INTRODUCTORY PARTICLE COSMOLOGY

Luca Doria

Institut für Kernphysik
J.J. Becher-Weg 45 55128 Mainz (Germany)
doria@uni-mainz.de

Chapter 1 | Preface

This document contains the notes of the course "Introductory Particle Cosmology" held in the 2018 Summer Term at the Johannes Gutenberg University in Mainz (Germany). The course is aimed at last year Bachelor and Master students which were not exposed before to the concepts of General Relativity (GR) but had introductory classes on quantum physics and the Standard Model (SM). In this sense, the course tries to be self-contained, starting from the mathematical tools and an introduction to GR and the SM. Often mathematical rigor or a more general treatment are sacrificed for the sake of clarity needed in an introductory course, which does not aim to substitute a full-fledged course on GR.

After having mastered some calculational techniques of GR, the applications to Cosmology are introduced. Emphasis will be devoted to the current standard cosmological model, its problems and challenges. In particular, Dark Matter and Dark Energy will be discussed, also in connection with the experimental approaches. The groundbreaking recent detection of gravitational waves opens a completely new opportunity for cosmology: the consequences of this new line of research in relation to the physics of the early universe will be discussed. If time allows, we will introduce the theory of cosmological perturbations, which is a fundamental tool for linking theory to experimental observations.

Contents

Chapter 2 | Geometrical Tools

In this chapter, we will review the most important tools for understanding GR. In particular, we have to develop a mathematical theory of curved spaces. We will try to study these topics sometimes abandoning mathematical rigor in favor of a more intuitive understanding, leaving a more abstract treatment in dedicated appendices. The concepts of vectors, vector spaces and vector operations are assumed to be known.

2.1 Vectors on a Generic Base

For simplicity, let us consider a two-dimensional space with two reference axes orthogonal to each other, x and y. We can consider unit basis vectors relative to these axes $e_x = (1, 0)$ and $e_y = (0, 1)$. A generic vector $\mathbf{v}$ can be represented in components using the basis vectors:

$$\mathbf{v} = v_x e_x + v_y e_y = v_x(1, 0) + v_y(0, 1) = (v_x, v_y) \quad . \tag{2.1}$$

The scalar product (here represented by a dot "·") can be used to "extract" a component from a vector. For example, $\mathbf{v} \cdot e_x = v_x$. Using the latter relation, we can rewrite the vector as

$$\mathbf{v} = (\mathbf{v} \cdot e_x)e_x + (\mathbf{v} \cdot e_y)e_y \quad . \tag{2.2}$$

This procedure works, because the basis vectors are orthogonal, i.e. $e_i \cdot e_j = 0$ for $i \neq j$ and $e_i \cdot e_j = 1$ for $i = j$ if the basis is also normalized to 1. In such cases, the basis is said to be *orthonormal* and in a space of arbitrary finite dimension the orthonormality property is expressed as

$$e_i \cdot e_j = \delta_{ij} \quad . \tag{2.3}$$

The scalar product characterizes also the length of a vector: $|v|^2 = \mathbf{v} \cdot \mathbf{v}$. What does it happen when we relax the requirements on our basis? Let's start dropping the normality property: now one or more basis vectors do not have unit length. We can still use the scalar product for expressing the single components if we take care of the different basis lengths changing Eq. 2.2 in the following way

$$\mathbf{v} = \mathbf{v} \cdot \frac{e_x}{|e_x|^2}e_x + \mathbf{v} \cdot \frac{e_y}{|e_y|^2}e_y \quad . \tag{2.4}$$

as it is easy to directly verify calculating $\mathbf{v} \cdot e_i = v_i$ with $i = x, y$. The similarity between Eq. 2.2 and Eq. 2.4 can be pushed further defining new basis components $e^i = e_i/|e_i|^2$ (i=x,y) obtaining

$$\mathbf{v} = (\mathbf{v} \cdot e^x)e_x + (\mathbf{v} \cdot e^y)e_y \quad . \tag{2.5}$$

Note that the new basis is indicated with upper indices. The latter expression can be rewritten identifying $(\mathbf{v} \cdot e^x)$ and $(\mathbf{v} \cdot e^y)$ with the new components v^x and v^y which are called **contravariant**, in contrast to **covariant** components with lower indices.

After normality, now we are ready to abandon also the requirement of orthogonality. Suppose that the basis $\{e_1, e_2\}$ (let's stay in two dimensions for simplicity) is not normal nor orthogonal. We have to try to generalize Eq. 2.4 in such a way that we still can rewrite the vector in contravariant components. This can be done introducing a proper basis. Since we do not have an orthogonal basis anymore, the idea is to construct a new basis $\{e^x, e^y\}$ which is orthogonal to the one at hand. This can be achieved requiring

$$e_i e^j = \delta_i^j \quad . \tag{2.6}$$

Sometimes this new basis is called *reciprocal basis*. It is straightforward to check that with the help of the new basis, we can still extract components as we did in the orthonormal and orthogonal cases. For example, we can try to extract v^x:

$$\mathbf{v} \cdot e^x = (v^x e_x + v^y e_y) \cdot e^x = v^x e_x \cdot e^x + v^y e_y \cdot e^x = v^x \cdot 1 + v^y \cdot 0 = v^x \tag{2.7}$$

Note that if we consider an orthonormal basis, the reciprocal basis coincides with the basis at hand and therefore covariant and contravariant components are the same. The differentiation makes sense only if the basis is not orthonormal.

$$E = \begin{pmatrix} -- & e_1 & -- \\ -- & e_2 & -- \\ -- & e_3 & -- \end{pmatrix} \quad \xrightarrow{U = A \cdot E} \quad \xleftarrow{E = A^{-1} \cdot U} \quad U = \begin{pmatrix} -- & u_1 & -- \\ -- & u_2 & -- \\ -- & u_3 & -- \end{pmatrix}$$

$$EE^R = 1 \qquad\qquad UU^R = 1$$

$$E^R = \begin{pmatrix} | & | & | \\ e^1 & e^2 & e^3 \\ | & | & | \end{pmatrix} \quad \xrightarrow{U^R = E^R A^{-1}} \quad \xleftarrow{E^R = U^R A} \quad U^R = \begin{pmatrix} | & | & | \\ u^1 & u^2 & u^3 \\ | & | & | \end{pmatrix}$$
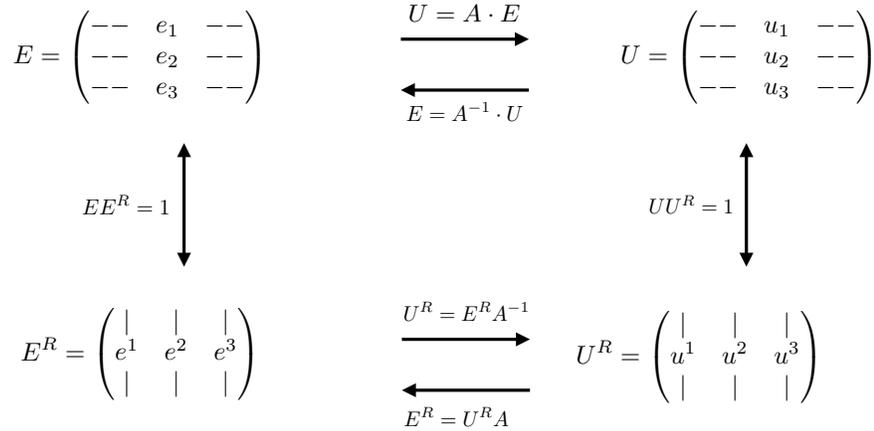
Figure 2.1: Transformations between bases and reciprocal bases.

2.2   Change of Basis

Two bases e_i and u_i are related by a transformation matrix

$$u_i = a_i^1 e_1 + a_i^2 e_2 + ... = a_i^j e_j \tag{2.8}$$

Changing notation, we can collect the N basis vectors of N components in NxN matrices E and U and rewrite the transformation as a matrix equation $U = A \cdot E$, where A is the matrix containing the transformation coefficients a_i^j. The rows of E and U are the single basis vectors. Solving for the transformation matrix:

$$A = U \cdot E^{-1} \quad . \tag{2.9}$$

Up to now we have expressed the new basis as linear combination of the old one, but also the other way around is possible: $E = A^{-1} \cdot U$. Solving for the transformation matrix we have $A^{-1} = E \cdot U^{-1}$.

Considering more closely the transformation in Eq. 2.8, we realize that the coefficients a_i^j are just the contravariant components of u_i in the e_i basis. As also Eq. 2.9 suggests, we can write

$$a_i^j = u_i e^j \quad , \tag{2.10}$$

3

where the contravariant basis vectors e^i come from the inverse matrix E^{-1}. The analog relation for the inverse transformation coefficients is

$$(a^{-1})_i^j = e_i u^j \quad .$$
(2.11)

Remembering the previous section, covariant and contravariant components belong to "reciprocal" bases, which were orthogonal to each other. Having now expressed u_i with contravariant components, we can ask how the components of the "reciprocal vector" u^i look like. In this case, the components must be covariant. Using Eq. 2.11

$$u^i = (u^i \cdot e_1)e^1 + (u^i \cdot e_2)e^2 + ... = (a^{-1})_j^i e^j$$
(2.12)

The above result means that the reciprocal basis vectors of the new basis are related to the reciprocal basis vectors of the old basis by the coefficients of the inverse transformation. If we invert the transformation and express E as a function of U, then the reciprocal basis transforms with coefficients from A and the original basis with A^{-1}. In matrix notation, calling the reciprocal bases U^R and E^R we have $U^R = E^R A^{-1}$ and $E^R = U^R A$. In the reciprocal basis, the columns (and not the rows as for the "normal" basis) are formed by the basis vectors. Recalling the property of reciprocal bases (Eq. 2.3) it is easy to verify that $U^R U = E^R E = 1$.

2.3 Covariant and Contravariant Vectors: Summary

Covariant and contravariant components are concepts arising naturally when we consider generic bases which are not orthonormal. The idea is to regain an orthonormality concept considering the reciprocal basis which has basis vectors orthogonal to the starting basis. This in turn is useful for maintaining a way to extract the components. In relation to basis transformations, all the dualities among bases are summarized in the diagram of Fig. 2.1. In the case when a basis is orthonormal, the reciprocal basis is exactly identical to the original base ($E = E^R$) and differentiating among contravariant and covariant components is not needed. As for the nomenclature, the terms "covariant" and "contravariant" are related to the transformation properties under a change of base. Covariant components transform with the coefficients of A, while contravariant components transform with A^{-1}.

2.4 Curves in the three-dimensional Space

Let's consider a curve f in the three dimensional space defined parametrically by

$$f(\xi) = \begin{cases} x = x(\xi) \\ y = y(\xi) \\ z = z(\xi) \end{cases}. \tag{2.13}$$

The parameter ξ belongs to an open set of $\mathbb{R}$ I=(a,b). We require also that the curve is regular, i.e. $\|df/d\xi\| \neq 0$ in I and it is a C^1 function. We define the **curvilinear abscissa**

$$S(\xi) - S(0) = \int_{\xi_0}^{\xi} \left\| \frac{df}{d\xi} \right\| d\xi = \int_{\xi_0}^{\xi} \sqrt{\left(\frac{dx}{d\xi}\right)^2 + \left(\frac{dy}{d\xi}\right)^2 + \left(\frac{dz}{d\xi}\right)^2} \tag{2.14}$$

The curvilinear abscissa S measures distances along the curve starting from ξ_0. What we would like to do now is to define an orthonormal basis of vectors on the curve. Since in general the curve is not a straight line, we expect that this basis will be different at different points along the curve. One of the vectors of this basis can be the tangent vector to the curve at a given point. The second vector will be the vector orthogonal to the tangent vector (the normal vector). As a third vector, using the vector product, we can construct a vector orthogonal to the plane defined by the tangent and normal vector:

$$\begin{aligned} \text{Tangent Vector:} \quad & \mathbf{t} = \frac{df}{dS} \\ \text{Normal Vector:} \quad & \mathbf{n} = \frac{d\mathbf{t}}{dS} / \|\frac{\mathbf{t}}{dS}\| \\ \text{Binormal Vector:} \quad & \mathbf{b} = \mathbf{t} \times \mathbf{n} \end{aligned} \tag{2.15}$$

The tangent vector has unit norm by construction, so $\mathbf{t}(S) \cdot \mathbf{t}(S) = 1$. If we show that $\mathbf{t}$ is orthogonal to $\mathbf{n}$, we have our orthonormal set of vectors. This can be done differentiating the latter normality equation:

$$\frac{d}{dS}(\mathbf{t}(S) \cdot \mathbf{t}(S)) = 2\mathbf{t} \cdot \frac{d\mathbf{t}}{dS} = 0 \quad . \tag{2.16}$$

Given the curve f, the defined orthonormal basis is fully determined and in this sense is called *intrinsic*.

It is now interesting to understand how the basis changes as we move along the curve. Actually we can picture the three orthonormal vectors rotating as the origin moves along the curve.
Starting with the binormal vector

$$\frac{db}{dS} = \frac{d}{dS}(t \times n) = \frac{dt}{dS} \times n + t \times \frac{dn}{dS} = Cn \times n + t \times \frac{dn}{dS} \quad . \qquad (2.17)$$

We introduced the new quantity $C = \frac{dt}{dS}$ called **curvature** of the curve. Moreover the latter equation states that $\tau = db/dS$ is orthogonal to t. Consider now the normal vector

$$\frac{dn}{dS} = \frac{d}{dS}(b \times t) = \frac{db}{dS} \times t + b \times \frac{dt}{dS} = \tau n \times t + b \times Cn = -\tau b - Ct \quad (2.18)$$

τ is called **torsion** of the curve. We can collect all the previous results in matrix form

$$\frac{d}{dS}\begin{pmatrix} t \\ n \\ b \end{pmatrix} = \begin{pmatrix} 0 & C & 0 \\ -C & 0 & -\tau \\ 0 & \tau & 0 \end{pmatrix} \begin{pmatrix} t \\ n \\ b \end{pmatrix} \qquad (2.19)$$

Note that the obtained transformation matrix is antisymmetric (like the angular velocity matrix: indeed the three axes "rotate" along the curve). Curvature and torsion are directly calculable from the parametric equation of the curve

$$C = \left\| \frac{dP}{dS} \times \frac{d^2P}{dS^2} \right\| \quad , \qquad (2.20)$$

$$\tau = \frac{dP}{dS} \times \frac{d^2P}{dS^2} \cdot \frac{d^3P}{dS^3} \quad . \qquad (2.21)$$

2.5 Vectors in Curved Geometries

In non-relativistic physics, space is assumed to have the natural structure of a three-dimensional space where there are no particular problems in defining vectors and operations among them. The same situation happens in special relativity with only minor complications due to the Lorentz structure. In curved geometries, vectors have to be defined in a different way, like also the operations among them. The intuitive idea is to consider a "tangent space" to a given point of the curved space (where

it is locally similar to a flat space) and define vectors there: in this way we regain the concept of a "flat" space where vectors as we used to know them can live. Having to deal with the concept of tangent space, we are forced to consider derivatives and indeed in these situations derivatives and vectors play basically the same role.

In a flat space, a vector can be described by two points A and B corresponding to the "tip" and the "tail" of it. We can look instead at the vector as a line between the two points, parametrically described by $P(\lambda) = A + \lambda(B - A)$ with $\lambda \in [0, 1]$. Differentiating

$$\frac{d}{d\lambda}(A + \lambda(B - A)) = B - A = V(1) - V(0) \tag{2.22}$$

This suggests that instead of defining a vector as a difference between two points, we can consider

$$v = \frac{dP}{d\lambda}|_{\lambda=0} \quad . \tag{2.23}$$

This point of view relates the vector to a single point and is more useful if we would like to define vectors on a curved space.

From a physical point of view, a vector looks naturally connected to the concept of derivative. If a body is moving along a generic curve parameterized by the time, its velocity can be calculated at every point as the time derivative of the displacement. So the derivative creates a tangent vector at every point on the curve.

Another important observation is that it is generally more useful to define objects on a curved space in a way that does not depend on any "embedding" (or coordinate charts) of it into a higher-dimensional space. Taking as curved space the bidimensional surface of a sphere, is easy to imagine it embedded into a three-dimensional euclidean flat space. What we would like to do, is to define vectors (or other objects) in a way which is dependent only from the characteristics of the surface, without the help of any embedding.

Tangent vectors as directional derivatives are the objects we are looking for. A tangent vector can be regarded as an infinitesimal displacement from a specific point of our curved space. All the tangent vectors at a point form a vector space which is called *tangent space*. Something like that is needed, since on a curved space the natual vector space structure

we have in Euclideand space is lost globally, but we can reconstruct it at least locally.

It can be showed that the tangent space has the same dimension D of the curved space and it is spanned by D linearly independent vectors which form a basis for it. If the curved space has coordinates x^μ ($\mu = 1..D$), the directional derivatives at a point are

$$v = v^\mu \frac{\partial}{\partial x^\mu} \quad . \tag{2.24}$$

Notice that this derivative has the structure of a vector, with components v^μ and vectors $\partial/\partial x^\mu$ which can be considered a basis for the tangent space. This is how we can see vectors on a curved space, without the need to go "outside" of it with an embedding: these vectors do not "stick out" from the curved space into a higher-dimension space.

Note that in general, the basis we have now is not orthonormal and this brings us back to the discussion about covariant and contravariant vectors. First of all, we have to construct the reciprocal basis, which has axes orthogonal to the ones of the starting basis. Calling dx^μ the reciprocal basis we have formally

$$dx^\mu \left(\frac{\partial}{\partial x^\nu} \right) = \delta^\mu_\nu \quad . \tag{2.25}$$

We can regard dx^μ as a linear function acting on vectors: this connects to the usual dual vectors known from basic linear algebra. We introduced the machinery of covariant and contravariant vectors for keeping intact important relations which otherwise would fail in non orthonormal bases, for example the extraction of the components. We can see that also in this case things work as expected. Let's apply dx^μ to the vector v defined in Eq. 2.24

$$dx^\mu(v) = dx^\mu(v^\nu \frac{\partial}{\partial x^\nu}) = v^\nu dx^\mu(\frac{\partial}{\partial x^\nu}) = v^\nu \delta^\mu_\nu = v^\mu \quad . \tag{2.26}$$

We can see things from another angle. As vectors applied to a function represent the derivative of the function in the direction of the vector (directional derivative), covectors (vectors of the reciprocal basis) applied to a function represent the gradient of it. If you remember from basic multidimensional analysis, gradients point in the orthogonal direction of

the surface where they are calculated, so we see again the idea of orthogonality between the normal base (here on the tangent plane) and the reciprocal base.

