that the quantities in Eq. 10.1.3-10.1.5 are physical quantities, they are modified when expressed in terms of comoving coordinates.

Theorem 10.1.1 (Newtonian approximation in FLRW background)

Cosmological perturbations in a FLRW background are governed by:

$$\dot{\rho} + 3H\rho + \frac{1}{a}\nabla \cdot (\rho\mathbf{v}) = 0 \quad (10.1.13)$$

$$\left(\frac{\partial}{\partial t} + \frac{\mathbf{v}}{a} \cdot \nabla \right) \mathbf{u} = -\frac{1}{a\rho}\nabla P - \frac{1}{a}\nabla\Phi \quad (10.1.14)$$

$$\Delta\Phi = 4\pi G a^2 \rho \quad (10.1.15)$$

where $\nabla \equiv \nabla_x$ and $\partial_t \equiv \partial_t|_x$.

Proof. First, consider the continuity equation Eq. 10.1.3:

$$\begin{aligned} 0 &= \left(\frac{\partial \rho}{\partial t} \right)_r + \nabla_r \cdot (\rho \mathbf{u}) = \left(\frac{\partial}{\partial t} \Big|_x - H \mathbf{x} \cdot \nabla_x \right) \rho + \frac{1}{a} \nabla_x \cdot (\rho \dot{\mathbf{x}} + \rho \mathbf{v}) \\ &\equiv \dot{\rho} - H \mathbf{x} \cdot \nabla \rho + H \mathbf{x} \cdot \nabla \rho + H \rho \nabla \cdot \mathbf{x} + \frac{1}{a} \nabla \cdot (\rho \mathbf{v}) = \dot{\rho} + 3H\rho + \frac{1}{a} \nabla \cdot (\rho \mathbf{v}) \end{aligned}$$

Then, consider the Euler equation Eq. 10.1.4:

$$\begin{aligned} \left(\frac{\partial}{\partial t} \Big|_r + \mathbf{u} \cdot \nabla_r \right) \mathbf{u} &= -\frac{1}{\rho} \nabla_r P - \nabla_r \Phi \\ \left(\frac{\partial}{\partial t} - H \mathbf{x} \cdot \nabla + H \mathbf{x} \cdot \nabla + \frac{\mathbf{v}}{a} \cdot \nabla \right) \mathbf{u} &= -\frac{1}{a\rho} \nabla P - \frac{1}{a} \nabla \Phi \end{aligned}$$

which is the thesis. Finally, the Poisson equation is trivially modified by $\Delta_r = a^{-2}(t)\Delta$. \square

Introducing the density contrast $\delta \ll 1$, these equations can be linearized¹. The continuity equation at zeroth order and at first order reads:

$$\dot{\bar{\rho}} + 3H\bar{\rho} = 0 \quad \frac{\partial(\bar{\rho}\delta)}{\partial t} + 3H\bar{\rho}\delta + \frac{1}{a} \nabla \cdot (\bar{\rho}\mathbf{v}) = 0$$

Using the zeroth order equation for the evolution of the background density, the evolution of the density contrast is found to be:

$$\dot{\delta} = -\frac{1}{a} \nabla \cdot \mathbf{v} \tag{10.1.16}$$

At linear order, the growth of density perturbations is sourced by the divergence of the fluid velocity. Then, the linearized Euler equation reads:

$$\frac{\partial(H\mathbf{r})}{\partial t} = -\frac{1}{a\bar{\rho}} \nabla \bar{P} - \frac{1}{a} \nabla \bar{\Phi} \quad \frac{\partial(H\mathbf{r} + \mathbf{v})}{\partial t} + \frac{\mathbf{v}}{a} H \cdot \nabla \mathbf{r} = -\frac{1}{a\bar{\rho}} \nabla(\bar{P} + \delta P) - \frac{1}{a} \nabla(\bar{\Phi} + \delta \Pi)$$

which, using $\mathbf{v} \cdot \nabla \mathbf{x} = \mathbf{v}$, reduces to:

$$\dot{\mathbf{v}} + H\mathbf{v} = -\frac{1}{a\bar{\rho}} \nabla \delta P - \frac{1}{a} \nabla \delta \Phi \tag{10.1.17}$$

In the absence of perturbations, this equation simply states that $\mathbf{v} \propto a^{-1}(t)$. Now, combine Eq. 10.1.16-10.1.17 and substitute the Poisson equation $\Delta \delta \Phi = 4\pi G a^2 \bar{\rho} \delta$ to obtain:

$$\ddot{\delta} + 2H\dot{\delta} - \frac{c_s^2}{a^2} \Delta \delta = 4\pi G \bar{\rho} \delta \tag{10.1.18}$$

This equation is equivalent to Eq. 10.1.9, but with two important differences caused by the Hubble flow: a redshift of the sound velocity $a^{-1}(t)c_s$ and the presence of friction term, which makes this a damped wave equation.

Note that the Jeans scale is still present, although now it is affected by the redshift due to the Hubble flow: if $k > ak_J$, the perturbations behave like damped waves, while for $k < ak_J$ the system presents Jeans' instability, which is no longer exponential due to the friction term (now it is a power-law instability).

¹Recall that, for the velocity field, the perturbation is the peculiar velocity \mathbf{v} .

For baryons, the sound velocity, and so the Jeans scale, depend on time. Before recombination, baryons are strongly coupled to photons, forming a relativistic photon-baryon fluid with $c_s \approx 1/\sqrt{3}$: the Jeans scale is then of the order of the comoving Hubble radius, suppressing the growth of baryonic fluctuations at subhorizon scales (oscillatory regime). After recombination, the baryonic fluid starts behaving like pressureless dust and baryonic fluctuations at all scales start to grow.

On the other hand, for cold dark matter (CDM) the sound speed is negligible at all times, so that subhorizon fluctuations can start to grow earlier.