2.6   Transformation Properties

Now that we discussed contravariant and covariant representations of vectors and how these objects look like on a curved space, we have to study their transformation properties under a change of coordinates. Using the chain rule from calculus the contravariant components of a vector transform as

$$v'^{\mu'} = \frac{\partial x'^{\mu'}}{\partial x^{\mu}} v^{\mu} \quad , \tag{2.27}$$

while the covariant components transform as

$$v'_{\mu'} = \frac{\partial x^{\mu}}{\partial x^{\mu'}} v_{\mu} \quad . \tag{2.28}$$

NOTE: In all our discussion about curved spaces, we always assumed that they can locally treated as flat. A generic surface does not necessarily have this property. This is why the curved surfaces we consider are of a specific kind, called **manifolds**. We note also that we discussed generic "curved" spaces appealing to an intuitive understanding of them and we have given a clear definition of curvature only for one-dimensional curves in D=3 space. We will see how the curvature is defined for generic manifolds later on.

2.7   Tensors

Scalars and vectors are not the only objects representing physical quantities but actually they can be regarded as specialized forms of more general objects called **tensors**. Here we introduce tensors in an informal way, leaving the more formal algebraic definition to the many existing textbooks on the subject.
As vectors (contravariant, or covariant co-vectors) can be thought as "rows"

of numbers representing their components on a certain basis, we can build squares of numbers ("matrices") or "cubes" of numbers and so on. Labeling the components with indices, a vector has one index, matrices have two and a tensor have in general more indices.

At this point, it is useful to introduce some notation. Contravariant vector components are written with upper indices (v^i) while covariant components have lower indices (v_i). A tensor with a general number of indices can have both lower and upper indices: $T^{\alpha,\beta,\gamma,\cdots}_{i,j,k,\cdots}$. If a tensor has only lower (upper) indices is called covariant (contravariant), while tensors with both kinds of indices are sometimes called mixed tensors. Another notation is the so-called Einstein convention for the summation of repeated indices. For example

$$\sum_i v_i w^i = v_i w^i \quad .\tag{2.29}$$

The summation symbol is omitted if the sum runs over repeated indices (covariant and contravariant in this case). The key property of tensors is their transformation rule as coordinates change. An equation written in tensor notation does not change if coordinates are changed. The components can change, but the relations among tensors do not.

2.8 The Metric Tensor

As first example of tensor, we introduce the **metric tensor** which enables to calculate distances (hence the name) on a generic (curved) space. Recall how we calculate distances s in a flat D=3 euclidean space between two points x_1 and x_2

$$s^2 = (x_1^1 - x_2^1)^2 + (x_1^2 - x_2^2)^2 + (x_1^3 - x_2^3)^2 = \sum_i (x_1^i - x_2^i)^2 \quad .\tag{2.30}$$

This is just Pythagoras'Theorem. A curved space does not have in general straight lines as we know them in euclidean geometry and it is better to work with infinitesimal displacements dx_i in the neighborhood of a point, where the space is close to a flat one. The above expression can be rewritten as

$$ds^2 = \sum_i (dx_i)^2 = \delta_{ij} dx^i dx^j \tag{2.31}$$

10

The last step seems unnecessary, but one starts to see its usefulness when considering different spaces. In special relativity for example, the distance among events is given by the proper time τ:

$$\tau^2 = (x_1^0 - x_2^0)^2 - (x_1^1 - x_2^2)^2 + (x_1^2 - x_2^2)^2 + (x_1^3 - x_2^3)^2 = \sum_i (x_1^i - x_2^i)^2 \quad .$$
$$(2.32)$$

which in its infinitesimal version is

$$d\tau^2 = dx^0 - \sum_{i=1}^{i=3} (dx_i)^2 = \eta_{\mu\nu} dx^\mu dx^\nu \qquad (2.33)$$

with $\eta_{\mu\nu}$=diag(1,-1,-1,-1) with $\mu, \nu = 0, 1, 2, 3$. In a generic curved space, Pythagoras'Theorem assumes in general a different form:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu \qquad (2.34)$$

and the tensor $g_{\mu\nu}$ is called the **metric tensor** and tells us how to compute distances. The euclidean ($g_{\mu\nu} = \delta_{\mu\nu}$) and the Lorentz ($g_{\mu\nu} = \eta_{\mu\nu}$) cases have diagonal metric tensors, while a generic curved space can have also non-zero off-diagonal components which "mix" the coordinates when we calculate distances.

Notice that the metric tensor has to be a **symmetric tensor** i.e.: $g_{\mu\nu} = g_{\nu\mu}$. This is a necessary requirement since the distance ds cannot be different if we exchange dx_μ with dx_ν.

Expressing vectors on a basis e_i, it is clear that the metric tensor can be written also as

$$g_{ij} = e_i e_j \quad , \qquad (2.35)$$

therefore representing all the possible scalar products among all the components of the basis vectors.

2.9 Covariant Differentiation

We would like now to construct the derivatives of tensor quantities on a general curved space, or in other words understand how tensors change along a certain direction given by a tangent vector. In this sense, we would like to generalize the concept of directional derivative from the classical multidimensional calculus.

Let's start with the derivative of a vector v in the direction of w, which is $\nabla_w v(P)$. The latter expression means that we would like to calculate the derivative of v in the direction of w and the vector v is applied at the point P of our generic space. The result of this operation will be another vector applied in P. The important property we would like to have for ∇ is that the result do not change if we change coordinates. In this sense this derivative is called **covariant derivative**.

Another property the new derivative should respect is also the Leibnitz rule, but again we do not go into too many details here. Thinking about how we calculated derivatives of vectors in flat spaces, we recognize that we have to consider the difference of two vectors at two different points ($\nabla v \sim (v(x) - v(x+h))/h$). The difference can be calculated translating one vector to the same place of the other one and then calculating the difference. This procedure does not work in general in a curved space since we do not have a notion of how to transport vectors from one place to another keeping it "parallel" to its original location. Therefore we need a notion of **parallel transport** on a curved space. The parallel transport is realized by an **affine connection** on the curved space. A connection is called like that because it "connects" vectors living in different tangent spaces. Here we avoid a general discussion on connections which will bring us too far and introduce covariant differentiation in a more concrete (and less general) way using coordinates.

Consider a basis $e_i = \partial/\partial x^i$ for a tangent space at a certain point P. We would like to understand how a basis vector e_i changes in the direction given by another basis vector e_j. In the notation of directional derivatives, we would like to calculate $\nabla_{e_j} e_i$. The result of this calculation will be another vector and in particular, it will be proportional to the combination of other basis vectors e_k:

$$\nabla_{e_j} e_i = \Gamma^k_{ij} e_k \quad . \tag{2.36}$$

The three-indices object $\Gamma^k_{ij} e_k$ is called **Christoffel symbol** and we antici-pate here that it is not a tensor (i.e. it does not transform like a tensor). We know now how a basis vector changes in the direction of another ba-sis vector: what about the change of a generic vector w in the direction of

another vector v? Let's calculate it (for the contravariant vector case):

$$\nabla_v w = \nabla_{v^j e_j}(w^i e_i) = v^j \nabla_{e_j}(w^i e_i) = v^j w^i \nabla_{e_j} e_i + v^j e_i \nabla_{e_j} w^i \qquad (2.37)$$

$$= v^j w^i \Gamma_{ij}^k e_k + v^j \frac{\partial w^i}{\partial x^j} e_i = \left(v^j w^i \Gamma_{ij}^k + v^j \frac{\partial w^k}{\partial x^j} \right) e_k \qquad (2.38)$$

The result is interesting: the part containing the Christoffel symbol tells us how the basis vectors "turn" as we move them in the direction of v, while the other part tells us how the components of w change along v. While the Christoffel symbol is not a vector, it can be proved that the combination present in the covariant derivative has the right transformation properties.

A general mixed tensor has the following covariant derivative:

$$(\nabla_{e_c} T)_{b_1..b_s}^{a_1..a_r} = \frac{\partial \nabla_{e_c} T_{b_1..b_s}^{a_1..a_r}}{\partial x^c} \qquad (2.39)$$

$$+ \Gamma_{dc}^{a_1} T_{b_1..b_s}^{da_2..a_r} + ... + \Gamma_{dc}^{a_r} T_{b_1..b_s}^{a_1..a_{r-1}d} \qquad (2.40)$$

$$- \Gamma_{b_1 c}^{d} T_{db_2..b_s}^{a_1..a_r} - ... - \Gamma_{b_s c}^{d} T_{b_1..b_{s-1}d}^{a_1..a_r} \qquad (2.41)$$

Notice that the derivation of covariant indices brings a negative sign in front of Γ.

2.10   Calculation of the Christoffel Symbols

Having a procedure for calculating derivatives of tensors along a direction given by a tangent vector, we need now a way to calculate the Christoffel symbols.

First, we have to know the following property

$$\nabla_\rho g_{\mu\nu} = 0 \quad . \qquad (2.42)$$

The covariant derivative of the metric tensor is zero. This property is called *metric compatibility* and it is always assumed valid in our context, but nothing prevents to develop covariant derivatives without compatibility. A physically appealing property of compatibility is that given two vectors, their mutual angle does not change if they are parallely transported along a curve by a compatible derivative. In other words, the

scalar product of two vectors does not change during the parallel transport (i.e. the metric is a constant for the covariant differentiation). From Eq. 2.42, we can write the following three expressions, which are the same except for a cyclic index permutation ($\rho \to \mu \to \nu$)

$$\nabla_\rho g_{\mu\nu} = \partial_\rho g_{\mu\nu} - \Gamma^\lambda_{\rho\nu} g_{\lambda\nu} - \Gamma^\lambda_{\rho\nu} g_{\mu\lambda} = 0$$
$$\nabla_\mu g_{\nu\rho} = \partial_\mu g_{\nu\rho} - \Gamma^\lambda_{\mu\nu} g_{\lambda\rho} - \Gamma^\lambda_{\mu\rho} g_{\nu\lambda} = 0 \tag{2.43}$$
$$\nabla_\nu g_{\rho\mu} = \partial_\nu g_{\rho\mu} - \Gamma^\lambda_{\nu\rho} g_{\lambda\mu} - \Gamma^\lambda_{\nu\mu} g_{\rho\lambda} = 0 \quad .$$

Calculating $\nabla_\rho g_{\mu\nu} + \nabla_\mu g_{\nu\rho} - \nabla_\nu g_{\rho\mu} = 0$ we obtain

$$\partial_\rho g_{\mu\nu} - \partial_\mu g_{\nu\rho} - \partial_\nu g_{\rho\mu} + 2\Gamma^\lambda_{\mu\nu} = 0 g_{\lambda\rho} \quad , \tag{2.44}$$

where we used the symmetry $\Gamma^k_{ij} = \Gamma^k_{ji}$ (evident from Eq. 2.36). Multiplying the last equation by $g^{\sigma\rho}$ we can solve for Γ:

$$\Gamma^\sigma_{\mu\nu} = \frac{1}{2} g^{\sigma\rho} \left(\partial_\mu g_{\nu\rho} + \partial_\nu g_{\rho\mu} - \partial_\rho g_{\mu\nu} \right) \quad . \tag{2.45}$$

This very nice result tells us that given the metric tensor we can calculate the Christoffel symbols. Said in another way, the metric tensor is all what is needed to calculate covariant derivatives of tensors on a curved space.

2.11 Geodesics

As straight lines are the shortest lines in flat space, we can ask which curve connecting two points in a curved space is the shortest. These particular curves are called **geodesics** and in this sense are the "straight lines" in curved spaces. One way to see the problem of finding the geodesics is the following: the most "straight line" between two points is the one where a tangent vector T on the first point does not change (remains parallel to itself) if transported in the direction pointed by the vector itself. Translating in mathematical language the previous statement:

$$\nabla_{T^\mu} T^\nu = T^\mu \nabla_\mu T^\nu = 0 \quad , \tag{2.46}$$

and substituting the definition of covariant derivative

$$\frac{dT^\mu}{dt} + \Gamma^\mu_{\alpha\beta} T^\alpha T^\beta = 0 \quad . \tag{2.47}$$

The components of a tangent vector to a curve parameterized by a parameter t are just $T^\mu = dx^\mu/dt$, therefore we obtain the equation a geodesic curve must satisfy

$$\frac{d^2 x^\mu}{dt^2} + \Gamma^\mu_{\alpha\beta} \frac{dx^\alpha}{dt} \frac{dx^\beta}{dt} = 0 \quad . \tag{2.48}$$

The last result is a coupled system of second order linear differential equations and we know that there exists an unique solution given initial values for x and dx/dt.

Another way to obtain the geodesic equation is to directly calculate what is the shortest path between two points in a curved space. The squared distance is given by

$$L = \int_{t_1}^{t_2} g_{ij} \dot{x}^i \dot{x}^j dt \tag{2.49}$$

where t is a parameter describing the curve and $\dot{x} = dx/dt$ for a more compact notation. The shortest path is the one with vanishing variation $\delta L = 0$:

$$\delta L = \int_{t_1}^{t_2} \left(\delta g_{ij} \dot{x}^i \dot{x}^j + 2 g_{ij} \delta(\dot{x}^i) \dot{x}^j \right) dt = 0 \quad , \tag{2.50}$$

where we used the symmetry of g_{ij} for the second term in the integrand. Again in the second term, exchanging δ with d/dt and integrating by parts

$$\delta L = \int_{t_1}^{t_2} \left(\frac{\partial g_{ij}}{\partial x^k} \dot{x}^i \dot{x}^j \delta x^k + 2 g_{ij} \dot{x}^i \frac{d}{dt} \delta x^j \right) dt = \tag{2.51}$$

$$\int_{t_1}^{t_2} \left(\frac{\partial g_{ij}}{\partial x^k} \dot{x}^i \dot{x}^j - 2 \frac{d}{dt} \left[g_{ij} \dot{x}^i \right] \right) \delta x^k dt + g_{ij} \frac{dx^i}{dt} \delta x^j |_{t_1}^{t_2} = 0 \quad .$$

The last term calculated at the start and end points of the curve vanishes because the curve has no variation there (the extrema are kept fixed). The

last integral must vanish, and this happens when the integrand is zero:

$$\frac{\partial g_{ij}}{\partial x^k}\dot{x}^i\dot{x}^j - 2\frac{d}{dt}\left[g_{ij}\dot{x}^i\right] \tag{2.52}$$

$$= \left(\frac{\partial g_{ij}}{\partial x^k} - 2\frac{\partial g_{ij}}{\partial x^k}\right)\dot{x}^i\dot{x}^j - 2g_{ij}\frac{d^2x^i}{dt^2}$$

$$= \frac{d^2x^k}{dt^2} + \frac{1}{2}g^{kl}\left(\frac{\partial g_{il}}{\partial x^j} + \frac{\partial g_{il}}{\partial x^i} - \frac{\partial g_{ij}}{\partial x^l}\right)\dot{x}^i\dot{x}^j$$

$$= \frac{d^2x^k}{dt^2} + \frac{1}{2}g^{kl}\Gamma^k_{ij}\dot{x}^i\dot{x}^j = 0$$

We obtained again the geodesic equation as the shortest path between two points on a curved space. On a side-note, we can observe that if our space is **compact**, for every couple of points in it, there is always a geodesic connecting them.

2.12 Curvature

Up to now we always referred to a generic curved space, but now we have the tools for clearly defining what the curvature is. We remind that we defined it for a one-dimensional curve on three-dimensional euclidean space. It can be showed, that the curvature at a point P is the inverse of the radius of a circle tangent to P. In this sense, the curvature is also often called "radius of curvature". The idea is that if the curve is indeed very "curved" in the neighborhood of a point, the tangent circle will be "small". A small radius corresponds to a high curvature. Rising the dimension by one, a similar concept can be thought about bidimensional surfaces. In this case the curvature is characterized by two *principal curvatures*, as famously demonstrated by C.F. Gauss in his *Theorema Egregium* (1827). The theorem is indeed *egregium*, since it shows that the curvature is an *intrinsic* property of a surface and as such it can be calculated measuring properties on it (like angles and distances), without referring to an external embedding. In other terms, we do not need to know that the earth is a sphere immersed in a three-dimensional space: we can just make measurements on its surface for discovering that it is not flat! In the following we would like to characterize the concept of curvature to spaces of arbitrary dimension.

$$V^\mu(BC) = -\int_a^{a+\delta a} d\theta V^\beta \Gamma^\mu_{\theta\beta}|_{\phi=b+\delta b}$$

ϕ

$\theta = a$ $\theta = a + \delta a$

$\phi = b + \delta b$ **B**   **C** $\phi = b + \delta b$

$$V^\mu(AB) = -\int_b^{b+\delta b} d\phi V^\beta \Gamma^\mu_{\phi\beta}|_{\theta=a}$$ $$V^\mu(CD) = -\int_{b+\delta b}^b d\phi V^\beta \Gamma^\mu_{\phi\beta}|_{\theta=a+\delta a}$$

θ

A   **D**

$\theta = a$ $\theta = a + \delta a$

$\phi = b$ $\phi = b$

$$V^\mu(DA) = -\int_{a+\delta a}^a d\theta V^\beta \Gamma^\mu_{\theta\beta}|_{\phi=b}$$
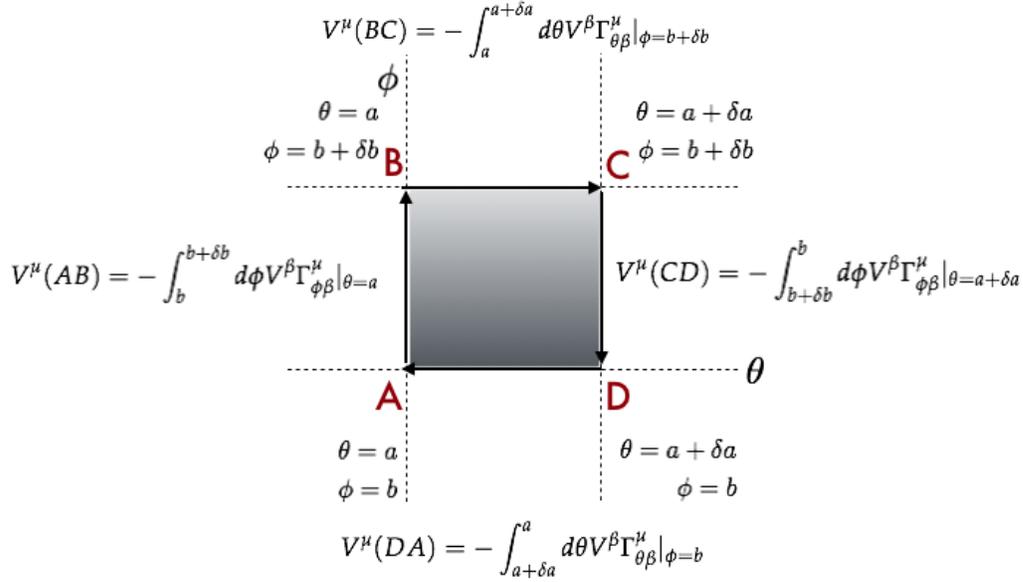
Figure 2.2: Parallel transport along a closed path.