§10.1.2.1 Matter era

In matter era $a \propto t^{2/3}$ and $H(t) = \frac{2}{3t}$. The density contrast is just $\delta \simeq \delta_m \equiv \delta\rho_m/\bar{\rho}_m$, since:

$$\delta = \frac{\delta\rho}{\bar{\rho}} \simeq \frac{\delta\rho_r + \delta\rho_m}{\bar{\rho}_m} \equiv \delta_m + \delta_r \frac{\bar{\rho}_r}{\bar{\rho}_m} \simeq \delta_m$$

By Eq. 7.3.12, then, $4\pi G\bar{\rho}_m = \frac{3}{2}H^2 = \frac{2}{3t^2}$, and Eq. 10.1.18 for subhorizon modes (i.e. $\lambda < (aH)^{-1}$) becomes:

$$\ddot{\delta}_m + \frac{4}{3t^2}\dot{\delta}_m - \frac{2}{3t^2}\delta_m = 0 \quad (10.1.19)$$

since $c_s^2 \approx w = 0$ in matter era. With a power-law ansatz $\delta_m \propto t^\alpha$, two solutions are found:

$$\delta_m \propto \begin{cases} t^{-1} \propto a^{-2/3}t^{2/3} \propto a \\ t^1 \end{cases} \quad (10.1.20)$$

The first solution is a *decaying solution*, while the second one is a *growing solution*. The growing solution is $\delta \propto a$, which during matter era varies from $a_{eq} \simeq 1/3401$ to $a_\Lambda \simeq 1/1.4$, i.e. by $\sim 10^3$ times: assuming initial perturbations of order $\delta_m \sim 10^{-5}$ (from CMB), at the end of matter era these perturbations have grown to $\delta_m \sim 10^{-2}$, still not enough to explain the cosmic structures observed today. This is due to the fact that, after the CMB emission, cosmological perturbations enter the non-linear regime, and non-linear collapse is faster than the linear one. Moreover, note that a smaller matter era determines a smaller growth of fluctuations: this means that the existence of dark matter (i.e. an additional dominant contribution to Ω_m other than baryonic matter) is necessary to explain the current cosmic structures.

§10.1.2.2 Radiation era

In radiation era $a \propto t^{1/2}$ and $H(t) = \frac{1}{2t}$. Now, the density contrast is $\delta \simeq \delta_r + \delta_m \frac{\bar{\rho}_m}{\bar{\rho}_r}$: since the radiation fluid has a large sound velocity $c_s \approx 1/\sqrt{3}$, it is expected to oscillate over scales smaller than the comoving Hubble horizon, so that, averaging over many oscillation periods, $\langle \delta_r \rangle = 0$ and the time-average density contrast reduces to the matter contribution. Then, Eq. 10.1.18 becomes:

$$\ddot{\delta}_m + 2H\dot{\delta}_m - 4\pi G\bar{\rho}_m\delta_m = 0$$

Since $\dot{\delta}_m \sim H\delta_m$, the first two terms are $\sim H^2 \sim \bar{\rho}_r \gg \bar{\rho}_m$, hence the last term is negligible and the evolution of δ_m is given by:

$$\ddot{\delta}_m + \frac{1}{t}\dot{\delta}_m = 0 \quad (10.1.21)$$

With a power-law ansatz, two solution are found:

$$\delta_m \propto \begin{cases} \text{const.} \\ \ln t \propto \ln a \end{cases} \quad (10.1.22)$$

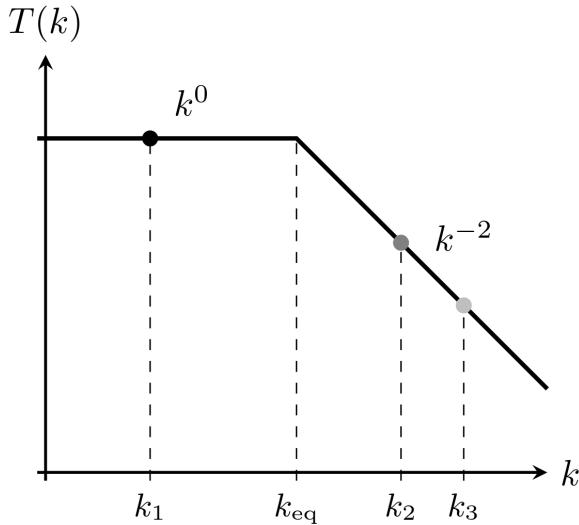


Figure 10.2: Transfer function for matter perturbation: the shape of $\mathcal{T}(k)$ is a consequence of the suppressed growth of subhorizon modes during the radiation era.

Therefore, matter perturbations grow slower in radiation era, compared to matter era (*Mészáros effect*). On the other hand, superhorizon modes (i.e. $\lambda > (aH)^{-1}$) cannot causally interact, and their evolution can be studied with a relativistic analysis.

To show how matter perturbations at different scales evolve, it is necessary to study their power spectrum $\delta_m(k)$. In particular, Inflation predicts that, when perturbations enter the comoving Hubble horizon, they have the primordial power spectrum $\delta_{m,\text{initial}}(k) \propto k^{-1}$ (*Harrison-Zel'dovich spectrum*), and then they evolve according to a transfer function of the form (see Fig. 10.2):

$$\mathcal{T}(k) \simeq \begin{cases} 1 & k < k_{\text{eq}} \\ \left(\frac{k_{\text{eq}}}{k}\right)^2 \ln \frac{k}{k_{\text{eq}}} & k > k_{\text{eq}} \end{cases} \quad (10.1.23)$$

where $k_{\text{eq}} \sim 60$ is the wavenumber of perturbations entering the comoving Hubble horizon at radiation-matter equality ($\lambda_{\text{eq}} \sim 100$ Mpc, approximately the size of a galaxy cluster). This means that large-scale perturbations grow less than small-scale perturbations, since they enter the Hubble radius later (see Fig. 10.1).

Combining these, it is possible to compute the predicted (linear) matter power spectrum for matter fluctuations observed today:

$$\delta_m^2 \propto \begin{cases} k & k < k_{\text{eq}} \\ k^{-3} \ln^2 \frac{k}{k_{\text{eq}}} & k > k_{\text{eq}} \end{cases} \quad (10.1.24)$$

which has excellent agreement with experimental observations (Fig. 10.3), except for the high- k (i.e. small- λ) regime, a discrepancy which is due to the limited applicability of the linear approximation to small-scale fluctuations. Indeed, the linear analysis tends to underestimate the formation of small-scale structure fuelled by the faster non-linear collapse.

§10.1.2.3 Vacuum era

Finally, in vacuum era $a \propto e^{Ht}$, with $H \approx \text{const.}$ during the dark-matter-dominated era. Again $4\pi G \bar{\rho}_m \ll H^2 \propto \rho_\Lambda$ (the density of dark energy does not fluctuate), hence Eq. 10.1.18 reduces

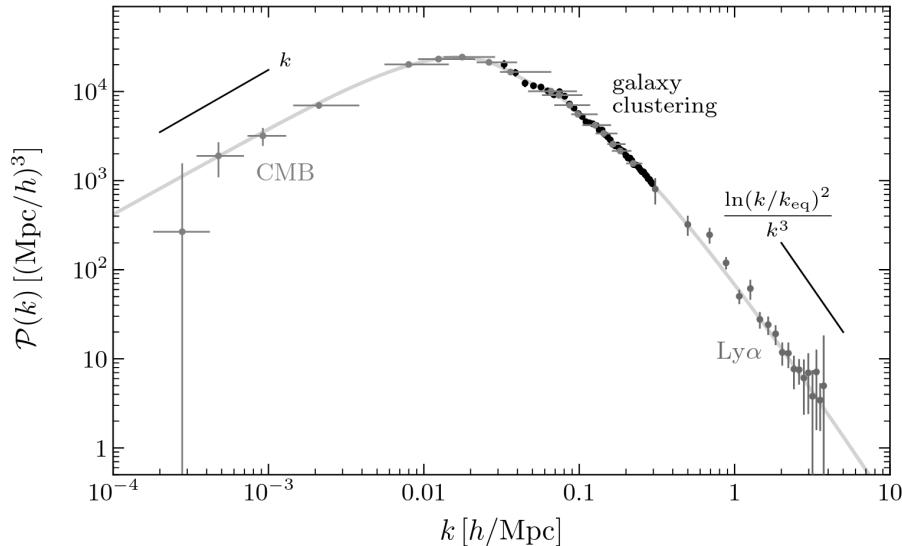


Figure 10.3: Measurements of the linear matter power spectrum.

to:

$$\ddot{\delta}_m + 2H\dot{\delta}_m = 0 \quad (10.1.25)$$

which, with a power-law ansatz, has two solutions:

$$\delta_m \propto \begin{cases} e^{-2Ht} \propto a^{-2} \\ \text{const.} \end{cases} \quad (10.1.26)$$

which shows that, during the vacuum era, the growth of cosmological perturbations (i.e. of structures) halts.

§10.2 Relativistic perturbation theory

The Newtonian analysis is inadequate on scales larger than the Hubble radius and for relativistic fluids (e.g. photons and neutrinos). Before photon decoupling, photons electrons and protons are strongly coupled by electromagnetic interactions, forming a single (relativistic) photon-baryon fluid: the matter perturbations then oscillate, since gravity makes perturbations collapse, causing an increase in radiation density that makes them re-expand. At photon decoupling, modes with different wavelengths are captured at different phases in their evolution, giving rise to a characteristic pattern in the CMB. After photon decoupling, the baryons lose the pressure support of the photons and fall into the gravitational potential wells created by dark matter, eventually forming stars and galaxies. These dark matter clusters only start to form after radiation-matter equality, since before that they are suppressed by the large radiation pressure of photons.

§10.2.1 Perturbations

The Einstein field equations Eq. 4.5.6 couple perturbations in the stress-energy tensor to those in the metric, which then need to be studied simultaneously. In particular, working in conformal time, the perturbations with respect to the homogeneous background can be written as:

$$g_{\mu\nu}(\eta, \mathbf{x}) = \bar{g}_{\mu\nu}(\eta) + \delta g_{\mu\nu}(\eta, \mathbf{x}) \quad (10.2.1)$$

$$T_{\mu\nu}(\eta, \mathbf{x}) = \bar{T}_{\mu\nu}(\eta) + \delta T_{\mu\nu}(\eta, \mathbf{x}) \quad (10.2.2)$$

To avoid clutter, from now on the dependence of perturbations from (η, \mathbf{x}) is suppressed. Note that the unperturbed metric is assumed to be the flat FLRW metric, so that:

$$ds^2 \equiv \bar{g}_{\mu\nu} dx^\mu dx^\nu = a^2(\eta) [d\eta^2 - d\mathbf{x}^2]$$

§10.2.1.1 Metric perturbations

Since the metric tensor of a 4D manifold has 10 independent components, $\delta g_{\mu\nu}$ can be written as:

$$ds^2 \equiv g_{\mu\nu} dx^\mu dx^\nu = a^2(\eta) [(1 + 2A) d\eta^2 - 2B_i dx^i d\eta - (\delta_{ij} + h_{ij}) dx^i dx^j] \quad (10.2.3)$$

where A is a scalar, B_i is a spatial 3-vector and h_{ij} is a symmetric spatial 3-tensor. Then, by convention $B^i = \delta^{ij} B_j$ and $h^i_j = \delta^{ik} h_{kj}$.

Lemma 10.2.1 (SVT decomposition of vectors)

Given a 3-vector B_i , it can be decomposed as:

$$B_i = \partial_i B + \hat{B}_i \quad (10.2.4)$$

where B is a scalar and $\hat{B}_i : \partial^i \hat{B}_i = 0$ is a divergenceless 3-vector.