2.12.1 Parallel Transport in Closed Loops

We start investigating what is happening to a tangent vector transported on a closed loop. For simplifying the discussion, let's think about a two-dimensional surface (e.g. a sphere) where we have two coordinates (longitude θ and latitude ϕ). In a small neighborhood (where we can use linear dispacements), let's consider a closed loop which resembles a square with sides AB, BC, CD, and DA. We would like to transport a vector V^μ from A to B and consider its difference:

$$V^\mu(AB) = V^\mu(B) - V^\mu(A) = \int_A^B d\phi \frac{\partial V^\mu}{\partial \phi}|_{\theta=a} \quad . \tag{2.53}$$

The vector is parallel transported along AB, so its covariant derivative is zero: $\nabla_i V^\mu = \frac{\partial V^\mu}{\partial x^i} + \Gamma^\mu_{i\beta} V^\beta = 0$ ($x^{1,2} = \theta, \phi$). Using the last formula and the values of the coordinates (see Fig.2.12)

$$V^\mu(AB) = -\int_b^{b+\delta b} d\phi V^\beta \Gamma^\mu_{\phi\beta}|_{\theta=a} \quad . \tag{2.54}$$

17

A similar calculation can be done for the other three segments:

$$V^\mu(BC) = -\int_a^{a+\delta a} d\theta\, V^\beta \Gamma^\mu_{\theta\beta}|_{\phi=b+\delta b} \quad ,$$

$$V^\mu(CD) = -\int_{b+\delta b}^{b} d\phi\, V^\beta \Gamma^\mu_{\phi\beta}|_{\theta=a+\delta a} \quad ,$$

$$V^\mu(DA) = -\int_{a+\delta a}^{a} d\theta\, V^\beta \Gamma^\mu_{\theta\beta}|_{\phi=b} \quad .$$

Adding together all the four parts of the closed path and using the definition of partial derivative

$$\Delta V^\mu = \int_b^{b+\delta b} d\phi \left[V^\beta \Gamma^\mu_{\phi\beta}|_{\theta=a+\delta a} - V^\beta \Gamma^\mu_{\phi\beta}|_{\theta=a} \right] \qquad (2.55)$$

$$- \int_a^{a+\delta a} d\theta \left[V^\beta \Gamma^\mu_{\theta\beta}|_{\phi=b+\delta b} - V^\beta \Gamma^\mu_{\theta\beta}|_{\phi=b} \right]$$

$$= - \left[\int_a^{a+\delta a} d\theta \delta b \frac{\partial}{\partial \phi} \left(V^\beta \Gamma^\mu_{\theta\beta} \right) - \int_b^{b+\delta b} d\phi \delta a \frac{\partial}{\partial \theta} \left(V^\beta \Gamma^\mu_{\phi\beta} \right) \right]$$

$$= -\delta a \delta b \left[\frac{\partial}{\partial \phi}(V^\beta \Gamma^\mu_{\theta\beta}) - \frac{\partial}{\partial \theta}(V^\beta \Gamma^\mu_{\phi\beta}) \right] \qquad . \qquad (2.56)$$

Substituting again the covariant derivative formula and noting that $\delta A = \delta a \delta b$ is the small area enclosed in the loop

$$\Delta V^\mu = -\delta A \left(\frac{\partial V^\beta}{\partial \phi} \Gamma^\mu_{\theta\beta} + V^\beta \frac{\partial \Gamma^\mu_{\theta\beta}}{\partial \phi} - \frac{\partial V^\beta}{\partial \theta} \Gamma^\mu_{\phi\beta} - V^\beta \frac{\partial \Gamma^\mu_{\phi\beta}}{\partial \theta} \right) \qquad (2.57)$$

$$= -\delta A \left(-V^\nu \Gamma^\beta_{\phi\nu} \Gamma^\mu_{\theta\beta} + V^\beta \frac{\partial \Gamma^\mu_{\theta\beta}}{\partial \phi} + V^\nu \Gamma^\beta_{\theta\nu} \Gamma^\mu_{\phi\beta} - V^\beta \frac{\partial \Gamma^\mu_{\phi\beta}}{\partial \theta} \right)$$

$$= \delta A V^\beta \left(\Gamma^\nu_{\phi\beta} \Gamma^\mu_{\theta\nu} - \frac{\partial \Gamma^\mu_{\theta\beta}}{\partial \phi} - \Gamma^\nu_{\theta\beta} \Gamma^\mu_{\phi\nu} + \frac{\partial \Gamma^\mu_{\phi\beta}}{\partial \theta} \right) \qquad .$$

We discovered that the variation in the vector components after a parallel transport around a closed loop is proportional to the enclosed area, times a combination of Christoffel symbols (depending at the end from the metric tensor). The quantity in parentheses in the last line of the last equation is actually a tensor and if we generalize it, replacing θ and ϕ

with the N coordinates x^α of a space of dimension N, we can rewrite it as

$$R^\mu_{\beta\gamma\lambda} = \Gamma^\nu_{\lambda\beta}\Gamma^\mu_{\gamma\nu} - \frac{\partial\Gamma^\mu_{\gamma\beta}}{\partial x^\lambda} - \Gamma^\nu_{\gamma\beta}\Gamma^\mu_{\lambda\nu} + \frac{\partial\Gamma^\mu_{\lambda\beta}}{\partial x^\gamma} \tag{2.58}$$

The quantity $R^\mu_{\beta\gamma\lambda}$ is called the **Riemann Tensor** and it characterizes the curvature of a generic multidimensional space. In the obtained form, it has one contravariant and three covariant indices. If a space is flat, then all the components of R vanish (and vice-versa), so this is a necessary and sufficient condition for flatness, *with the caveat that we do not have any information about the topology of the space* [1]. In the next section, we will see how this new tensor is connected with the Gaussian curvature.

2.12.2 Properties of the Riemann Tensor

It turns out that the Riemann tensor is the only tensor that can be constructed using the metric tensor and its derivatives: in this sense the Riemann tensor is unique. Another characterization of the Riemann tensor is the following

$$\nabla_d\nabla_c V^a - \nabla_c\nabla_d V^a = [\nabla_d, \nabla_c]V^a = R^a_{bcd}V^b \tag{2.59}$$

Therefore, the tensor describes the non-commutativity of covariant derivatives. In a flat space the covariant derivatives are directional derivatives which commute, so all the tensor components are zero. Other useful properties are

$$R_{abcd} = -R_{bacd} = -R_{abdc} = R_{badc} \quad \text{(symmetry)} \tag{2.60}$$
$$R_{abcd} + R_{adbc} + R_{acdb} = 0 \quad \text{(antisymmetry)} \tag{2.61}$$
$$R_{abcd} + R_{adbc} + R_{acdb} = 0 \quad \text{(cyclicity)} \tag{2.62}$$

Another key property of the Riemann tensor is the so-called **Bianchi identity**

$$\nabla_e R_{abcd} + \nabla_d R_{abec} + \nabla_c R_{abde} = 0 \tag{2.63}$$

---

[1]Although a connection to the topology of the manifold can be provided under certain requirements through the Gauss-Bonnet theorem, which links the integral of the curvature over the manifold to the Euler characteristic (or the genus) of it.

Contracting (summing) the indices a and c and remembering that the covariant derivative of the metric tensor is zero

$$\nabla_e R_{bd} + \nabla_d R_{be} + \nabla_c R^c_{bde} = 0 \tag{2.64}$$

Contracting again b and d

$$\nabla_e R_{bd} + \nabla_d R_{be} + \nabla_c R^c_{bde} = \nabla_b(R^b_e - \frac{1}{2}\delta^b_e R) = \nabla_b G^b_e = 0 \tag{2.65}$$

Let's rewrite the last equation with other indices' names and introducing the metric tensor for having only covariant indices

$$\nabla_\mu G_{\mu\nu} = \nabla_\mu(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R) = 0 \tag{2.66}$$

The tensor $G_{\mu\nu}$ is called **Einstein Tensor** and has the interesting property of having zero covariant derivative. This is a central result in the development of the general theory of relativity.

2.13 A Worked-out Example

If the material covered up to now might have looked too abstract, then it is time to try a direct application of it to a simple case. Let's consider a two-dimensional surface as case study. A further simplification is to consider a surface of constant curvature, i.e. a surface where the curvature does not depend from the point where it is calculated. We are going to consider the surface of a sphere with fixed radius R. The sphere is a manifold (although we never really discussed its exact definition) and as such it can be mapped in a smooth way on a flat plane with the aid of at least two maps (which, in the language of manifolds, form an *atlas*). The usual map is

$$\begin{cases} x = R\cos\theta\sin\phi \\ y = R\sin\theta\sin\phi \\ z = R\cos\phi \quad, \end{cases}$$

with $\theta \in [0°, 180°]$ and $\phi \in [0°, 360°)$. Inverting the map

$$\begin{cases} \theta = \tan^{-1}\frac{y}{x} \\ \phi = \cos^{-1}(z/R) \quad . \end{cases}$$

The Jacobian of this coordinate transformation is zero at $\sin\theta = 0$, so two points are excluded from the mapping: in this sense a single map is not sufficient to describe this surface and we need a more complete atlas (actually in this case the smallest atlas contains two maps). Another way to see the problem is to recall that it is not possible to map the entire surface of the earth on a piece of paper: at least two "poles" will always escape our projection.

Our plan is now to build the metric tensor: this is the only object we need for deriving the Christoffel coefficients, and calculate covariant derivatives, the Riemann tensor, the Ricci tensor and the curvature scalar.

In a flat space, the squared infinitesimal distance is calculated through Pythagoras's Theorem: $ds^2 = dx^2 + dy^2 = \delta_{ij}dx^idx^j$. In this case the metric tensor is just the Kronecker δ symbol. If we perform a coordinate transformation to coordinates y^i

$$ds^2 = \delta_{ij}\frac{\partial y^i}{\partial x^k}dx^k\frac{\partial y^j}{\partial x_l}dx^l = \delta_{ij}\frac{\partial y^i}{\partial x^k}\frac{\partial y^j}{\partial x^l}dx^kdx^l = \frac{\partial y^k}{\partial x^i}\frac{\partial y^k}{\partial x^j}dx^idx^j \quad . \tag{2.67}$$

where in the last step the δ forces $i = j$ and then we changed the name of the indices. The last equation tells us how to calculate the metric tensor, given the coordinate transformations in Eq. 2.67. For example

$$g_{\theta\theta} = \frac{\partial x}{\partial\theta}\cdot\frac{\partial x}{\partial\theta} + \frac{\partial y}{\partial\theta}\cdot\frac{\partial y}{\partial\theta} + \frac{\partial z}{\partial\theta}\cdot\frac{\partial z}{\partial\theta} = \tag{2.68}$$

$$R^2\sin^2\theta\sin^2\phi + R^2\cos^2\theta\sin^2\phi + 0 = R^2\sin^2\phi \quad . \tag{2.69}$$

After calculating also $g_{\theta\phi} = g_{\phi\theta}$ and $g_{\phi\phi}$ we obtain

$$g_{ij} = \begin{pmatrix} g_{\theta\theta} & g_{\theta\phi} \\ g_{\phi\theta} & g_{\phi\phi} \end{pmatrix} = \begin{pmatrix} R^2 & 0 \\ 0 & R^2\sin^2\phi \end{pmatrix} \tag{2.70}$$

Note also that $g^{ij} = (g^{ij})^{-1}$. Having the metric tensor, we can calculate the Christoffel symbols with Eq. 2.45. For example

$$\Gamma_\theta^{\theta\theta} = \frac{1}{2}g^{\theta\theta}\left(\frac{\partial g_{\theta\theta}}{\partial\theta} + \frac{\partial g_{\theta\theta}}{\partial\theta} - \frac{\partial g_{\theta\theta}}{\partial\theta}\right) = \tag{2.71}$$

$$\frac{1}{2}\frac{1}{R^2}(0 + 0 + 0) = 0 \quad .$$

Calculating all the coefficients we have

$$\Gamma^{\theta}_{ij} = \begin{pmatrix} 0 & 0 \\ 0 & -\sin\theta\cos\phi \end{pmatrix} \quad ; \quad \Gamma^{\phi}_{ij} = \begin{pmatrix} 0 & \frac{\cos\phi}{\sin\phi} \\ \frac{\cos\phi}{\sin\phi} & 0 \end{pmatrix} \tag{2.72}$$

This allows us to calculate covariant derivatives (parallel transport of vectors) on the sphere's surface and also write the geodesic equations. Parameterizing a curve on the surface with a parameter t, the geodesic equations Eq. 2.53 become

$$\frac{d^2\theta}{dt^2} + 2\frac{\cos\phi}{\sin\phi}\frac{d\theta}{dt}\frac{d\phi}{dt} = 0$$

$$\frac{d^2\phi}{dt^2} - \sin\phi\cos\phi\left(\frac{d\theta}{dt}\right)^2 = 0 \quad .$$

We know already intuitively that geodesics on a sphere are maximal circumferences (meridians) and this can be seen from the obtained equations, for example fixing one angle and looking at the curve described by the other.

Let's try now to calculate the components of the Riemann tensor. In the two-dimensional case the Riemann tensor has dimension$^{\text{\#indices}} = 2^4 = 16$ components but only one is not zero. Inserting the Christoffel symbols (Eq. 2.72) into the definition of the Riemann tensor (Eq. 2.58) we find

$$R^{\phi}_{\theta\phi\theta} = \frac{\partial\Gamma^{\phi}_{\theta\theta}}{\partial\phi} - \Gamma^{\phi}_{\theta\theta}\Gamma^{\theta}_{\theta\phi} = \sin^2\theta \quad . \tag{2.73}$$

Contracting the Riemann tensor with the metric tensor we can calculate the Ricci tensor. The only non-zero components are

$$R_{\phi\phi} = g^{ij}R_{i\phi j\phi} = g^{\phi\phi}R_{\phi\phi\phi\phi} = \sin^2\phi$$

$$R_{\theta\theta} = 1 \quad .$$

Contracting the Ricci tensor with the metric tensor we obtain the Ricci scalar ^{2}

$$R = g^{ij}R_{ij} = g^{\theta\theta}R_{\theta\theta} + g^{\phi\phi}R_{\phi\phi} = \frac{2}{R^2} \tag{2.74}$$

---

^{2}It turns out that for two-dimensional surfaces, the Ricci scalar is just twice the Gaussian curvature.

The Ricci scalar is interpreted as the curvature of the surface: indeed at high curvatures correspond a small R, or put in another way, we consider a small sphere more curved than a large one. From another point of view: if R is very large, the spherical surface looks like a plane and this is what we experience every day on earth. As a side-note, the Ricci scalar is the equivalent to the curvature of surfaces found by Gauss in his research on bidimensional spaces. The Riemann formalism presented here is just more general and applicable to curved spaces of any dimension.

2.14   A Flatness Test

In his famous book, S. Weinberg proposes a flatness test for Tolkien's Middle Earth. In Fig. 2.13, we can try a similar exercise applied to the country where this course was first given. On the map there are four points corresponding to the cities A=(München), B=(Bremen), C=(Rostock), D=(Dresden). The distances from each other are:

AB = 584 km
AC = 664 km
AD = 360 km
BC = 245 km
BD = 406 km
CD = 365 km

One way to check if the usual euclidean geometry works, is to calculate one of the distances, for example AC, using the available data and then confront it with the value AC=664 km. For example, we can calculate the angle at the corner D between the segments AD and DC (without using AC). After obtaining the angle, we can try to calculate AC with the Cosines'Theorem and see if it agrees with the data. A more formal approach is based on the formula for calculating the (square) volume of a tetrahedron in three dimensions where each side corresponds to one of
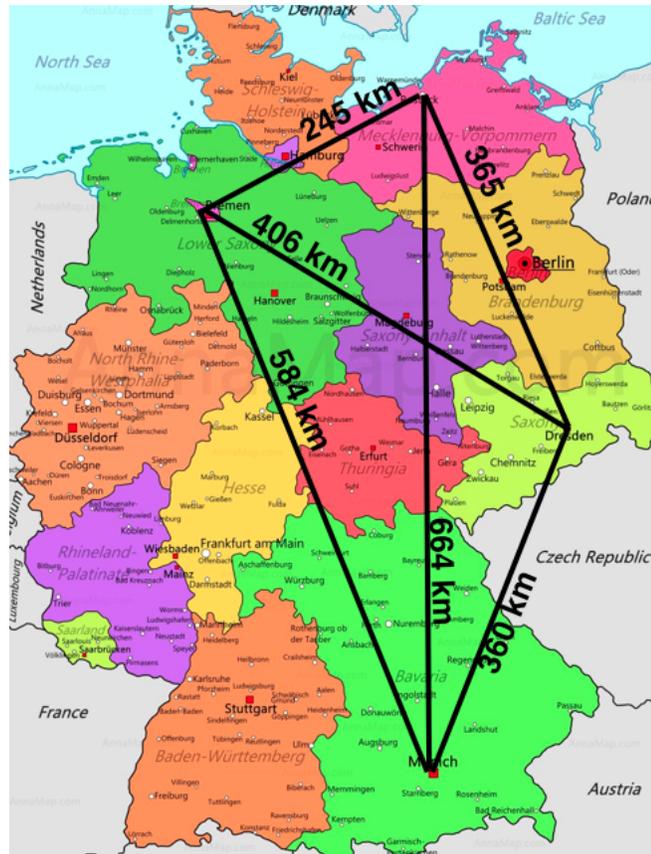
Figure 2.3: Is Germany flat? In the picture four points (A=München, B=Bremen, C=Rostock, D=Dresden) are showed on the map, with their reciprocal distances. If the Earth would be flat, the usual euclidean geometry should not work.

our city distances

$$V^2 = \frac{1}{288} det \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & \bar{A}B^2 & \bar{A}C^2 & \bar{A}D^2 \\ 1 & \bar{A}B^2 & 0 & \bar{B}C^2 & \bar{B}D^2 \\ 1 & \bar{A}C^2 & \bar{B}C^2 & 0 & \bar{C}D^2 \\ 1 & \bar{A}D^2 & \bar{B}D^2 & \bar{C}D^2 & 0 \end{pmatrix} \tag{2.75}$$

If the above determinant (the *Cayley-Menger determinant*) is zero, the tetrahedron collapses and the relations we obtain among the distances are the conditions for the four cities to lay on a plane.