Lemma 10.2.2 (SVT decomposition of symmetric tensors)

Given a symmetric 3-tensor h_{ij} , it can be decomposed as:

$$h_{ij} = 2C\delta_{ij} + 2\partial_{(i}\partial_{j)}E + 2\partial_{(i}\hat{E}_{j)} + 2\hat{E}_{ij} \quad (10.2.5)$$

where C, E are scalars, $\hat{E}_i : \partial^i \hat{E}_i = 0$ is a divergenceless 3-vector, $\hat{E}_{ij} : \partial^i \hat{E}_{ij} = 0 \wedge \hat{E}_i^i = 0$ is a divergenceless traceless symmetric 3-tensor and:

$$\partial_{(i} \partial_{j)} := \partial_i \partial_j - \frac{1}{3} \delta_{ij} \Delta \quad (10.2.6)$$

With the SVT decomposition, the 3 degrees of freedom of B_i are decomposed into 1 + 2 and the 6 of h_{ij} into 1 + 1 + 2 + 2, resulting in a complete decomposition of the 10 degrees of freedom of $\delta g_{\mu\nu}$ into 4 + 4 + 2: scalars A, B, C, E , vectors \hat{B}_i, \hat{E}_i and tensor \hat{E}_{ij} .

The SVT decomposition is convenient since, at linear order, the Einstein equations for scalars, vectors and tensors do not mix and can therefore be treated separately. In particular, in the following vector perturbations are ignored, since they are not produced by Inflation, and even if they were, they would decay quickly with the expansion of the universe. On the other hand, \hat{E}_{ij} is a transverse traceless tensor, hence tensor perturbations are truly gravitational waves (Eq. 5.2.5) and will be analyzed later.

Gauge fixing The remaining scalar perturbations can be further simplified with gauge invariance. Indeed, the perturbations in Eq. 10.2.3 are not uniquely defined, but depend on the choice of coordinates, i.e. on the choice of gauge: the chosen form of the perturbed metric has been determined by the implicit introduction of a specific time slicing of the spacetime and definition of specific coordinates on these time slices. In general, choosing different coordinates introduces “fictitious perturbations”, known as *gauge modes*.

In order to avoid these gauge problems, there are two solutions: either define special combinations of the metric perturbations that do not transform under a change of coordinates, called *Bardeen variables*, or fix a particular gauge. In the following, the **Newtonian gauge** is used, where two of the four scalar perturbations are set to zero:

$$B = E \equiv 0 \quad (10.2.7)$$

The perturbed metric then becomes:

$$ds^2 = a^2(\eta) [(1 + 2\Psi) d\eta^2 - (1 - 2\Phi) \delta_{ij} dx^i dx^j] \quad (10.2.8)$$

where the Bardeen potentials $\Psi \equiv A$ and $\Phi \equiv -C$ have been introduced. For perturbations that decay at spatial infinity, the Newtonian gauge is unique (i.e. the gauge is fixed completely). Note that, in this gauge, time slices are orthogonal to the worldlines of observers at rest in these coordinates (since $B = 0$), and the induced geometry on time slices is isotropic (since $E = 0$), thus making the physics transparent.

The metric in the Newtonian gauge bears similarity to the weak-field limit in Minkowski space Eq. 5.1.13: this suggests that Ψ plays the role of a gravitational potential.

§10.2.1.2 Matter perturbations

The perturbations of the stress-energy tensor can be written in term of the perturbed energy density $\rho(\eta, \mathbf{x}) = \bar{\rho}(\eta) + \delta\rho(\eta, \mathbf{x})$ and pressure $P(\eta, \mathbf{x}) = \bar{P}(\eta) + \delta P(\eta, \mathbf{x})$:

$$\delta T_0^0 = \delta\rho \quad (10.2.9)$$

$$\delta T_0^i = (\bar{\rho} + \bar{P}) v^i \quad (10.2.10)$$

$$\delta T_j^i = -\delta P \delta_j^i - \Pi_j^i \quad (10.2.11)$$

where v^i is a 3-vector, known as *bulk velocity*, and Π^i_j is a symmetric traceless 3-tensor, known as *anisotropic stress*. The 10 degrees of freedom of δT^μ_ν are then correctly decomposed as $1 + 1 + 3 + 5$. In the following analysis, the anisotropic stress is ignored: $\Pi^i_j \equiv 0$.

§10.2.2 Evolution

Lemma 10.2.3 (Christoffel symbols in perturbed FLRW)

The Christoffel symbols associated to the metric Eq. 10.2.8 are:

$$\Gamma_{00}^0 = \mathcal{H} + \Psi' \quad (10.2.12)$$

$$\Gamma_{i0}^0 = \partial_i \Psi \quad (10.2.13)$$

$$\Gamma_{00}^i = \delta^{ij} \partial_j \Psi \quad (10.2.14)$$

$$\Gamma_{ij}^0 = \mathcal{H} \delta_{ij} - [\Phi' + 2\mathcal{H}(\Phi + \Psi)] \delta_{ij} \quad (10.2.15)$$

$$\Gamma_{j0}^i = [\mathcal{H} - \Phi'] \delta_j^i \quad (10.2.16)$$

$$\Gamma_{jk}^i = -2\delta_{(j}^i \partial_{k)} \Phi + \delta_{jk} \delta^{i\ell} \partial_\ell \Phi \quad (10.2.17)$$

where $\mathcal{H} \equiv \frac{a'}{a}$ and $a' := \frac{da}{d\eta}$.

The equations of motion of matter perturbations follow from the covariant conservation of the stress-energy tensor: $\nabla_\mu T^\mu_\nu = 0$. The $\nu = 0$ component is the continuity equation, which at zeroth and linear order in perturbations is:

$$\bar{\rho}' + 3\mathcal{H}(\bar{\rho} + \bar{P}) = 0 \quad (10.2.18)$$

$$\delta\rho' + 3\mathcal{H}(\delta\rho + \delta P) - 3\Phi'(\bar{\rho} + \bar{P}) + \nabla \cdot [(\bar{\rho} + \bar{P})\mathbf{v}] = 0 \quad (10.2.19)$$

The first equation is just the continuity equation Eq. 7.3.5 for the background, while the second equation describes the evolution of the density perturbations: in particular, the first term is the universal dilution due to the expansion, the second term is a local dilution due to the “local scale factor” $(1 - \Phi)a$ in the spatial part of the metric, and the third term is the local divergence of the fluid flow. It is also possible to rewrite the linearized continuity equation in terms of the density contrast $\delta \equiv \delta\rho/\bar{\rho}$.

Proposition 10.2.1 (Continuity equation for the density contrast)

$$\delta' = (1 + c_s^2)(\Phi' - \nabla \cdot \mathbf{v}) \quad (10.2.20)$$

Proof. Since the background is homogeneous, i.e. $\bar{\rho} = \bar{\rho}(\eta)$ and $\bar{P} = \bar{P}(\eta)$, Eq. 10.2.19 becomes:

$$\delta\rho' + 3(1 + c_s^2)(\mathcal{H}\delta\rho - \Phi'\bar{\rho}) + (1 + c_s^2)\bar{\rho}\nabla \cdot \mathbf{v} = 0$$

where the sound velocity $c_s^2 \equiv \bar{P}/\bar{\rho} = w$ has been introduced. Then, recall Eq. 10.2.18, so that:

$$\delta' = \frac{\delta\rho'}{\bar{\rho}} - \delta\frac{\bar{\rho}'}{\bar{\rho}} = \frac{\delta\rho'}{\bar{\rho}} + 3(1 + c_s^2)\mathcal{H}\delta$$

Inserting the above expression for $\delta\rho'$ yields the thesis. \square

The $\nu = i$ component is the Euler equation, which reads (using Eq. 10.2.18):

$$\mathbf{v}' + (1 - 3c_s^2)\mathcal{H}\mathbf{v} = -\frac{1}{\bar{\rho} + \bar{P}}\nabla \cdot \delta P - \nabla \cdot \Psi \quad (10.2.21)$$

where $c_s^2 \equiv \bar{p}'/\bar{\rho}' = w$. Note the equivalence with Eq. 10.1.4 and the fact that, in absence of energy density and pressure, this equation is equivalent to $v \propto a^{-1}$ as expected.

The equations of motion of metric perturbations follow instead from the Einstein field equations: $G^\mu_\nu = 8\pi G T^\mu_\nu$. The (00)-component at zeroth and linear order is:

$$\mathcal{H}^2 = \frac{8\pi G}{3}a^2\bar{\rho} \quad (10.2.22)$$

$$\Delta\Psi - 3\mathcal{H}(\Phi' + \mathcal{H}\Psi) = 4\pi Ga^2\delta\rho \quad (10.2.23)$$

The first equation is just the Friedmann equation Eq. 7.3.12, while the second is the relativistic generalization of a sourced Poisson equation for Ψ , coupled to Φ : for subhorizon modes, i.e. with $k \gg aH = \mathcal{H}$ (since $\partial_\eta = a\partial_t$), it reduces to the Poisson equation in the Newtonian limit $\Delta\Psi = 4\pi Ga^2\delta\rho$. Then, the $(0i)$ -component reads:

$$\partial_i(\Phi' + \mathcal{H}\Phi) = -4\pi Ga^2(\bar{\rho} + \bar{P})v_i$$

Applying a SVT decomposition to $v^i = \partial^i v + \hat{v}^i$ and ignoring vector perturbations (as already stated), this equation reduces to a scalar equation:

$$\Phi' + \mathcal{H}\Phi = -4\pi Ga^2(\bar{\rho} + \bar{P})v \quad (10.2.24)$$

which is again a relativistic sourced Poisson equation for Φ . There only remain the spatial components, which are best studied in terms of their trace and traceless parts. The traceless component ($ij - \frac{1}{3}ii$) simply is:

$$\Phi = \Psi \quad (10.2.25)$$

while the trace component (ii) is:

$$\Phi'' + 3\mathcal{H}\Phi' + (2\mathcal{H}' + \mathcal{H}^2)\Phi = 4\pi Ga^2\delta P \quad (10.2.26)$$

Note that these equations are only valid with $\Pi^i_j \equiv 0$: if the anisotropic stress cannot be ignored, then $\Phi \neq \Psi$ and the last equation has an additional source term.

§10.2.2.1 Metric evolution

The evolution of the potential Φ differs between radiation era and matter era.

Radiation era In radiation era $\mathcal{H} = \eta^{-1}$ (Eq. 7.3.19), so that $2\mathcal{H}' + \mathcal{H}^2 = -\mathcal{H}^2$. Moreover, since $\delta P = \frac{1}{3}\delta\rho$, combining Eq. 10.2.23 and Eq. 10.2.26 results into:

$$\Phi'' + 4\mathcal{H}\Phi' - \frac{1}{3}\Delta\Phi = 0 \quad (10.2.27)$$

which is a damped wave equation with $c_s = 1/\sqrt{3}$ (as expected from $c_s^2 = w$). Now, insert the Fourier expansion of the metric perturbation:

$$\Phi(\eta, \mathbf{x}) = \int_{\mathbb{R}^3} \frac{d^3k}{(2\pi)^{3/2}} \Phi_{\mathbf{k}}(\eta) e^{i\mathbf{k} \cdot \mathbf{x}} \quad (10.2.28)$$

so that the wave equation becomes an ODE:

$$\Phi''_{\mathbf{k}} + \frac{4}{\eta} \Phi'_{\mathbf{k}} + \frac{k^2}{3} \Phi_{\mathbf{k}} = 0$$

The solutions to this equation have a distinct behavior in the superhorizon and subhorizon regimes:

$$\Phi_{\mathbf{k}}(\eta) \approx \begin{cases} \text{const.} & \eta \ll \frac{1}{k} \\ -G\Phi_{\mathbf{k}}(0) \frac{1}{(k\eta)^2} \cos \frac{k\eta}{\sqrt{3}} & \eta \gg \frac{1}{k} \end{cases} \quad (10.2.29)$$

This shows that outside the comoving Hubble horizon no linear evolution happens, and the metric perturbations only evolve with small non-linear terms. On the other hand, inside the comoving Hubble horizon, in radiation era the metric perturbations oscillate with a damped amplitude that decreases as a^{-2} (since $a \propto \eta$).