Chapter 3 | General Relativity

In this chapter, we review the central concepts at the foundations of General Relativity (GR) and derive its fundamental equations. We will make use of all the geometrical tools developed in the previous chapter.

3.1 The Equivalence Principle

Gravity has the unique property of impressing the same acceleration to bodies regardless from their mass. This observation leads to the indistinguishably between gravity or an accelerated reference frame. Following an ideal experiment suggested by Einstein, if we are inside an uniformly accelerated elevator, we might think to be immersed in a gravitational field (or the other way around: we think that the elevator is accelerating, but it just hangs still over the surface of a planet). There is a key observation to make here: if the elevator is large enough, we can understand if it is accelerated or at rest into a gravitational field. In the latter case, we could for example note that the acceleration vectors in the elevator are not all parallel. For example, if the elevator is immersed in the gravitational field of a planet, the field lines converge to a single point (the center of the planet). Therefore, an accelerated reference frame or a gravitational field are indistinguishable only *locally*.

We can state now the so called **(strong) Equivalence Principle (EP)**: *at every space-time point in a gravitational field, it is possible to choose a locally inertial coordinate system such that in a sufficiently small neighborhood of that point the laws of nature are expressed in the same form as in an unaccelerated Cartesian (flat) coordinate system.* There is also a weaker version of the principle, called **(weak) Equivalence Principle** which instead of referring to all the laws of nature, it refers only to the laws of motion of free-falling

bodies. Clearly the strong version implies the weak on but not vice-versa.

3.2 Free-Falling Bodies

Let's try to translate in mathematical formulas the ideas contained in the EP. Consider a free-falling body: according to the EP, there must exist locally a coordinate system where the effect of gravitation vanishes (inside Einstein's elevator) and the equation of motion in flat space time is

$$\frac{d^2\xi^\mu}{d\tau^2} = 0 \quad , \tag{3.1}$$

where $d\tau = -\eta_{\mu\nu}d\xi^\mu d\xi^\nu$ is the proper time and $\eta_{\mu\nu}$ the Lorentz metric. Let's get out from the elevator and change to new coordinates x^μ, which can be whatever we want (a curvilinear system, an accelerated or rotating system, etc..)

$$\frac{d}{d\tau}\left(\frac{\partial\xi^\alpha}{\partial x^\mu}\frac{dx^\mu}{d\tau}\right) = \frac{\partial\xi^\alpha}{\partial x^\mu}\frac{d^2x^\mu}{d\tau^2} + \frac{\partial^2\xi^\alpha}{\partial x^\mu\partial x^\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = 0 \quad . \tag{3.2}$$

Multiplying the last equation by $\partial x^\lambda/\partial\xi^\alpha$ and recognizing the presence of the Christoffel symbol

$$\frac{d^2x^\lambda}{d\tau^2} + \Gamma^\lambda_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = 0 \quad . \tag{3.3}$$

The last result is quite interesting: the presence of the gravitational field can be seen as a curvature of space-time where free-falling particles follow a geodetic in that space. From the coordinate transformation formula and $d\tau = -\eta_{\mu\nu}d\xi^\mu d\xi^\nu$, it is clear that the metric tensor of the curved space is related to the Lorentzian flat space by

$$g_{\mu\nu} = \frac{\partial\xi^\alpha}{\partial x^\mu}\frac{\partial\xi^\beta}{\partial x^\nu}\eta_{\alpha\beta} \quad . \tag{3.4}$$

3.3 Non-Relativistic Limit

So far our calculations were relativistic and now we would like to see if what we derived can be reduced to the know non-relativistic result, which

should be classical Newtonian gravity. The non-relativistic limit refers to velocities smaller than the speed of light ($v \ll c$) and weak, stationary gravitational fields. Remembering the composition of the velocity four-vector $dx^\mu/d\tau = (dt/d\tau, dx/d\tau)$, in this limit $dx/d\tau \ll dt/d\tau$ so the only non-zero component of the velocity is $\mu = 0$ and therefore

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma^\mu_{00}\left(\frac{dt}{d\tau}\right)^2 = 0 \quad . \tag{3.5}$$

For a weak stationary gravitational field, the space-time geometry must be almost flat: $g_{\alpha\beta} = \eta_{\alpha\beta} + h_{\alpha\beta}$ with $|h_{\alpha\beta}| \ll 1$.
Substituting this metric tensor into the Christoffel symbol we obtain

$$\Gamma^\alpha_{00} = -\frac{1}{2}\eta^{\alpha\beta}\frac{\partial g_{00}}{\partial x_\beta} \quad . \tag{3.6}$$

Reinserting the Christoffel symbol in Eq. 3.5 and separating the $\mu = 0$ "time" and $\mu = 1,2,3$ "space" parts we have

$$\frac{d^2 t}{d\tau^2} = 0 \quad ; \quad \frac{d^2 \mathbf{x}}{d\tau^2} - \frac{1}{2}\left(\frac{dt}{d\tau}\right)^2 \nabla h_{00} = 0 \quad . \tag{3.7}$$

The first equation tells us that $dt/d\tau$ is a certain constant K and we choose K=1. Substituting the first equation in the second and comparing it with the equation of motion with the gravitational potential $d^2 x/dt^2 = -\nabla\phi$ we find $h_{00} = -2\phi + C$ where C is another constant. Since the potential must be zero at infinity, we can fix also the second constant C=0. Reinserting the result for h in the original metric tensor we finally have

$$g_{00} = -(1 + 2\phi) \tag{3.8}$$

which is its the only non-zero component in the low-velocity, low-gravitational field approximation. We have showed here that there is a choice of the metric tensor which in the non-relativistic limit makes the relativistic geodesic equation is consistent with Newtonian gravity.

3.4 Energy-Momentum Tensor

Our aim is to find general relativistic equations for the gravitational field which are valid in every reference frame (not only the inertial ones). Such

equation must have tensorial character, and since gravitational fields are created by matter and energy distributions, it it meaningful to find a (relativistic) tensorial description for them. The object we are looking for is the energy-momentum tensor. We know already from electrodynamics that in the 4-dimensional formalism charge density and current vector can be organized in a single four-vector. We are going to do something similar for the 4-momenta p^α of a system of N particles labeled with the index n. The momentum density is

$$T^{\alpha 0} = \sum_n p_n^\alpha \delta^3(x - x_n) \quad . \tag{3.9}$$

Note that in this definition we are already thinking at the density as the zeroth-component of a tensor T. In this case T is a tensor with two indices, since one index spans the 4-vector components, while the other one will label the density and the three components of the current which we define as

$$T^{\alpha i} = \sum_n p_n^\alpha \frac{dx_n^i}{dt} \delta^3(x - x_n) \quad . \tag{3.10}$$

where the latin index i runs only on the "space" coordinates 1,2,3. Merging the last two equations into a single tensor

$$T^{\alpha \beta} = \sum_n p_n^\alpha \frac{dx_n^\beta}{dt} \delta^3(x - x_n) = \sum_n \frac{p_n^\alpha p_n^\beta}{E_n} \delta^3(x - x_n) \quad , \tag{3.11}$$

where we used the known relativistic result $v = p/E$. From the last expression, it is clear that $T^{\alpha\beta} = T^{\beta\alpha}$, and therefore the energy-momentum tensor is symmetric.

As in classical physics the time derivative of the momentum gives the force, in this context we have

$$\frac{\partial T^{\alpha\beta}}{\partial x^\beta} = F^\alpha \quad , \tag{3.12}$$

where F^α is a density of forces' 4-vector. In absence of forces, $\partial T^{\alpha\beta}/x^\beta = 0$ and this represents the energy-momentum conservation law. On a generic curved space, the partial derivative is substituted by the covariant one: $\nabla_\alpha T_{\alpha\beta} = 0$.

3.5 The Einstein Equations

We are now in the position to derive the generally covariant equations for the gravitational field. The distribution of matter and energy exerts a gravitational force, and we have seen that gravitation itself is related to the space-time metric. So we expect that the energy matter distribution is somehow related to space-time geometry. For guessing the correct equations, we can list first the requirements they have to obey:

1. The equations have to be tensor equations, thus retaining their form in any coordinate system. This requirement is connected to the equivalence principle.

2. In analogy to the other field equations of physics, they have to be partial differential equations of (at most) second order in the variable expressing the gravitational potential. In this case such variable is $g_{\mu\nu}$, the metric tensor, as we have seen in the approximate case of the Newtonian non-relativistic limit.

3. The equations must reduce to the Poisson equation for the gravitational potential in the non-relativistic weak-field limit.

4. $T^{\mu\nu}$ must be the source of the gravitational field, since it encodes the energy-matter distribution.

5. If the space-time is flat (no gravitational field), then $T^{\mu\nu} = 0$.

From requirements 1. and 4., the equations must have a form like

$$G^{\mu\nu} \propto T^{\mu\nu} \quad . \tag{3.13}$$

Since we know that from energy-momentum conservation $\nabla_\mu T^{\mu\nu} = 0$, we require that $\nabla_\mu G^{\mu\nu} = 0$. Moreover, since T is symmetric, also G must be symmetric. From the previous chapter, we know already a symmetric, two-indices tensor which contains $g_{\mu\nu}$ and its derivatives up to second order: the Einstein tensor. So we can guess the following form

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = k \cdot T_{\mu\nu} \quad . \tag{3.14}$$

where now we use covariant indices, $R_{\mu\nu}$ is the Ricci tensor, R is the Ricci scalar and k is a constant. We have to check now if requirement 3. holds.

We have to use the weak-field and $v \ll c$ approximations together with stationarity $\partial g_{\mu\nu}/\partial t = 0$. In this limit, the only non-zero component of the Ricci tensor is R_{00}. The energy-momentum tensor reduces also to only the energy density component T_{00} which in the non-relativistic limit is just the matter density ρ. The approximate equation is therefore

$$R_{00} = \frac{1}{2}\nabla^2 h_{00} = k\rho \quad , \tag{3.15}$$

which has to be compared to the Poisson equation for the gravitational potential

$$\nabla^2 \phi = 4\pi G \rho \quad . \tag{3.16}$$

where G is the Newton constant. Calculating the constant k, we can write the general relativistic Einstein equation (in natural units $G = c = 1$)

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi T_{\mu\nu} \quad . \tag{3.17}$$

In "normal units", $8\pi \to 8\pi G/c^4$.

The equation has been checked against many astronomical data, and laboratory and satellite experiments, always finding good agreement up to now.

Eq. 3.17 is not the most general form allowed by our requirements. Since the covariant derivative of both sides of the equation vanishes and this happens also for the metric tensor, we can also add a term which is proportional to $g_{\mu\nu}$

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi T_{\mu\nu} \quad . \tag{3.18}$$

The new constant Λ is called **cosmological constant**.

Given the symmetry of the tensors in the equation, there are only 10 independent components. This means that the Einstein equation represents a coupled system of 10 non-linear second-order partial differential equations and finding analytical general solutions is possible only in few cases characterized by high symmetry content. Besides the trivial flat solution, a particularly important space-time satisfying the Einstein equations is the **Schwarzschild solution** which finds wide applications in physics problems involving a spherically symmetric gravitational field.

3.6 Trace-Reversed Form of the Einstein Equations

There is another equivalent form for Eq. 3.18 which can be obtained taking first its trace

$$-R + 4\Lambda = 8\pi T \quad , \tag{3.19}$$

where R and T are the traces of the Ricci tensor and energy-momentum tensor respectively. Multiplying the last trace formula by $g_{\mu\nu}/2$ and substituting the result again in Eq. 3.18, we obtain the "trace-reversed" form of the Einstein equations

$$R_{\mu\nu} - \Lambda g_{\mu\nu} = 8\pi(T_{\mu\nu} - \frac{1}{2}T g_{\mu\nu}) \quad . \tag{3.20}$$

This version of the equation allows some interesting observations. First, in absence of the cosmological constant, matter and energy we have $R_{\mu\nu} = 0$, which represent a Ricci-flat space-time. Ricci-flat spacetimes are the solutions of GR for the completely empty space. The flat space-time is a trivial example of Ricci-flat space-time. A classical non-trivial example of vacuum solution is the Schwarzschild solution describing the space-time around a spherical mass. In absence of matter and energy, and $\Lambda \neq 0$ we have

$$R_{\mu\nu} = \Lambda g_{\mu\nu} \quad , \tag{3.21}$$

and it is tempting to do the identification $T_{\mu\nu} = \Lambda g_{\mu\nu}$ and thinking at the cosmological constant as the energy content of the vacuum itself.

3.7 Geodesic Deviation

There is another interpretation of the role played by the Riemann tensor in General Relativity. If a free-falling observer observes a nearby free-falling object, if there is no gravity, he should see it at rest. If gravity is present, the observer and the object should move with respect to each other. The free-falling observer follows the trajectory

$$\frac{d^2 x^\mu}{d\tau^2} + \Gamma^\mu_{\nu\lambda} \frac{dx^\nu}{d\tau} \frac{dx^\lambda}{d\tau} = 0 \quad . \tag{3.22}$$

Another observer is closeby, at $x^\mu(\tau) + \delta x^\mu(\tau)$ thus following the trajectory

$$\frac{d^2}{d\tau^2}(x^\mu + \delta x^\mu) + \Gamma^\mu_{\nu\lambda}(x^\mu + \delta x^\mu)\frac{d}{d\tau}(x^\nu + \delta x^\nu)\frac{d}{d\tau}(x^\lambda + \delta x^\lambda) = 0 \quad . \quad (3.23)$$

The difference between the last two equations at first order in δx^μ is

$$\frac{d^2\delta x^\mu}{d\tau^2} + \frac{\partial\Gamma^\mu_{\nu\lambda}}{\partial x^\rho}\delta x^\rho\frac{dx^\nu}{d\tau}\frac{dx^\lambda}{d\tau} + 2\Gamma^\mu_{\nu\lambda}\frac{dx^\nu}{d\tau}\frac{dx^\lambda}{d\tau} = 0 \quad , \quad\quad\quad (3.24)$$

which in terms of a covariant derivative along a curve [1] can be written as

$$\frac{D^2}{D\tau^2}\delta x^\lambda = R^\lambda_{\nu\mu\rho}\delta x^\mu\frac{dx^\nu}{d\tau}\frac{dx^\rho}{d\tau} \quad . \quad\quad\quad (3.25)$$

In absence of gravity, the Riemann tensor is identically zero and two close geodetics stay "parallel" to each other. If gravity is present, two nearby particles will not conserve their distance along the motion. In this sense, the Riemann tensor can be regarded as quantifying the amount of *geodesic deviation*. The relative acceleration detected among two nearby particles can be thought to be caused by a *tidal force*.

3.8   Summary

Eq. 3.17 was written first by Einstein in 1915 (with Riemann almost contemporarily providing a derivation based on a variational method). General Relativity can thus be summarized as follows:

Space-time is described by a manifold M equipped with a Lorentz metric. The curvature of M (computable from the metric) is related to the matter/energy distribution in M by the Einstein equation.

Eq. 3.17 represent 10 non-linear partial differential equations of the hyperbolic kind (like the wave equation) and they are the gravitational analog to the Maxwell equations written with the relativistic formalism where

---

[1]If V^μ is a vector, its derivative along a curve x^μ parameterized by a parameter τ is $DV^\mu/D\tau = dA^\mu/d\tau + \Gamma^\mu_{\nu\lambda}dx^\lambda/d\tau A^\nu$.

the scalar and vector potentials are arranged into a fourvector A_μ and charge density and current into a four-vector J_μ: $\partial^2 A_\mu = -4\pi J_\mu$. A fundamental difference among these two theories is the following: while in Maxwell theory once the charges/currents are given, all the potentials can be calculated, in General Relativity the metric enters on both sides of the equation. This means that we cannot specify $T_{\mu\nu}$ and then calculate $g_{\mu\nu}$, since also for constructing $T_{\mu\nu}$ we need the metric. Intuitively this means that gravitation influences the matter/energy distribution, which in turn modifies the gravitational field ("backreaction"). A famous quote summarizing the complexity of Einstein equation is *"Space tells matter how to move, matter tells space how to curve"* (from Gravitation, Misner, Thorne, Wheeler).

Chapter 4 | Cosmological Models

In this chapter, we will apply the Einstein equations to the Universe as a whole discussing some possible models and their consequences. The standard cosmological model presently favored by the available data will be presented.

4.1 The Cosmological Principle

The Einstein equations allow to calculate the space-time geometry if the distribution of matter is known. This task is in general quite complex if we would like to find analytical solutions. Symmetry principles can simplify the problem greatly. In cosmological applications, the so-called **Cosmological Principle**:

The Universe is spatially homogeneous and isotropic

is assumed. With "homogeneous" we mean that the universe is invariant under spatial translations, while with "isotropic" we assume that the universe looks the same in every direction, or that it is spherically symmetric around us. The principle implies that every observer at every point of the universe observes the same properties (a modern version of the Copernican principle which stated that we do not occupy any privileged position in the Universe.). Another way to state the principle, is that the universe can be foliated in space-like surfaces which are spherically symmetric

about any point.

Of course, on a small scale the universe is not homogeneous, since there are stars, galaxies and even clusters of galaxies. The Cosmological Principle is assumed to be approximately realized on scales larger than - say - 10^8 or 10^9 light-years where many galaxy clusters are contained. The Cosmological Principle is not a completely proved fact, but it is supported (besides from our philosophical beliefs) by observations of the matter distribution on the largest scales and by the existence of a rather homogeneous cosmic microwave background (which is of cosmological origin).

If you like to be more mathematically precise, then a space-time is said to be spatially homogeneous if there exist a one-parameter family of space-like hypersurfaces Σ_t foliating the space-time, such that for every t, P and Q in Σ_t, there exists an isometry[1] of the space-time metric which maps P into Q.