Matter era In matter era $\mathcal{H} = 2\eta^{-1}$, so that $2\mathcal{H}' + \mathcal{H}^2 = 0$. Moreover, since $w = 0$, Eq. 10.2.26 reduces to:

$$\Phi'' + 3\mathcal{H}\Phi' = 0 \quad (10.2.30)$$

which has two solutions:

$$\Phi(\eta) \propto \begin{cases} \text{const.} \\ \eta^{-5} \propto a^{-5/2} \end{cases} \quad (10.2.31)$$

In matter era, then, the growing solution is constant independently on the wavelength of the fluctuation.

Putting everything together, it is possible to outline the linear evolution of curvature perturbations. Modes with $k > k_{\text{eq}}$, which enter the Hubble radius in radiation era, are affected by the damped oscillations described by Eq. 10.2.29 before freezing in matter era. On the contrary, modes with $k < k_{\text{eq}}$ enter the Hubble radius in matter era, thus freezing immediately (see Fig. 10.4).

§10.2.2.2 Radiation evolution

The evolution of perturbations in the radiation energy density, encapsulated in the radiation density contrast $\delta_r \equiv \delta\rho_r/\bar{\rho}_r$, is of particular interest, since it predicts the shape of the CMB power spectrum (which is still described by the linear approximation): indeed, since $\rho_r \propto T^4$, then $\delta\rho \propto 4T^3\delta T$ and:

$$\delta_r = 4\delta_T \quad (10.2.32)$$

where $\delta_T \equiv \delta T/\bar{T}$ is the temperature contrast, which is measured in CMB anisotropies. To study the evolution of δ_r , combine Eq. 10.2.22-10.2.23:

$$\delta = \frac{2}{3} \frac{\Delta\Phi}{\mathcal{H}^2} - \frac{2\Phi'}{\mathcal{H}} - 2\Phi \quad (10.2.33)$$

After introducing the Fourier expansion Eq. 10.2.28, this equation can be studied both in radiation era and in matter era.

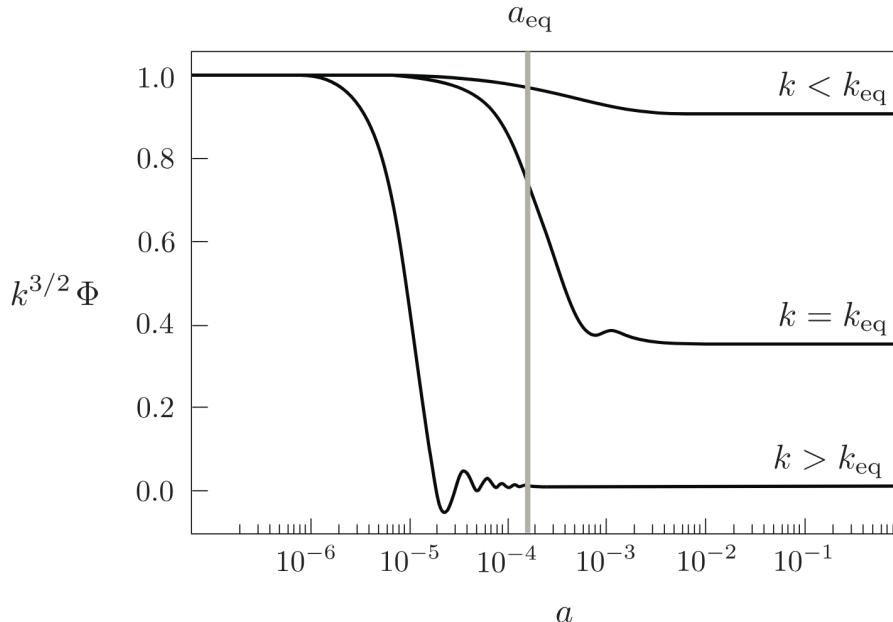


Figure 10.4: Linear evolution of metric perturbations. The small suppression of modes $k > k_{eq}$ is a minor effect due to the radiation-to-matter transition.

Radiation era In radiation era δ_r is dominant, i.e. $\delta \approx \delta_r$, and $\mathcal{H} = \eta^{-1}$, so that:

$$\delta_r = -\frac{2}{3}(k\eta)^2 \Phi_k - 2\eta \Phi'_k - 2\Phi_k$$

Recalling the expression Eq. 10.2.29 for the metric perturbation, the solution for δ_r is found:

$$\delta_r(\eta) \approx \begin{cases} \text{const.} & \eta \ll \frac{1}{k} \\ 4\Phi_k(0) \cos \frac{k\eta}{\sqrt{3}} & \eta \gg \frac{1}{k} \end{cases} \quad (10.2.34)$$

Note that superhorizon modes are still frozen in the linear regime, while subhorizon radiation fluctuations oscillate without damping (contrary to metric perturbations, which are damped).

Matter era In matter era δ_r is subdominant, and its evolution has to be inferred from the continuity and the Euler equations: subhorizon modes still oscillate with the same frequency as in radiation era, however now with a decreased amplitude and a shift due to the gravitational potential of the dominant matter:

$$\delta_r(\eta) \approx \begin{cases} \text{const.} & \eta \ll \frac{1}{k} \\ -4\Phi_m + A \cos \frac{k\eta}{\sqrt{3}} & \eta \gg \frac{1}{k} \end{cases} \quad (10.2.35)$$

As seen in Fig. 10.1, radiation modes with $k > k_{eq}$, i.e. those which enter the comoving Hubble horizon in matter era, do not evolve significantly before photon decoupling, while radiation modes with $k < k_{eq}$ are affected by oscillations, hence arrive at the last-scattering surface with different phases. The power spectrum of the CMB is then expected to have a plateau at low $k < k_{eq}$ (i.e. at high wavelengths): this is experimentally observed and is known as the **Sachs–Wolfe plateau** (see Fig. 10.6).

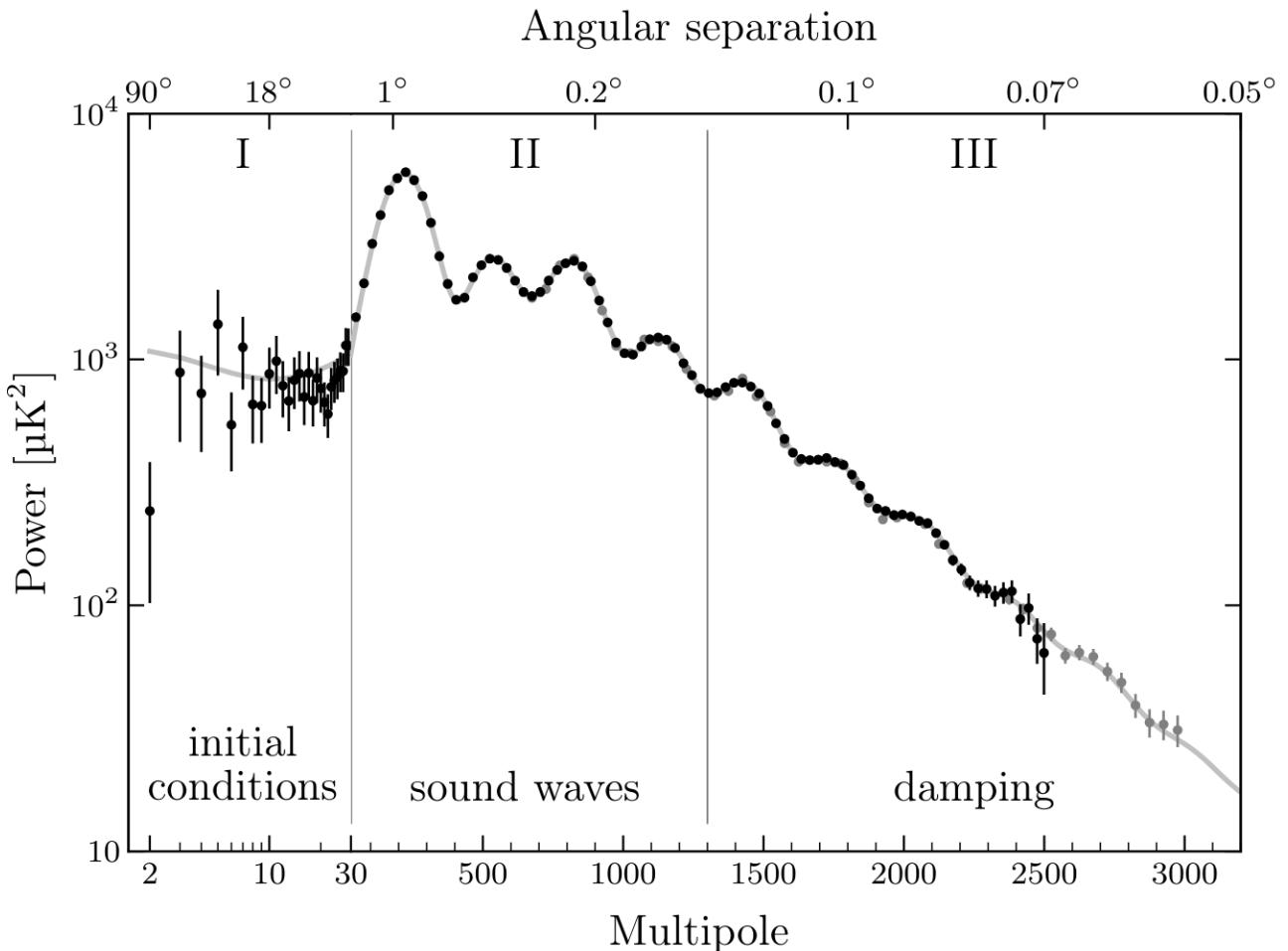


Figure 10.5: Power spectrum of the CMB anisotropies.