A space-time is said to be spatially isotropic at each point if there exists a congruence of time-like curves[2] ("observers"), with tangents denoted by v^α filling the space-time and satisfying the following property. Given any point P and any two unit spatial tangent vectors (orthogonal to v^α) s_1^α, s_2^α there is an isometry of the metric which leaves p and u^α at P fixed, but rotates s_1^α, s_2^α. This means that if isotropy is assumed, it is not possible to construct a preferred tangent vector orthogonal to v^α. Constructing a preferred vector is possible of the tangent space to Σ_t does not coincide to the tangent space containing the vectors orthogonal to u^α. This fact also shows that isotropy requires that Σ_t is orthogonal to u^α.

4.2 Metric of Homogeneous and Isotropic Space-times

The requirements of homogeneity and isotropy from the Cosmological Principle greatly constrain the class of metrics compatible with it.

The spatial isotropy requirement implies that $g_{0i} = 0$ ($i = 1, 2, 3$): in this way no direction is privileged (we have no "mixing" between directions

---

[1]An isometry is a transformation which preserves the lengths.

[2]A congruence of curves is the set of integral curves of a nowhere-vanishing vector field.

and time). We assume that all the observers on a spatial surface are able to measure the time ("cosmic time") in the same way, so we can choose $g_{00} = -1$ (a constant not depending on the space-time point). It turns out that all these requirements lead to a space-time with constant curvature (in particular such space-times are called *maximally symmetric*[3]). It can be showed that the requirements of homogeneity and isotropy in four dimensions lead to the following metric

$$- ds^2 = g(v)dv^2 + f(v)\left[d\mathbf{u}^2 + \frac{k(\mathbf{u}d\mathbf{u})^2}{1 - k\mathbf{u}^2}\right] \quad , \tag{4.1}$$

where v is a coordinate and $\mathbf{u}$ is a vector of three coordinates and f an unknown function of v only. This metric is clearly rotationally invariant as required. Introducing new "spherical coordinates" r, θ, ϕ for the "space" variables $\mathbf{u}$ and a time coordinate $t = \int -\sqrt{-g(v)}dv$ we obtain

$$ds^2 = dt^2 - R^2(t)\left[\frac{dr^2}{1 - kr^2} + r^2 d\theta^2 + r^2 \sin^2\theta d\phi^2\right] \quad , \tag{4.2}$$

where $R(t)$ function of time to be determined, and k is a constant representing the curvature. This metric is invariant under the rescaling

$$R \to \frac{R}{\lambda}$$
$$r \to \lambda r \tag{4.3}$$
$$k \to \frac{k}{\lambda^2} \quad ,$$

so choosing $\lambda = \sqrt{k}$, the curvature k can assume only the values $k = -1, 0, 1$. The function R(t) is usually normalized to its present value $a(t) = R(t)/R(0)$. and $a(t)$ is called **cosmic scale factor**.

Eq 4.2 is called the **Friedman-Lemaître-Robertson-Walker** metric (FLRW) and represents the most general homogeneous and isotropic space-time in four dimensions.

It interesting to look at the spatial geometry. If $k = 0$, space is flat and equivalent to a three-dimensional "plane". If $k = 1$, space is equivalent

---

[3]Mathematically, a maximally symmetric space M of dimension D is a space with a metric admitting D(D+1)/2 Killing vectors. Killing vectors form a vector field describing the infinitesimal isometric transformations in M.
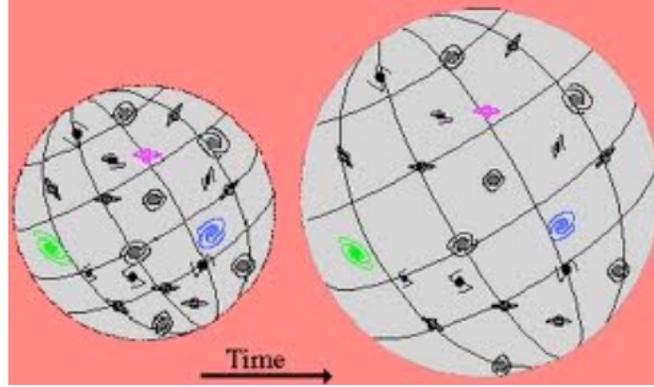
Figure 4.1: Comoving Coordinates

to the surface of a sphere with radius a(t). If $k = -1$, the space is hyperbolic. An important observation is that *the requirements of homogeneity and isotropy (the Cosmological Principle) allowed us to write the metric without solving or considering the Einstein Equations.*

4.3 Properties of the FLRW Universe

Considering the FLRW metric for an homogeneous and isotropic universe, we can directly calculate some relevant geometric quantities. In the $k = 1$ case, it is possible to calculate the spatial volume of the universe

$$^{(3)}V = \int \sqrt{|-^{(3)}g|}d^3x = a^3(t) \int_0^{2\pi} d\phi \int_0^{\pi} \sin\theta d\theta \int_o^{R_U} \frac{r^2 dr}{\sqrt{1-kr^2}} \quad,$$

(4.4)

where R_U is the "radius of the universe". R_U is finite only in the spherical $k = 1$ case, while in the other two cases $k = 0, -1$ it is infinite. Therefore, only in the $k = 1$ case we obtain a finite spatial volume $^{(3)}V = 2\pi^2 a^3(t)$ and $a(t)$ can be interpreted as a "radius". A $k = 1$ Universe is said to be **closed**, while the other two cases correspond to **open** universes.

Let's discuss now what the coordinates mean in the FLRW metric. These coordinates are called **comoving**, because the position of an observed does not change with respect to them. The idea is clearer looking at

Fig. 4.2: if the space-time is expanding (or contracting), the coordinates stretch according to the scale factor $a(t)$ and an object (say, a galaxy) keeps its position with respect to the axes.

An extremely simplified case is the one where $k = 0$ and $a(t)$=constant. Substituting the FLRW metric in the Einstein Equations gives $T_{\mu\nu} = 0$, since the components of the Ricci tensor $R_{\mu\nu}$ are zero (and therefore also the curvature scalar R). In this case, the equations of General Relativity describe an empty universe with a Lorentz (flat) metric. This case is trivial but it is a check that the Einstein Equations have an additional correct limiting case.

4.4 Friedmann Equations

Now it is time to use the FLRW metric in the Einstein equations. This will allow us to extract the exact form of the still unknown scale factor $a(t)$. The left side of the equation contains only the geometric quantities which can be calculated from the metric $g_{\mu\nu}$ which non-zero components are

$$
\begin{aligned}
g_{00} &= 1 \\
g_{11} &= -\frac{a^2(t)r^2}{1 - kr^2} \\
g_{22} &= -a^2(t)r^2 \\
g_{33} &= -a^2(t)r^2 \sin^2 \theta \quad .
\end{aligned}
\tag{4.5}
$$

From the metric tensor, we can directly calculate the Christoffel symbols with Eq. 2.45. The non-zero Christoffel symbols are

$$
\Gamma^0_{11} = \frac{\dot{a}a}{1 - kr^2} \quad ; \quad \Gamma^0_{22} = a\dot{a}r^2 \quad ; \quad \Gamma^0_{33} = a\dot{a}r^2 \sin^2 \theta
$$

$$
\Gamma^1_{01} = \Gamma^1_{10} = \Gamma^2_{02} = \Gamma^2_{20} = \Gamma^3_{03} = \Gamma^3_{30} = \frac{\dot{a}}{a}
$$

$$
\Gamma^1_{22} = -r(1 - kr^2) \quad ; \quad \Gamma^1_{33} = -r(1 - kr^2) \sin^2 \theta
\tag{4.6}
$$

$$
\Gamma^2_{12} = \Gamma^2_{21} = \Gamma^3_{13} = \Gamma^3_{31} = \frac{1}{r}
$$

$$
\Gamma^2_{33} = -\sin\theta\cos\theta \quad ; \quad \Gamma^3_{23} = \Gamma^3_{32} = \cot\theta \quad ,
$$

and the Riemann tensor with Eq. 2.58. Contracting the Riemann tensor we can calculate the Ricci tensor which non-zero components are

$$
\begin{aligned}
R_{00} &= -3\frac{\ddot{a}}{a} \\
R_{11} &= \frac{a\ddot{a} + 2\dot{a}^2 + 2k}{1 - kr^2} \\
R_{22} &= r^2(a\ddot{a} + 2\dot{a}^2 + 2kr) \\
R_{33} &= r^2(a\ddot{a} + 2\dot{a}^2 + 2kr)\sin^2\theta \quad ,
\end{aligned}
\tag{4.7}
$$

where the "dot" represents the total derivative with respect to the time ($\dot{x} = dx/dt$). Finally, contracting the Ricci tensor ($g^{\mu\nu}R_{\mu\nu}$) we obtain the curvature scalar

$$
R = \frac{6}{a^2}(a\ddot{a} + 2\dot{a}^2 + k) \quad .
\tag{4.8}
$$

Now we have all the ingredients for calculating the left side of the Einstein equations and we just need $T_{\mu\nu}$. Again the Cosmological Principle can help us: an homogeneous and isotropic distribution of matter is the one corresponding to a perfect fluid with density ρ and pressure P:

$$
T_{\mu\nu} = (\rho + P)v_\mu v_\nu - Pg_{\mu\nu} \quad ,
\tag{4.9}
$$

where v is the four-velocity vector with components $v = (1,0,0,0)$, since the "fluid" is at rest with respect to the comoving coordinates. Again, thinking at the fluid as made of point-like galaxies, they retain their place with respect to the coordinate axes: it is just the distance among them which changes through $a(t)$. Density and pressure can be time-dependent, but not space-dependent, otherwise this would be against the Cosmological Principle.
Consider now the energy-momentum conservation

$$
\nabla_\mu T^\mu_\nu = \partial_\mu T^\mu_\nu + \Gamma^\mu_{\mu\beta}T^\beta_\nu - \Gamma^\beta_{\mu\nu}T^\mu_\beta = 0 \quad ,
\tag{4.10}
$$

and the zero component (i.e. the energy component) of the above equation

$$
\partial_0\rho(t) + 3\frac{\dot{a}(t)}{a(t)}(\rho(t) + P(t)) = 0 \quad .
\tag{4.11}
$$

This equation expresses **energy conservation** in the FLRW Universe. Notice that for obtaining Eq. 4.11 we had to use the space-time geometry

through the covariant derivative and this brings us back to the complexity of the Einstein Equations, where curvature and energy/matter distribution are in general directly connected.

At his point we are stuck: we have the scale factor and other two unknowns in a single equation. The only way forward is to postulate a relationship between density and pressure, i.e. an **equation of state** for the energy/matter content of the universe. It can be showed that basically all the cosmologically relevant perfect fluids have an equation of state like $P = w\rho$ where w is a constant characteristic of the specific fluid. Substituting this generic equation of state in Eq. 4.11 we obtain

$$\frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{a}}{a} \quad , \tag{4.12}$$

which has solutions like

$$\rho(t) \propto a(t)^{-3(1+w)} \quad . \tag{4.13}$$

Now we have an equation that tells us how the density behaves while the universe expands or contracts. We still have to determine the dynamics of the scale factor.

Using the expressions for the Ricci tensor, the curvature scalar and the energy-momentum tensor, we can substitute them into the Einstein equations Eq. 3.17 finding (we leave $\Lambda = 0$ for the moment)

$$-3\frac{\ddot{a}}{a} = 4\pi G(\rho + 3P) \quad \text{for} \quad (\mu, \nu) = (0,0)$$

$$\frac{\ddot{a}}{a} + 2\left(\frac{\dot{a}}{a}\right)^2 + 2\frac{k}{a^2} = 4\pi G(\rho - P) \quad \text{for} \quad (\mu, \nu) = (i,j) \quad . \tag{4.14}$$

For the spatial components $i, j = 1, 2, 3$ there is only one equation as it should be, given the requirement of isotropy. Substituting the second derivative of $a(t)$ from the first of the Eq. 4.14 into the second we obtain the **Friedmann Equations**

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P)$$

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \quad . \tag{4.15}$$

The Friedmann equations are basically the Einstein equations for an homogeneous and isotropic universe filled with a perfect fluid. For fully solve them and obtain the function $a(t)$, we need to specify the equation of state (or - almost equivalently - the parameter w).

4.5 The Cosmological Parameters

The expansion rate (how much a(t) changes in units of a(t) itself) is called the **Hubble Parameter**

$$H(t) = \frac{\dot{a}(t)}{a(t)} \quad . \tag{4.16}$$

The definition of H makes sense, since the scale of a itself is not important: in the ratio, we obtain a meaningful measurable quantity. The today's value of this parameter is called H_0.

It is also useful to rewrite the energy conservation equation in the FLRW Universe (Eq. 4.11) using the Hubble parameter as

$$\dot{\rho} + 3H(\rho + P) = 0 \quad . \tag{4.17}$$

Another relevant quantity is the **deceleration parameter**

$$q = -\frac{a\ddot{a}}{\dot{a}^2} \quad , \tag{4.18}$$

which quantifies the rate of change of H. Substituting the Hubble parameter in the second Friedmann equation

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \Rightarrow \frac{8\pi G}{3H^2}\rho - 1 = \frac{k}{H^2 a^2} \quad . \tag{4.19}$$

Defining the **critical density** $\rho_c = \frac{3H^2}{8\pi G}$ (today's value $\approx 10^{-29} g/cm^3 \sim 1.05 \times 10^{-4} eV/cm^3$) and the **density parameter** $\Omega = \rho/\rho_c$ we have

$$\Omega - 1 = \frac{k}{H^2 a^2} \quad . \tag{4.20}$$

The density parameter is quite important since it determines the geometry (through the curvature k) of the universe:

$$\begin{array}{lll} \rho < \rho_c \Longleftrightarrow \Omega < 1 \Longleftrightarrow k = -1 & \text{(Open)} & \\ \rho = \rho_c \Longleftrightarrow \Omega = 1 \Longleftrightarrow k = 0 & \text{(Flat)} & \\ \rho > \rho_c \Longleftrightarrow \Omega > 1 \Longleftrightarrow k = 1 & \text{(Closed)} \end{array} \tag{4.21}$$

4.6 Cosmological Models

In order to treat all the possible cases, we will use the most general form of the Einstein equations, thus including the cosmological constant Λ. When the cosmological constant is taken into account, the Friedmann equations become

$$\frac{\ddot{a}}{a} = \dot{H} + H^2 = -\frac{4\pi G}{3}(\rho + 3P) + \frac{\Lambda}{3}$$

$$H^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3} \quad .$$

(4.22)

Regarding the equation of state, the most relevant cases are

$$w = 0 \Rightarrow \rho \sim \frac{1}{a^3} \quad \text{"Dust"}$$

$$w = 1/3 \Rightarrow \rho \sim \frac{1}{a^4} \quad \text{"Radiation"}$$

$$w = -1 \Rightarrow \rho \sim const. \quad \text{"Vacuum"} \quad .$$

(4.23)

In an universe filled with "dust", i.e. point-like massive particles, there is not interaction between them so the pressure is zero and therefore w=0. It is a standard physics result that for a volume filled with radiation only, $P = \rho/3$. The last case, where $P = -\rho$ (a sort of negative pressure) can arise in different models. It is called "vacuum energy" since this equation of state can arise from quantum field theory. We are now in the position to discuss some specific cosmological models.

4.6.1 The Einstein Universe

This cosmological model was first proposed by Einstein, who tried to obtain a static universe without expansion or contraction. This goal can be achieved only if $\Lambda \neq 0$. The static conditions are $\dot{a} = \ddot{a} = 0$ and from Eq. 4.22 we obtain

$$a = \sqrt{\frac{\Lambda}{k}} \Rightarrow k = 1$$

$$\rho = \frac{\Lambda}{4\pi G} \quad .$$

(4.24)

The previous equations describe a spherical universe with a constant radius (scale factor), a constant density and a non-zero cosmological constant. An unpleasant characteristic of this universe is its instability: a small perturbation would make it expand or contract. In this sense this model is not as static as its name may imply.

4.6.2 The Matter-dominated Universe

Suppose that the universe is uniformly filled only with non-interacting bodies. In this case, $P = 0$ (or $w = 0$), $\Lambda = 0$ and from Eq. 4.13 we have $\rho \cdot a(t)^3 = A$ with $A =$ constant. The second of the Friedman equations Eq. 4.22 becomes

$$\dot{a}^2 = \frac{8\pi G}{3}\frac{A}{a} - k \quad .\tag{4.25}$$

Introducing the *conformal time* η instead of the time t such that $d\eta/dt = 1/a(t)$ the last equation becomes

$$a'^2 = \frac{8\pi G}{3}Aa - ka^2 \quad .\tag{4.26}$$

The apex in a' indicates differentiation with respect to the conformal time. The last equation can be easily integrated. For example, choosing as initial condition $a(0) = 0$ we have

$$
\begin{aligned}
k = 1 &\Rightarrow a = \frac{4\pi GA}{3}(1 - \cos\eta) &;\quad t = \frac{4\pi GA}{3}(\eta - \sin\eta) \\
k = 0 &\Rightarrow a = \frac{2\pi GA}{3}\eta^2 &;\quad t = \frac{2\pi GA}{9}\eta^3 \\
k = -1 &\Rightarrow a = \frac{4\pi GA}{3}(\cosh\eta - 1) &;\quad t = \frac{4\pi GA}{3}(\sinh\eta - \eta) \quad .
\end{aligned}
\tag{4.27}
$$

In this model, a closed ($k = 1$) universe expands and eventually collapses again. Open universes ($k <= 1$) expand forever. In the boundary case $k = 0$, the expansion continues forever, but the expansion rate approaches zero for infinite times ($H \to 0$ when $t \to +\infty$).

4.6.3 The Radiation-dominated Universe

For radiation, we have already seen that $w = 1/3$, so $\rho = 3P$ and $\rho a^4 = A$ with A=constant. Setting $\Lambda = 0$, the second of the Friedman equations

Eq. 4.22 becomes

$$\dot{a}^2 = \frac{8\pi G}{3}\frac{A}{a^2} - k \quad .$$
(4.28)

Using the initial condition a(0)=0, the solutions are

$$k = 1 \Rightarrow a = \sqrt{2\sqrt{\frac{8\pi G A}{3a^2}}t - t^2}$$

$$k = 0 \Rightarrow a = \sqrt{2\sqrt{\frac{8\pi G A}{3a^2}}t}$$
(4.29)

$$k = -1 \Rightarrow a = \sqrt{2\sqrt{\frac{8\pi G A}{3a^2}}t + t^2} \quad .$$

As in the matter-dominated universe, the closed solution expands and then recollapses, while the other cases expand forever.