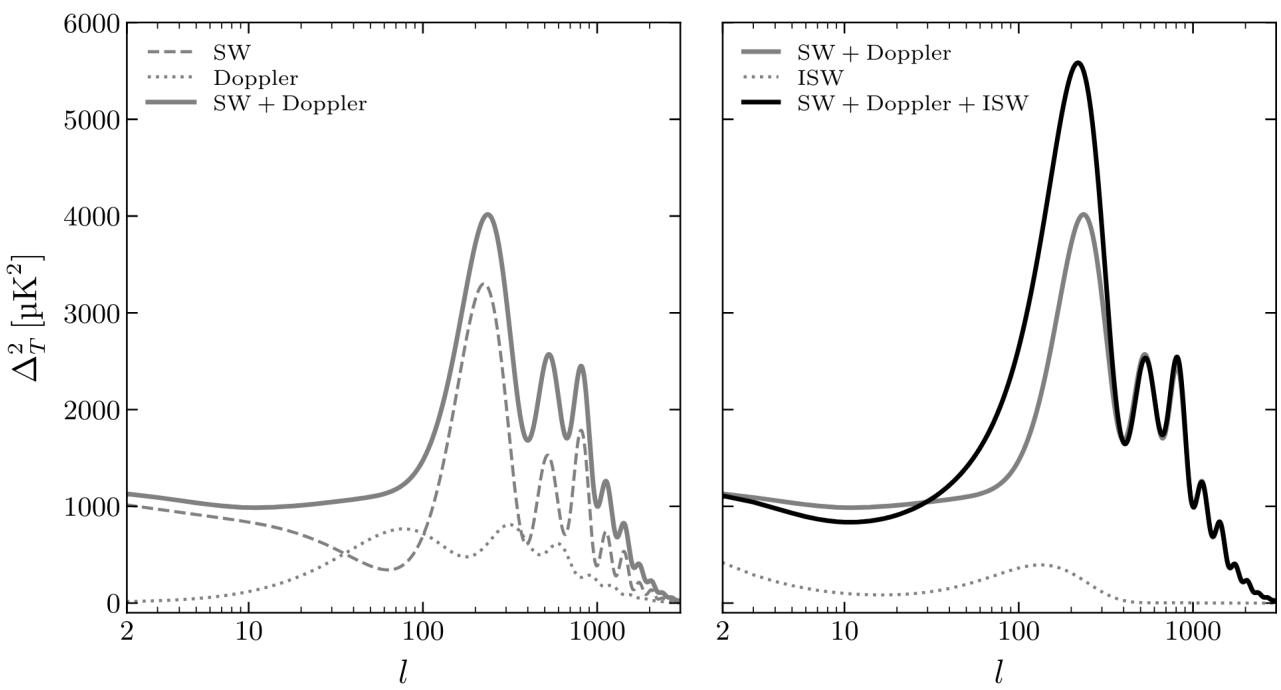


Figure 10.6: Illustration of the different contributions to the CMB power spectrum.

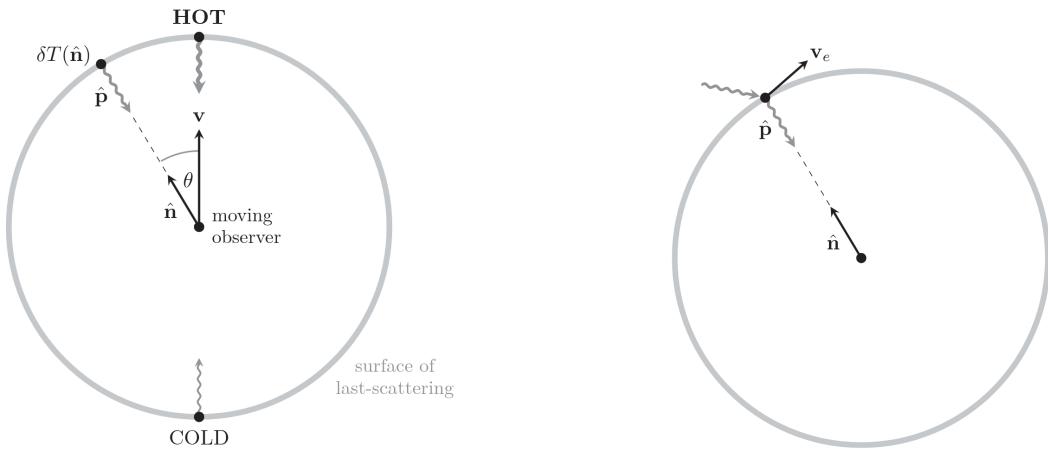


Figure 10.7: Doppler effect on the last-scattering surface.

§10.2.2.3 CMB power spectrum

The radiation density contrast is not the only contribution to the temperature anisotropies of the CMB. Indeed, the complete expression for δ_T as a function of the direction of observation is:

$$\delta_T(\hat{\mathbf{n}}) = \left(\frac{1}{4} \delta_r + \Psi \right)_* - (\hat{\mathbf{n}} \cdot \mathbf{v}_b)_* + \int_{\eta_*}^{\eta_0} d\eta (\Phi' + \Psi') \quad (10.2.36)$$

where \mathbf{v}_b is the bulk velocity of baryons relative to the observer RF, and the subscript * denotes the evaluation at η_* , i.e. at last scattering. These three terms are separate contributions:

- the first term is the *Sachs–Wolfe term*, which combines the intrinsic temperature fluctuations associated to the radiation density contrast with the induced temperature perturbations arising from the gravitational redshift of the photons. With the sign convention adopted in Eq. 10.2.8, $\Psi < 0$ corresponds to an overdensity (since, in the Newtonian limit, $\Psi \propto -M$), hence $\Psi_* < 0$ leads as expected to a temperature decrement, since a photon loses energy when climbing out of a potential well;
- the second term is a *Doppler term*, due to the relative motion of the last-scattering surface with respect to the observer (see Fig. 10.7). This contribution is subdominant on superhorizon scales, while its oscillations are out of phase with the oscillations in the Sachs–Wolfe term, thus reducing the contrast between the peaks in the spectrum;
- the last term describes the additional gravitational redshift due to the evolution of the metric potentials along the line-of-sight, determining what is known as the *integrated Sachs–Wolfe effect* (ISW). Since for most of the universe's history (namely, when it was matter dominated) these potential were constants, thus yielding no ISW effect. On the contrary, at early times the residual amount of radiation gives time-dependent gravitational potentials, resulting in an *early ISW effect*: this contribution raises the height of the first peak in the power spectrum (see Fig. 10.6).

The CMB power spectrum Fig. 10.5 imposes some constraints on the amount of matter, and in particular of dark matter (since baryonic matter is observed), in the universe: indeed, the extension of the Sachs–Wolfe plateau is determined by k_{eq} , and changing the amount of matter changed the time at which matter-radiation equality happens. Moreover, the second and third peaks of the power spectrum are sensible to the gravitational potential of matter: varying the amount of matter causes a variation in the height difference between the peaks.

Bibliography

- [1] L. Cerasi. 2025. URL: <https://github.com/LeonardoCerasi/notes/tree/master/personal/qft/main.pdf>.