4.6.4 Vacuum-Dominated Universe

In the vacuum-dominated model, there is no matter present, so $P = \rho = 0$. In this case, only the cosmological constant plays a role ($\Lambda > 0$). A possible interpretation of this model comes from quantum field theory (QFT), where there are non-zero quantum fluctuations even in the vacuum where fields have zero average. QFT predicts a term analogous to the cosmological constant. In this case the second of the Friedman equations Eq. 4.22 reduces to

$$\dot{a}^2 = \frac{\Lambda a^2}{3} - k \quad ,$$
(4.30)

and the solutions are

$$k = 1 \Rightarrow a = \sqrt{\frac{3}{\Lambda}}\cosh\left(\sqrt{\frac{\Lambda}{3}}t\right)$$

$$k = 0 \Rightarrow a = \sqrt{\frac{3}{\Lambda}}\exp\left(\sqrt{\frac{\Lambda}{3}}t\right)$$
(4.31)

$$k = -1 \Rightarrow a = \sqrt{\frac{3}{\Lambda}}\sinh\left(\sqrt{\frac{\Lambda}{3}}t\right) \quad .$$

The $k = 1$ case is also known as *deSitter Universe*.

The $k = 0$ case predicts an exponential growth of the Universe and as we will see later on, this solution has relevance in the framework of *inflationary models*. If we assume $\Lambda < 0$, there are no solutions for $k = 0, 1$, while for $k = -1$

$$a = \sqrt{-\frac{3}{\Lambda}} \cos\left(\sqrt{-\frac{\Lambda}{3}} t\right) \quad . \tag{4.32}$$

It can be verified that for $\Lambda = 0$ this model reduces to the flat spacetime case with $k = 0$ and $a =$ constant.

4.6.5 Mixed Models

The models investigated so far contained only one type of matter/radiation. A more realistic model could contain different kinds of them in different proportions. In this case, the energy density will be the sum of the different components

$$\rho_{TOT}(a) = \sum_i \rho_i(a) = \rho_C \sum_i \Omega_i a^{-3(1+w_1)} \quad , \tag{4.33}$$

where ρ_c is the critical density and $\Omega_{TOT} = \rho_{TOT}/\rho_C$. Considering all the cases we treated so far and introducing the appropriate critical densities, the second Friedmann equation (Eq. 4.22) can be rewritten as

$$\frac{k}{a^2} = H^2(\Omega_{TOT} - 1) \quad . \tag{4.34}$$

Introducing in the last equations the density parameters observed today, together with the present Hubble parameter H_0 and the present scale a_0 we have

$$\frac{k}{a_0^2} = H_0(\Omega_m + \Omega_r + \Omega_\Lambda - 1) \quad , \tag{4.35}$$

where the density parameters describe pure matter (the "dust" universe), relativistic matter (radiation) and the effect of the cosmological constant.

Chapter 5 | Observational Cosmology and the ΛCDM Model

5.1 Cosmological Red Shift

Most of the information we gather on earth about the cosmos comes in the form of electromagnetic radiation. Here we would like to investigate how the universe's dynamics can affect a generic light signal of wavelength λ (or frequency $\nu = c/\lambda$). According to the Cosmological Principle, we can place the origin of our coordinate system where we would like to, so we choose $r = 0$ and for simplicity we forget about the angular coordinates θ and ϕ (equivalently, we can think of keeping them constant). An electromagnetic wave traveling from a distant star towards us (in the $-r$ direction) has the following equation of motion in an FLRW universe

$$d\tau^2 = dt^2 - a^2(t)\frac{dr^2}{1 - kr^2} = 0 \quad . \tag{5.1}$$

If the wave (say, a certain crest of the wave) leaves the star at time t_1 and reaches our telescope at time t_0, integrating the last equation we have

$$\int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_o^{r_1} \frac{dr}{\sqrt{1 - kr^2}} = \begin{cases} \sin^{-1} r_1 & k = 1 \\ r_1 & k = 0 \\ \sinh^{-1} r_1 & k = -1 \end{cases} \quad . \tag{5.2}$$

If our star belongs to a galaxy (as basically always it is the case), it has fixed coordinates, so $\int_{t_1}^{t_0} \frac{dt}{a(t)}$ is a time-independent function, as it is clear

from Eq. 5.2. This means that if we consider another crest of the electromagnetic wave leaving the star at a slight different time $t + \delta t$ we will find the same result as before for the integral $\int_{t_1+\delta t_1}^{t_0+\delta t_0} \frac{dt}{a(t)}$. Subtracting the two integrals and assuming that $a(t)$ does not vary much between the two crests,

$$\int_{t_1+\delta t_1}^{t_0+\delta t_0} \frac{dt}{a(t)} - \int_{t_1}^{t_0} \frac{dt}{a(t)} = \int_{\delta t_0}^{\delta t_1} \frac{dt}{a(t)} = 0 \Rightarrow \frac{\delta t_0}{a(t_0)} - \frac{\delta t_1}{a(t_1)} = 0 \qquad (5.3)$$

and therefore

$$\frac{\delta t_0}{a(t_0)} = \frac{\delta t_1}{a(t_1)} \qquad (5.4)$$

Frequencies and times are inversely proportional, so

$$\frac{\delta t_1}{\delta t_0} = \frac{a(t_1)}{a(t_0)} = \frac{\nu_0}{\nu_1} = \frac{\lambda_1}{\lambda_0} \qquad . \qquad (5.5)$$

We can now introduce the **red-shift parameter** and relate it to the scale factor

$$z = \frac{\lambda_0 - \lambda_1}{\lambda_1} = \frac{a(t_0)}{a(t_1)} - 1 \qquad . \qquad (5.6)$$

The wavelength λ_1 is the original one emitted from the star, as measured by a nearby observer, while λ_0 is what we will observe on the earth (at $r = 0$ of our coordinate system).
If $z > 0$, then $\lambda_0 > \lambda_1$: this is called **red-shift** and corresponds to an expanding universe.
If $z < 0$, then $\lambda_0 < \lambda_1$: this is called **blue-shift** and corresponds to a contracting universe.

5.2   Age of the Universe

We would like to use the Friedmann equations for determining how old the universe is as a function of the curvature and of the matter/energy content. The scale factor has no dimension and can be thought as the ratio of two lengths, or "radii": $R(t)$ at a time t, and $R(0)$ at $t = 0$ which is a reference time, for example "today". For $t = 0$ we have therefore a(0)=1. Multiplying and dividing the second Friedmann equation by the critical

density $\rho_c = 3H_0^2/8\pi G$ (H_0 is the Hubble parameter's value today) we have

$$H^2 = \frac{8\pi G}{3}\rho_c \left[\sum_i \frac{\rho_i}{\rho_c} + \frac{\rho_k}{\rho_c} \right] \quad , \tag{5.7}$$

where we have defined $\rho_k = k/(a^2)$ which looks like a density due to the global curvature of the universe.

Strictly speaking, this is just a formal analogy, and we should not think at ρ_k as a contribution to the energy density: this definition just helps in writing the equations in a more appealing way.

Remembering how the different energy densities scale ($\rho(t) \propto a^{-3(1+w)}$) and introducing the present-day density factors $\Omega_x^0 = \rho_x^0/\rho_c$, Eq. 5.7 becomes

$$H^2 = H_0^2 \left[\frac{\Omega_m^0}{a^3} + \frac{\Omega_r^0}{a^4} + \frac{\Omega_k^0}{a^2} + \Omega_\Lambda^0 \right] \quad . \tag{5.8}$$

For making better contact with measurements, we can introduce the red-shift parameter $z = a_0/a - 1$ ($a_0 = 1$ in our convention):

$$H^2 = H_0^2 \left[\Omega_m^0(1+z)^3 + \Omega_r^0(1+z)^4 + \Omega_k^0(1+z)^2 + \Omega_\Lambda^0 \right] \quad . \tag{5.9}$$

The last equation can be further simplified with the following steps

- We can rewrite $H = \dot{a}/a$ using the red-shift parameter, obtaining $H = -\frac{1}{1+z}\frac{dz}{dt}$.

- Approximate $\Omega_r^0 \approx 0$ since the era when the universe was radiation-dominated was much shorter than the matter domination and vacuum-energy domination.

- Remember that the sum of all the density parameters is equal to 1 by definition, so we set $\Omega_k^0 = 1 - \Omega_m^0 - \Omega_\Lambda^0$.

obtaining after some algebra

$$\Delta t = \frac{1}{H_0} \int_0^z \frac{dz'}{1+z'} \frac{1}{\sqrt{(1+\Omega_m^0 z')(1+z')^2 - z'(2+z')\Omega_\Lambda^0}} \quad . \tag{5.10}$$

The integral over the red-shift factor extends from today ($z = 0$) to some era when the factor was equal to z. For obtaining the total age of the

universe, we have to extend the integration to $z \to \infty$. Since the integral in Eq. 5.10 is of order one, a quick estimate of the age A of the universe is $A \sim 1/H_0 \approx 14$ Gyr.

The exact calculation can be carried out only numerically, but it is interesting to investigate the simple analytic result were the universe is flat ($k = 0 \Rightarrow \Omega_\Lambda^0 = 1 - \Omega_m^0$) and just dust-filled ($\Omega_\Lambda^0 = 0$):

$$A = \frac{1}{H_0} \int_0^\infty \frac{dz'}{(1+z')^{5/2}} = \frac{2}{3H_0} \sim 10 \text{ Gyr} \quad . \tag{5.11}$$

5.3 Measurement of Cosmological Distances

Measuring distances at the cosmic scale requires different methods, each of which is appropriate in a certain range. If the various methods partially overlap, it is possible to construct a **distance ladder** which calibration can be checked.

The distance has to be put in correspondence to the cosmological parameters and variables we studied so far.

Considering a light source at a certain distance d from us emitting a flux Φ (power/surface), it scales with the distance as

$$\Phi = \frac{L}{4\pi d^2} = \frac{L}{4\pi(a_0 r)^2} = \frac{L}{4\pi a_0^2 r^2 (1+z)^2} \quad , \tag{5.12}$$

where we introduced the actual distance as a function of the FLRW variable r (the "detector" is at $r = 0$). In the last step, we took into account the following relativistic effects. The flux has dimensions Energy/(time×surface): the energy has to be red-shifted (reduced) by a factor $a/a_0 = 1 + z$, while the time interval at emission that we measure has to be dilatated by $a_0/a = 1 + z$. Thus, the flux has to be multiplied by an additional $1/(1+z)^2$ factor.

We can now introduce the so-called **luminosity distance** d_L

$$d_L = a_0 r (1+z) = \sqrt{\frac{L}{4\pi\Phi}} \quad . \tag{5.13}$$

Knowing the absolute luminosity of an astronomical object and measuring the flux on earth, the distance can be calculated. Now we would like

to know how the distance depends from the cosmological parameters. Since we are considering the propagation of light rays ($d\tau^2 = 0$), disregarding the angular FLRW variables (i.e. looking only along the line of sight) we have

$$dt^2 = a\frac{dr^2}{1 - kr^2} \Rightarrow (1+z)dt = \frac{a_0}{a}dt = a_0\frac{dr}{\sqrt{1 - kr^2}} \quad . \tag{5.14}$$

Using Eq. 5.10 we have

$$\frac{1}{H_0}\int_0^z \frac{dz'}{\sqrt{(1 + \Omega_m^0 z')(1 + z')^2 - z'(2 + z')\Omega_\Lambda^0}} = \frac{1}{H_0}\int_0^z dz'F(z') =$$

$$a_0\int_0^r \frac{dr'}{\sqrt{1 - kr'^2}} = a_0 S_k^{-1}(r) = a_0 \begin{cases} \sin^{-1} r \ (k = 1) \\ r \quad (k = 0) \\ \sinh^{-1} r \ (k = -1) \end{cases} \tag{5.15}$$

After choosing k, the last equation can be inverted for extracting r as a function of H_0 and the density parameters. The result for r can finally be substituted in Eq. 5.13 obtaining the luminosity distance

$$d_L(z, H_0, \Omega_m^0, \Omega_\Lambda^0) = \frac{1 + z}{H_0\sqrt{\Omega_k^0}}S_k\left(\sqrt{\Omega_k^0}\int_0^z dz'F(z')\right) \quad . \tag{5.16}$$

Measuring the flux and knowing the absolute luminosity, the density and Hubble parameters can be determined.

5.4 Cosmography

With the word "Cosmography" we refer to the geography of the Universe, in the sense that we deal only with the measurement of distances without making any hypotheses about the matter content. Given a set of astronomical distance and red-shift data, we can ask what we can say about the scale factor $a(t)$ without using the Einstein equations, or, equivalently, without postulating a specific form of the energy-momentum tensor. Such a question is interesting, since there is a large number of models for $T_{\mu\nu}$, developed for explaining the data.
The idea is to expand the scale factor in power series and then relate it to

observable quantities (see e.g. M.Visser, gr-qc/0411131).
Defining the following expansion coefficients

$$H(t) = \frac{1}{a}\frac{da}{dt} \quad \text{Hubble Parameter}$$

$$q(t) = -\frac{1}{a}\frac{d^2a}{dt^2} \Big/ \left(\frac{1}{a}\frac{da}{dt}\right)^2 \quad \text{Deceleration}$$

$$j(t) = \frac{1}{a}\frac{d^3a}{dt^3} \Big/ \left(\frac{1}{a}\frac{da}{dt}\right)^3 \quad \text{"Jerk"}$$

$$s(t) = \frac{1}{a}\frac{d^4a}{dt^4} \Big/ \left(\frac{1}{a}\frac{da}{dt}\right)^4 \quad \text{"Snap" (or Jounce)} \tag{5.17}$$

$$c(t) = \frac{1}{a}\frac{d^5a}{dt^5} \Big/ \left(\frac{1}{a}\frac{da}{dt}\right)^5 \quad \text{"Crackle"}$$

$$p(t) = \frac{1}{a}\frac{d^6a}{dt^6} \Big/ \left(\frac{1}{a}\frac{da}{dt}\right)^6 \quad \text{"Pop"}$$

...

we can write the scale factor as

$$a(t) = a_0 \left[1 + H_0(t - t_0) - \frac{1}{2}q_0 H_0^2(t - t_0)^2 + \frac{1}{6}j_0 H_0^3(t - t_0)^3 + ...\right] \tag{5.18}$$

The red-shift is (here we keep $c \neq 1$)

$$1 + z = \frac{a(t_0)}{a(t_0 - D/c)} \quad , \tag{5.19}$$

where D/c is the time at which the light signal was emitted and D is the distance traveled since the emission. Substituting the expansion in Eq. 5.18 in Eq. 5.19 and expanding it in the parameter $x = H_0 D/c$ with

$$\frac{1}{1 + x + ax^2/2 + bx^3/6 + cx^4/24 + ...} =$$
$$1 - x + (1 - \frac{a}{2})x^2 + (a - \frac{b}{6} - 1)x^3 \tag{5.20}$$
$$+ \frac{1}{24}(6a^2 - 36a + 8b - c + 24)x^4 + \mathcal{O}(x^5) \quad ,$$

we obtain

$$z(D) = \frac{H_0 D}{c} + \frac{2 + q_0}{2}\frac{H_0^2 D^2}{c^2} + \frac{6(1 + q_0) + j_0}{6}\frac{H_0^3 D^3}{c^3}$$
$$+ \frac{24 - s_0 + 8j_0 + 36q_0 + 6q_0^2}{24}\frac{H_0^4 D^4}{c^4} + \dots \quad ,$$

(5.21)

and inverting it we can write the distance as a function of the red-shift

$$D(z) = \frac{cz}{H_0}\left[1 - \left(1 + \frac{q_0}{2}\right)z + \left(1 + q_0 + \frac{q_0^2}{2} - \frac{j_0}{6}\right)z^2\right.$$
$$\left. - \left(1 + \frac{3}{2}q_0(1 + q_0) + \frac{5}{8}q_0^3 - \frac{1}{2}j_0\frac{5}{12}q_0 j_0 - \frac{s_0}{24}\right)z^3 + \dots\right]$$

(5.22)

What is usually measured, is the luminosity distance defined in Sec. 5.3, so we have to convert the physical distance D to d_L. Using Eq. 5.12, the luminosity distance can be written as

$$d_L = \frac{a_0^2 r_0}{a(t_0 - D/c)} \quad .$$

(5.23)

where r_0 is the "radius" at the detection point (r=0 is the emission point). The radial coordinate r_0 is given by inverting Eq. 5.2 for the three possible values of the curvature k:

$$r_0(D) = \begin{cases} \sin\int_{t_0-D/c}^{t_0} A(t)dt & k = 1 \\ \int_{t_0-D/c}^{t_0} A(t)dt & k = 0 \\ \sinh\int_{t_0-D/c}^{t_0} A(t)dt & k = -1 \end{cases}$$

(5.24)

with $A(t) = cdt/a(t)$. The last expression can be expanded in Taylor series with the integral as parameter. Remembering that $\sin x \sim x - \frac{1}{3!}x^3 + \frac{1}{5!}x^5 - \dots$ and $\sinh x \sim x + \frac{1}{3!}x^3 + \frac{1}{5!}x^5 + \dots$ we can write

$$r_0(D) = \left(\int_{t_0-D/c}^{t_0} A(t)dt\right) - \frac{k}{3!}\left(\int_{t_0-D/c}^{t_0} A(t)dt\right)^3 + \dots \quad .$$

(5.25)

We can substitute the same expansion which lead to Eq. 5.21 (multiplying and dividing by a_0) and obtain

$$\int_{t_0-D/c}^{t_0} A(t)dt = \frac{D}{a_0}\left[1 + \frac{1}{2}\frac{H_0 D}{c} + \frac{2+q_0}{6}\frac{H_0^2 D^2}{c^2} + ...\right] \quad . \tag{5.26}$$

Combining Eq. 5.22, 5.23, 5.25, and 5.26 we can express the luminosity distance as a function of the redshift

$$d_L(z) = \frac{cz}{H_0}\left[1 + \frac{1}{2}(1-q_0)z - \frac{1}{6}\left(1 - q_0 - 3q_0^2 + j_0 + \frac{kc^2}{H_0^2 a_0^2}\right)z^2\right.$$

$$\left. + \frac{1}{24}\left(2 - 2q_0 - 15q_0^2 - 15q_0^3 + 5j_0 + 10q_0 j_0 + s_0 + \frac{2kc^2(1+3q_0)}{H_0^2 a_0^2}z^3\right)\right.$$

$$+ ... \quad . \tag{5.27}$$

The last expression relates two measurable quantities (z and d_L) without specifying the energy-momentum tensor (the matter-energy content of the universe). Fitting distance data with Eq. 5.27 allows the estimation of the Hubble parameter, the deceleration parameter and if the dataset is precise enough, even higher-order terms (j_0,...). Note that if we keep only the term linear in z, we have the original Hubble law $d_L \sim H_0 z$, where the red-shift is proportional to the distance.

Accurate measurements can therefore (at least in principle) help reconstructing the scale factor function $a(t)$. This information can be fed back into the Friedmann equations for extracting the equation of state of the cosmological fluid, in a sort of inverse process with respect to what is commonly done (postulate an equation of state and then extract $a(t)$). In particular, the present value of the parameter $w = P/\rho$ is connected to the deceleration parameter q_0, while the slope parameter $dP/d\rho$ is connected to the jerk and so on.

5.5   The ΛCDM Model

With ΛCDM Model, we intend the current Standard Cosmological Model which best fits the available observations. An important source of experimental information is the Cosmic Microwave Background (CMB), which

we will discuss later on. Since up to now we discussed cosmic distances, we mention here the most important experimental method which lead to the estimate of the density parameters.

5.5.1 Type Ia Supernovae

As we have seen, the calculation of the luminosity distance requires the knowledge of the absolute luminosity of an object. Astronomical objects for which the absolute luminosity is fairly well known are called **standard candles**. One of the most important standard candles are the Type Ia supernovae. These stars are actually belonging to a binary system where one of the two companions is a carbon-oxygen white dwarf. The white dwarf, thanks to its intense gravitational field, accretes matter from the companion star until a runaway nuclear reaction makes it turning into a supernova. The supernova explosion is extremely bright and visible from large distances. The important point is that the efficiency of the supernova mechanism is dominated by the temperature of the core and therefore ultimately from the mass of the star. It turns out that the Type Ia supernovae can happen only in a restricted mass range, so they deliver always a very similar light curve. That's why they are useful as standard candles.

5.5.2 Accelerated Expansion

Fitting the last result of Sec. 5.3 yields estimates for Ω_m^0 and Ω_Λ^0. The striking result is that if we take into account the CMB result $\Omega_m^0 + \Omega_\Lambda^0 \approx 1$, Type Ia supernova data suggests

$$\begin{aligned} \Omega_m^0 &\sim 0.3 \\ \Omega_\Lambda^0 &\sim 0.7 \quad , \end{aligned} \tag{5.28}$$

leading to an universe dominated by the cosmological constant term, where matter has a quite smaller weight. In particular, Type Ia supernova surveys suggest a negative deceleration parameter q_0, which means that the universe is accelerating its expansion.

5.5.3 The ΛCDM Model

The ΛCDM is the "concordance model" emerged from data coming from different experimental approaches. The Λ refers to the accelerated expansion driven by a term analogous to the cosmological constant. CDM means "Cold Dark Matter" and refers to a large component of Ω_m composed by some kind of matter we did not have identified yet. The normal observed "barionic" matter (Ω_b) is the smallest component. The ΛCDM model can be summarized by the following values of the density parameters

$$\Omega_b^0 \sim 0.05$$
$$\Omega_{CDM}^0 \sim 0.25 \tag{5.29}$$
$$\Omega_\Lambda^0 \sim 0.7 \quad,$$

leading to the puzzling conclusion that most of the universe is filled with yet unidentified forms of matter/energy, while the matter described by the Standard Model of particle physics makes up just 5% of its content. CMB, Dark matter, and Dark Energy will be discussed in the next chapters.

Chapter 6 | Inflation

In this chapter, we will introduce the idea of Inflation. Inflation was put forward in order to solve various problems of the standard Big-Bang cosmological scenario.

Some of these "problems" are the so-called horizon and flatness problems, which we will describe first (the fact that these are real problems is matter of still open debate). Besides giving a solution to the aforementioned problems (and others connected to specific theories like grand-unified theories (GUTs)), inflation has other advantages. For example, one of the successes of the inflationary paradigm, is that it provides a mechanism for generating cosmological fluctuations on microscopic scales which, during cosmological evolution, evolve into the perturbations observed in the large-scale structure of the universe and the anisotropies in cosmic microwave background (CMB).

6.1 Conformal Time and the Hubble Horizon

As we know already, the Hubble constant is defined as $H = \dot{a}/a$ which can be read as the ratio between velocity and space. Considering a particle traveling at the speed of light ($c = 1$), we can define the Hubble radius $R_H(t) = 1/H(t)$. In an expanding Universe, two light sources separated by a distance greater than the Hubble radius, will never be able to communicate. Today's Hubble radius R_0 is estimated to be $1/H_0 \sim 4.1$ Gpc. Another useful definition is the *comoving Hubble radius*

$$R_{cH} = \frac{1}{a(t)H(t)} \quad . \tag{6.1}$$

In the previous chapter, we defined the *conformal time* via $d\eta = dt/a(t)$. Using this new variable, the FLRW metric can be rewritten like

$$ds^2 = a^2(\eta)\left[-d\eta^2 + d\chi^2 + S_k^2(\chi)d\Omega^2\right] \quad , \tag{6.2}$$

where Ω collects the two "angular" variables, $d\chi^2 = dr^2/(1 - kr^2)$ and

$$S_k(\chi) = \begin{cases} \sinh \chi & k = -1 \\ \chi & k = 0 \\ \sin \chi & k = 1 \end{cases} \tag{6.3}$$

The interesting fact about this metric is that besides the *conformal factor* $a^2(\eta)$, it looks like a Minkowski space-time in Cartesian coordinates

$$ds^2 = a^2(\eta)\eta_{\mu\nu}dx^\mu dx^\nu \quad . \tag{6.4}$$

In conformal coordinates, a light ray propagates along $45°$ lines in a space-conformal time plot: $\chi = \pm\eta + C$. Using the normal time in a curved space-time, light would have followed curves instead of straight lines: that's the advantage of introducing the conformal time.
The maximum distance a light ray can travel is therefore

$$\Delta\eta = r_{max} = \int_0^t \frac{dt'}{a(t')} = \int_{a_0}^{a_1} \frac{da}{a\dot{a}} = \int_{\ln a_0}^{\ln a_1} \frac{1}{aH}d\ln a = \int_{\ln a_0}^{\ln a_1} R_{cH}d\ln a \quad . \tag{6.5}$$

We have therefore rewritten the "horizon" r_{max} in terms of the comoving Hubble radius, or, in other words, the elapsed conformal time depends from the evolution of the Hubble radius (which in turn is governed by the Friedmann equations).

6.2 The Particle Horizon "Problem"

The problem is based on trying to find an explanation to the extremely high present homogeneity of the universe, even among regions which are causally disconnected. Two regions are said to be causally disconnected if even light did not have the time to travel from one region to the other while the universe is expanding. If no causal connection can be established among these regions, how is it possible that the universe is so homogeneous everywhere? Light started to travel free through the universe

at the **recombination time**, when the Universe was about 3×10^5 years old and expanding. This light (which today we observe as the cosmic microwave background) did not had the time to travel to every possible point of the universe, according to the evolution we can calculate with GR, but still, the universe is quite homogeneous. This observation might rise the question: how is it possible that about the same initial conditions were set everywhere in the same way, so that causally disconnected regions evolved in a very similar way? Like all the fine-tuning "problems", the fact that it is really a problem we have to solve, it is matter of debate. Let's see now more mathematically how such causally disconnected regions arise. A light ray propagates with zero proper time, so considering only radial rays $d\tau^2 = dt^2 - a^2(t)dr^2 = 0$ and therefore integrating by separation of variables

$$\int_0^{r_{max}} dr = \int_0^t \frac{dt'}{a(t')} \quad , \tag{6.6}$$

where r_{max} is the maximum r traveled by the light ray in a time t starting from $t = 0$. Let's consider now the time where light started traveling and the universe was radiation-dominated. From Eq. 4.29 we have $a \sim \sqrt{t}$. The **horizon** D_H is the distance traveled while the universe is expanding, so it is r times the scale factor

$$D_H = a(t) \int_0^{r_{max}} dr = a(t) \int_0^t \frac{dt'}{a(t')} = \sqrt{t} \int_0^t \frac{dt'}{\sqrt{t'}} = 2t \quad . \tag{6.7}$$

We obtained the following result: if we go back in time, the horizon shrinks proportionally to the time, while the scale factor (the distances) shrinks like $\sqrt{t}$. This means that the horizon is getting smaller faster than the dimension of the universe. In other words, there were portions of universe which could not "communicate" with each other, but still, the CMB looks with good accuracy the same everywhere. This is the essence of the horizon problem.

Now let's look at the Horizon Problem from another point of view, making use of the conformal time and the comoving Hubble radius discussed previously.

First, we rewrite things as a function of the red-shift $a = 1/(1+z)$ $(a_0 = 1)$

$$\frac{da}{a} = -\frac{dz}{1+z} \quad . \tag{6.8}$$

This allows to rewrite the the horizon as ($a_0 = 1$)

$$d(z) = \int_{z_1}^{z_2} \frac{dz'}{H(z')} \quad . \tag{6.9}$$

We would like to calculate the angle subtended by the horizon at the recombination time, which can be approximated by the ratio between the comoving particle horizon d_h at recombination and the comoving angular diameter distance from us d_0 (redshift z=0) to recombination (z$\sim$1090):

$$\theta_h = \frac{d_h}{d_0} = \frac{\eta_{rec} - \eta_\infty}{\eta_0 - \eta_{rec}} = \frac{\int_{z_{rec}}^{\infty} \frac{dz}{H(z)}}{\int_0^{z_{rec}} \frac{dz}{H(z)}} \quad , \tag{6.10}$$

where we introduced the comoving distance between two red-shifts. H(z) can be replaced with Eq. 5.8 with the approximate measured values of the density parameters: $\Omega_m \sim 0.3$, $\Omega_\Lambda = 1 - \Omega_m$, and $\Omega_r \sim 0$. The integrals can be numerically evaluated, yielding the interesting result $\theta_h \sim 1°$. This means that if we look at the sky, regions separated by about one degree were causally disconnected at recombination time, but now they look quite similar. This is another way to state the horizon problem.

6.3 The Flatness "Problem"

The so-called Flatness Problem is a fine-tuning problem of the Big-Bang theory based on the FLWR equations. Some people do not regard this as a real problem to solve, since we do not know the real probability distribution for possible initial conditions of the universe. After this disclaimer, let's see what this problem is.
Rearranging the Friedmann Eq. 4.15 introducing the critical density and the density parameter we can obtain

$$\left(\frac{1}{\Omega} - 1 \right) \rho_c a^2 = -\frac{3k}{8\pi G} \quad . \tag{6.11}$$

Considering only matter and radiation as content of the universe, during the expansion the density $\rho(t)$ drops faster that the growth of the scale factor $a(t)$. This means that since the right-hand side of Eq. 6.11 is a constant and ρa^2 decreases, $1/\Omega - 1$ increases. Taking as $t = 0$ the Planck

time ($t_P = \sqrt{\frac{\hbar G}{c^5}} \sim 0.5 \times 10^{-43}$s), ρa^2 should have dropped by a factor $\sim 10^{60}$ during the cosmic history until today. In turn, $1/\Omega - 1$ should have increased by the same factor.

Today we measure (for example with SN1a surveys and CMB measurements) $\Omega_0 \sim 1$ and therefore $1/(\Omega - 1) \sim 0$. Through Eq. 6.11 this means that the universe is nearly flat and very close to its critical density. This translates to an extremely tiny value ($\sim 10^{-62}$) for $|\Omega - 1|$ at the Planck time. This is the essence of the flatness "problem": in order to observe an almost flat universe today, we have to "fine-tune" the density to the critical density to high accuracy. Said in other words: if today we observe a flat universe, in the distant past, it had to be even flatter (by the huge $\sim 10^{60}$ factor).

Again, this might be not a problem at all, since we do not know how natural a similar choice for the initial condition is. At any rate, the introduction of an inflationary phase in the early universe removes the need of this fine-tuning "problem" predicting a nearly flat universe.

6.4 Dilution of Relics

For completeness, we mention also the so-called magnetic-monopole problem. Some theories beyond the SM (enlarging it with additional symmetries) predict the existence of magnetic monopoles (a sort of analog to the electric charges for the magnetic field). P.A.M. Dirac was the first putting this idea forward, deriving the quantization condition

$$q_m q_e = \frac{n}{2} \quad , \tag{6.12}$$

where n is an integer number. Theories predicting the existence of magnetic monopoles lead to an overabundance of such particles and inflation is a generic mechanism able to dilute them to the today's very small (if any) abundance. The same idea can work for other exotic particle species predicted theoretically.

6.5 Inflation

Generically, an inflationary phase of the universe is a phase where there is accelerated expansion, or $\ddot{a} > 0$.

From the Friedmann equation

$$\dot{H} + H^2 = \frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) \tag{6.13}$$

the inflation condition can be translated to $\rho + 3P < 0$. Thus, as can be easily verified, inflation can happen under the equivalent conditions

$$\ddot{a} > 0 \iff \frac{\dot{H}}{H^2} < 1 \iff \rho + 3P < 0 \iff \frac{d}{dt}\left(\frac{1}{aH}\right) < 0 \quad . \tag{6.14}$$

The last condition is quite interesting: inflation corresponds to a shrinking of the Hubble radius. This is exactly what it is needed for fixing the horizon problem, since otherwise the expansion of the Hubble radius leads to causally disconnected regions.
The condition $\rho + 3P < 0$ instead tells that we need a sort of negative pressure $P < -\rho/3$.

6.5.1 Solution to the Horizon Problem

Considering the version of the Friedmann equation obtained in Eq. 5.8 for just one generic matter/energy component Ω with equation of state $P = w\rho$ we have

$$H = \frac{\dot{a}}{a} = H_0\sqrt{\Omega}a^{-\frac{3}{2}(1+w)} \quad . \tag{6.15}$$

If $w \neq -1$ the solution is $a(t) \propto t^{2/3(1+w)}$. If $w = -1$, then $a(t) = e^{Ht}$. These solutions, using the conformal time $(d\eta = dt/a)$ are

$$a(\eta) \propto \begin{cases} \eta^{\frac{2}{1+3w}} & w \neq -1 \\ -\dfrac{1}{\eta} & w = -1 \end{cases} \tag{6.16}$$

This means that $\eta \propto \frac{2}{(1+3w)}a^{\frac{1}{2}(1+3w)}$ and since during inflation $1 + 3w < 0$, η goes to $-\infty$ if a goes to zero. Therefore, the Big Bang is pushed to negative conformal times. This in turn means that between the initial singularity and the decoupling time (when light started to be free to travel in the Universe) there is much more time than previously thought. In other words, the light cones which were separated at the decoupling time have

now time to merge.

This is the effect of the decreasing comoving horizon during the inflationary phase which solves the horizon problem. The comoving Hubble horizon

$$R_{cH} = \frac{1}{aH} = \frac{1}{H_0} a^{\frac{1}{2}(1+3w)} \quad, \tag{6.17}$$

in the inflationary case $1 + 3w < 0$ shrinks, while for normal "fluids", where $1 + 3w > 0$, it always grows.

6.5.2 Solution to the Flatness Problem

From the Friedman equation $\Omega - 1 = \frac{k}{a^2 H^2}$ we see that the density parameter is connected to the comoving Hubble radius. Since the radius decreases during inflation, the Universe is driven towards flatness.

A very nice way to see this is combining the two Friedmann equations with the inflationary condition $\dot{H} \ll H^2$, $P = w\rho$ and the definition of the density parameter, obtaining

$$\frac{d\Omega}{d(\ln a)} = (1 + 3w)\Omega(\Omega - 1) \quad. \tag{6.18}$$

Performing a stability analysis of the previous equation, it can be seen that $\Omega = 1$ is an attractor during inflation $(1 + 3w < 0)$, while it represents an unstable fixed point otherwise $(1 + 3w > 0)$.

Thus inflation produces naturally a flat universe, provided that the inflationary phase lasts for enough time.

A caveat to this discussion is the following. Inflation does not change the curvature k of the Universe. For example, if $k > 0$, this will not be modified by the inflationary expansion phase. It is the huge expansion of the Universe which reduces the curvature radius and makes space-time look almost flat, but the global curvature remains unchanged.

6.6 Scalar Fields

Before describing some inflationary models, we do a digression into the quantum field theory of a scalar field $\phi(x)$, where x is a space-time point.

The action of a scalar field in a generic space-time is

$$S = \int dx^4 \sqrt{-g}\, \mathcal{L} = \int dx^4 \sqrt{-g} \left[\frac{1}{2} g^{\mu\nu} \partial_\mu \phi \partial_\nu \phi - V(\phi) \right] \quad . \tag{6.19}$$

The field equations (the "equations of motion") can be obtained from the Euler-Lagrange equations for continuous systems

$$\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} - \frac{\partial \mathcal{L}}{\partial \phi} = 0 \quad , \tag{6.20}$$

and the result is

$$\partial^\mu \partial_\mu \phi - \frac{dV}{d\phi} = 0 \quad . \tag{6.21}$$

For V=0, the usual wave equation is recovered: $\partial^\mu \partial_\mu \phi = -\ddot{\phi} + \nabla^2 \phi = 0$. The energy-momentum tensor can be obtained (Nöther's theorem) with

$$T^{\mu\nu} = -\partial_\mu \frac{\partial \mathcal{L}}{\partial(\partial_\mu \phi)} \partial^\nu \phi + g^{\mu\nu} \mathcal{L} \quad . \tag{6.22}$$

The energy density is the ($\mu = 0, \nu = 0$) component of T, while the average pressure $\langle P \rangle$ is the average of the three spacial diagonal components ($\mu = \nu = 1, 2, 3$):

$$\rho = T^{00} = \frac{1}{2}\dot{\phi}^2 + \frac{1}{2}\nabla\phi^2 + V(\phi) \quad ,$$
$$P = \frac{1}{3}(P^{11} + P^{22} + P^{33}) = \frac{1}{6}\dot{\phi}^2 - \frac{1}{2}\nabla\phi^2 + V(\phi) \quad . \tag{6.23}$$

In the case of spacial homogeneity and isotropy, spacial gradients vanish and the two last equations are related by $P = -\rho$. We discover in this way that the equation of state of an homogeneous scalar field is characterized by $w = -1$ and can in principle have the correct characteristics, if dominating over other forms of matter, to drive an inflationary expansion.

6.7   Old Inflation

The first inflationary model was proposed by A. Guth at the beginning of the 80s. In this work, it was pointed out how an inflationary phase

could resolve the horizon and flatness problems. Similar observations were also made by A. Starobinski, who also predicted the generation of gravitational waves in the early universe. Mukhanov and Chibisiv (1981) were the first realizing that zero-point fluctuations in an initial vacuum state would be amplified by the expansion phase, leading to density perturbation which will act as seeds for galaxy formation.

The general idea was the one discussed before, i.e. assume a sort of negative pressure leading to an equation of state with $1 + 3w < 0$. The way this was realized is the following. Guth's idea was that the Universe began in a state characterized by higher symmetry, called a *false vacuum*, since it was not the lowest energy state allowed by the potential of a certain quantum field(s). He suggested that the supercooling of a first-order phase transition (of e.g. GUT models) can drive the inflationary phase. In this idea, the density ρ of the Universe is dominated by the difference in energy density between the false-vacuum and true-vacuum phases $\Delta\rho \sim T_{GUT}^4$. The energy density $\Delta\rho$ can then act as an effective cosmological constant leading to accelerated expansion.

Using a condensed matter analogy, the false vacuum corresponds to a superheated fluid, while the "true" vacuum is analogous to the vapor phase. The thermodynamic fluctuations were in this case the quantum fluctuations. During the transition to the true vacuum, bubble nucleation happens (like vapor bubbles in the superheated fluid).

The problem with this mechanism, is that the huge rate of expansion of the universe dominates the rate of production and growth of bubbles; the bubbles never merge to fully complete the transition. This means that the inflationary phase might last too long.

Another problem is that the collision of the nucleated bubbles can lead to large anisotropies.

6.8 New Inflation

The "new" inflation paradigm is based on a mechanism where the Universe cannot escape reaching the true vacuum. These new ideas were first developed by A. Linde (1982) and independently by A. Albrecht and P.J. Steinhardt (1982). The new inflationary models are based on a slowly evolving scalar field (the "**inflaton**").

A typical potential choosen for the scalar field has the the polynomial

form

$$V = \frac{1}{2}m^2\phi^2 + \frac{\lambda_\phi}{4}\phi^4 \quad . \tag{6.24}$$

The field theory defined by the last two equations is renormalizable and describes a self-interacting scalar field. Choosing for $g_{\mu\nu}$ the FLRW metric, the equation of motion for the field is [1]

$$\ddot{\phi} + 3H\dot{\phi} - \frac{\nabla^2\phi}{a^2} + \frac{dV}{d\phi} = 0 \quad . \tag{6.25}$$

The energy-momentum tensor is

$$T^{\mu\nu} = \frac{2}{\sqrt{-g}}\frac{\delta S}{\delta g_{\mu\nu}} = \partial^\mu\partial^\nu\phi - g^{\mu\nu}\left[\frac{1}{2}(\partial\phi)^2 - V(\phi)\right] \quad . \tag{6.26}$$

If ϕ varies slowly as a function of the space-time coordinates, then we can neglect the derivatives and the energy-momentum tensor is approximately

$$T^{\mu\nu} \approx g^{\mu\nu}V(\phi) \quad , \tag{6.27}$$

which resembles a cosmological constant term. After substitution into the Einstein (Friedmann in this case) equations, it leads to accelerated expansion, as seen in the vacuum-dominated Universe solution (Sec. 4.6.4).

A clear difference between the vacuum-dominated universe and the case at hand, is that in the former case the term in the energy-momentum tensor is really constant, while in the latter it varies slowly, moving towards the $dV/d\phi = 0$ equilibrium point. Usually the potential is defined such that at $dV/d\phi = 0$, $V = 0$: in this way the vacuum density disappears and the expansion stops.

A slow variation of the field can be achieved with a large value of the Hubble parameter H. Neglecting the ∇^2 term in Eq. 6.25, the equation of motion for the inflaton field looks like the Newton equation for a particle moving through a medium with friction (the $3H\dot{\phi}$ term). That's why sometimes the large-H assumption is called *Hubble friction*.

Under this condition, the "velocity" of the field must be small, so we can

---

[1]This means that $g = det(g_{\mu\nu}) = a^3$ ("cartesian" coordinates) and since we assume homogeneity and isotropy, spacial gradients are zero and the equation of motion reduces to $\frac{1}{\sqrt{-g}}\partial_t(-g\partial\phi) + \sqrt{-g}dV/d\phi = 0$.

assume $\ddot{\phi} \approx 0 \implies \dot{\phi} \approx const.$ This approximation is called **slow-roll approximation** and the equation of motion reduces to

$$\dot{\phi} = -\frac{dV/d\phi}{3H} \quad . \tag{6.28}$$

If in the inflationary phase the cosmological density is dominated by the slowly varying inflaton field, using the first Friedmann Equation with $\dot{H} = 0$ (true during inflation) and $w = -1$ (true for a scalar field), we have

$$H^2 = \frac{8\pi G}{3} V(\phi) \quad . \tag{6.29}$$

From the slow-roll approximation

$$dt = \frac{3H}{(dV/d\phi)d\phi} \quad , \tag{6.30}$$

so we can estimate

$$N = \int d(\ln a) = \int H dt = 8\pi G \int d\phi \frac{V(\phi)}{dV/d\phi} \quad . \tag{6.31}$$

The number N is the number of **e-foldings**, which is the number of times the Universe grew by a factor of e. In order to produce a simple estimate, we can assume $V \sim g\phi^n/n$ and calculate the number of e-foldings between two values of the inflaton field ϕ_1 and ϕ_2 which are respectively values at the beginning and end of the inflation phase

$$N = \frac{4\pi G}{n}(\phi_1^2 - \phi_2^2) \approx \frac{4\pi G}{n}\phi_1^2 = \frac{4\pi}{nM_P^2}\phi_1^2 \quad , \tag{6.32}$$

where we assumed $\phi_2 \ll \phi_1$ at the end of inflation and $M_P = \sqrt{1/G} \sim 1.2 \times 10^{19}$ GeV/c^2 is the Planck mass ($\hbar = c = 1$).
The largest scales in the CMB are produced at $N_{CMB} \sim 60$ before the end of inflation and therefore $N > N_{CMB}$ for solving the horizon problem. The inflationary phase can fix the standard cosmological theory problems if for example $N \sim 70$.
This implies the estimate for the initial value of the inflaton of

$$\phi_1^2 > 5.6 \cdot n \cdot M_P^2 \quad . \tag{6.33}$$

6.9 Reheating

During inflation most of the energy density in the universe is in the inflaton potential. The inflationary phase ends when the potential becomes steep and the inflaton field gains kinetic energy. The energy of the inflaton has to be transferred to the SM particles. This process is called **reheating** and corresponds to the start of the classical hot Big Bang. After reaching the minimum, the inflation starts to oscillate into it. Let's assume $V(\phi) = m^2\phi^2$ in the neighborhood of the minimum. With homogeneity we have

$$\ddot{\phi} + 3H\dot{\phi} + m^2\phi = 0 \quad . \tag{6.34}$$

The Universe expands and the expansion scale will become larger than the oscillation period of the inflaton. This situation is described by $H^{-1} \ll m^{-1}$ and it means that we can disregard the Hubble friction term in Eq. 6.34 and have oscillations of frequency m.

6.10 Double Scalar Field Inflation

A variant of the slow-roll mechanism introduced before is to consider two scalar fields ϕ and χ with the following potential

$$V(\phi, \chi) = \frac{1}{2}(a\phi^2 - v^2)\chi^2 + \frac{b}{4}\chi^4 + V(\phi) \quad , \tag{6.35}$$

where $V(\phi)$ is a slow-rolling potential and $a, b > 0$ are constants.
At the beginning, the inflaton field ϕ is large and evolves in the potential "valley" defined by $\chi \sim 0$.
When $\phi^2 < v^2/a$, the second field χ acquires a non-zero vacuum expectation value $\chi^2 \sim v^2/b$ and the effective mass of ϕ becomes large: $m_{eff}^2 \sim (a/b)v^2$. The large mass drives the inflaton towards the equilibrium point at $\phi = 0$. While going to zero, the inflaton converts in other particles and reheats the Universe starting again the normal expansion phase.

6.11 Starobinski R^2 Inflation

In this scenario, inflation is realized through a modification of the Einstein Equations instead of introducing a particular energy-momentum tensor. The idea of adding higher order terms to the Einstein-Hilbert legrangian was motivated by the observation that quantum corrections can induce such terms and quantum effects should have played a non-negligible role in the very early Universe.
The Einstein Equations can be obtained minimizing an action (the Hilbert action) $S \propto \int d^4x \sqrt{-g}R$. The idea is to consider higher-order actions and the simplest consists in adding a new R^2 term:

$$S = -\frac{1}{16\pi G} \int d^4x \sqrt{-g}R(1 - \frac{R}{6m^2}) \quad . \tag{6.36}$$

In the standard Einstein gravity, the Ricci scalar R is connected to the trace of the energy-momentum tensor

$$R = -8\pi G T^\mu_\mu \quad , \tag{6.37}$$

and therefore it is not a real dynamical variable of the theory. In the new R^2 version of the action, the scalar satisfies the equation of motion

$$\ddot{R} + 3H\dot{R} + m^2(R + 8\pi G T^\mu_\mu) = 0 \quad , \tag{6.38}$$

which looks like a Klein-Gordon equation for a scalar field (sometimes called the *scalaron*). The scalaron, in terms of the Hubble parameter is

$$R = -6\dot{H} - 12H^2 \quad . \tag{6.39}$$

In the absence of matter, the above equations describe an exponential expansion with an almost constant H, thus satisfying the requirements for inflation. After the inflationary phase, particles and reheating are produced by the decay of the scalaron.

6.12 Chaotic Inflation

This version of inflation was proposed by A. Linde. At that time, the motivation for introducing this scenario was to show that inflation is a

generic prediction of many theories including the Standard Model.
The idea is that during the Planck era, quantum fluctuations randomly
("chaotically") can drive the inflaton out of its minimum energy, starting
inflation.

Chapter 7 | The Cosmic Microwave Background

The Cosmic Microwave Background (CMB) is a relic electromagnetic radiation from the early Universe. It was predicted before its discovery in 1948 by R. Alpher and R.Herman. The CMB was finally measured by A. Penzias and R. Wilson in 1964 with a ground-based antenna, winning the Nobel price for the discovery in 1978.

Nowadays very precise measurements of the CMB are done with satellites.

The CMB originated at the time where the temperature of the Universe, through expansion, dropped at the point of allowing the capture of electrons by nuclei. The Universe then bacame transparent to the electromagnetic radiation, which then was red-shifted from the time of its production until now. The cosmological red-shift predicts a much colder relic radiation today with respect to its original temperature. Alpher and Herman gave 5K as the first estimate, which turned out to be not far away to the presently known value.

The CMB has today a density of about 514 photons per cm^3 and they traveled for 99.7% of the age of the Universe until they reached our detectors. At the time when the CMB was produced, the Universe was about 1000 times smaller and 1000 warmer than now.

The CMB appears as a rather uniform radiation compatible with a blackbody distribution with a temperature of about 2.7K. What is actually interesting are the deviations from this mean temperature as a function of the angular scale of the sky.

After reading this chapter you have definitely to google "Planck CMB simulator" or go to "http://strudel.org.uk/planck" and look interactively how the comological parameters affect the CMB.

7.1   Recombination

7.2   Multipole Decomposition of the CMB

The CMB has an average temperature (mediated over the whole sky) of about T_0=2.7K. After subtracting this average temperature, we can consider the relative fluctuations around the mean

$$\frac{\delta T}{T_0} = \frac{T - T_0}{T_0}(\theta, \phi) \quad , \tag{7.1}$$

which depend on the two angles θ and ϕ which constitute a coordinate system describing the sky around us. Since we are dealing with small fluctuations on the surface of a sphere (the sky), we can expand the relative fluctuations on the spherical harmonics basis $Y_{l,m}$ (this is analogous to a Fourier series expansion in a "flat" case on the sin / cos basis)

$$\frac{\delta T}{T_0}(\theta, \phi) = \sum_{l,m} a_{l,m} Y_{l,m}(\theta, \phi) \quad . \tag{7.2}$$

Since $Y_{l,m}$ is an orthonormal set of functions [1], we can invert the previous equation obtaining

$$a_{l,m} = \int Y_{l,m}^*(\theta, \phi) \frac{\delta T}{T_0}(\theta, \phi) d\Omega \quad , \tag{7.3}$$

where the integral in $d\Omega$ is done over all angles.
Since we subtracted the average temperature T_0 and $Y_{0,0}$=const, we should have $a_{0,0} = 0$ for the lowest multipole ($l = 0$). For $l = 1$ we have the *dipole contribution* which is due to the Doppler effect caused by the motion of the Earth with respect to the CMB.
Therefore, the interesting part of the CMB which should contain information about its origin at the decoupling time must be contained in the $l > 1$ multipoles.
The (in general, complex) components $a_{l,m}$ represent fluctuations around zero, therefore $\langle a_{l,m} \rangle = 0$. If they represent Gaussian random variables, the whole information about them should be contained in the variances

---

[1] $\int Y_{l,m} Y_{l',m'}^* d\Omega = \delta_{l,l'} \delta_{m,m'}$

$\langle |a_{l,m}|^2 \rangle$ which are connected to the power of the specific (l, m) mode. Given the isotropic nature of the CMB, we expect the variance to be dependent only from l, which is related to the angular size of the anisotropy pattern. Remembering the closure relation $\sum_m |Y_{l,m}|^2 = (2l + 1)/4\pi$ for the spherical harmonics, we can define the **angular power spectrum**

$$C_l = \frac{1}{2l + 1} \sum_m \langle |a_{l,m}|^2 \rangle \quad , \tag{7.4}$$

which is also called the **TT power spectrum**. Sometimes C_l is indicated as C_l^{TT}. If we assume that the $a_{l,m}$ are independent random variables, we have for the correlations

$$\langle a_{l,m} a_{l',m'} \rangle = \delta_{l,l'} \delta_{m,m'} C_l \quad . \tag{7.5}$$

If we assume that the spectrum of the density perturbations in the early Universe was Gaussian, the angular power spectrum contains all the statistical information about the CMB anisotropies and therefore we can proceed in calculating

$$\frac{\delta T}{T_0} = \langle \sum_{l,m} a_{l,m} Y_{l,m} \sum_{l,m} a_{l,m}^* Y_{l,m}^* \rangle = \sum_{l,l',m,m'} Y_{l,m} Y_{l,m}^* \langle a_{l,m} a_{l,m}^* \rangle =$$

$$\sum_l C_l \sum_m |Y_{lm}|^2 = \sum_l \frac{(2l + 1)}{4\pi} C_l \quad . \tag{7.6}$$

A subtle point here is the following: the averaging $\langle \rangle$ should be done over an ensemble of Universes, while we have only one realization of it. We can imagine that averaging over different directions might represent an averaging over an ensemble of different Universes. In practice, the *observed* power spectrum is calculated as follows

$$\frac{1}{4\pi} \int \left(\frac{\delta T}{T} \right)^2 d\Omega = \sum_l \frac{2l + 1}{4\pi} \hat{C}_l \quad , \tag{7.7}$$

with $\hat{C}_l = \sum_m |a_{l,m}|^2 / (2l + 1)$.
So if the theoretical power spectrum the angular average spectrum were the same, we should have $\langle \hat{C}_l \rangle = C_l \Rightarrow \langle \hat{C}_l - C_l \rangle = 0$. The averaged

squared difference between theory and observation is called **cosmic variance** and a direct calculation yields

$$\langle (\hat{C}_l - C_l)^2 \rangle = \frac{2}{2l+1} C_l^2 \quad . \tag{7.8}$$

The last formula shows that the variance is smaller for large l (small scales), while it is large for small l (large scales). The cosmic variance represents a limit on the accuracy of the comparison between theory and experiment.

7.3 Angular Scales

As we anticipated in the previous section, the multipole number l is connected to the angular scale in the sky. The spherical harmonics have an oscillatory pattern on the sphere in the following (approximate) sense: in a full great circle on the spherical surface, there are l wavelengths of oscillations. This means that the angular scale corresponding to the mode l is $\theta = 2\pi/l$. We can define the angular resolution as the angle connected to the distance from a crest and a valley of a wave $\theta_{res} = \pi/l$. A detector must have a resolution at least equal to θ_{res} in order to resolve scales up to l.

For comparison, the first high-resolution satellite mission (COBE) had $\theta_{res} = 7° \Rightarrow l < 26$. The follow-up experiment (WMAP) had $\theta_{res} = 0.23° \Rightarrow l < 783$. The latest (at the time of writing) and most precise satellite mission (Planck) improves the angular resolution about three times over WMAP.

The question we would like to answer now is: if there were density perturbations in the early Universe characterized by (comoving) wavenumbers k (*i.e.* a comoving wavelength $\lambda = 2\pi/k$), to which CMB multipole l will contribute the most? In other words, we would like to link the primordial perturbations to the pattern measured in the CMB.

Let's define the **angular diameter distance** as $d_A = D/\theta$, which is the same as defining the angle θ subtended by an object of width (length perpendicular to the line of sight) D placed at a distance d_A from us. Taking into account the expansion of the Universe we can define the comoving

version of the angular diameter distance

$$d_A^c = \frac{D^c}{\theta} = \frac{(a_0/a)D}{\theta} = \frac{(1+z)D}{\theta} = (1+z)d_A \quad . \qquad (7.9)$$

Considering now the comoving wavelength λ (associated with the comoving wavenumber k) of a density perturbation, the mode should be visible in the CMB at an angular size of

$$\theta_\lambda = \frac{\lambda}{d_A^c} = \frac{2\pi}{k d_A^c} = \frac{2\pi}{l} \quad , \qquad (7.10)$$

which gives the relation $l = k d_A^c$. This result follows from a rather simplified treatment, since clearly it is not possible that a single density perturbation mode contributes to just one single CMB harmonic. The full calculation must take into account all the modes but the basic result we obtained still holds, in the sense that only the modes close to k contribute significantly.

7.4   CMB Polarization

The CBM can be polarized because of different reasons. Thomson scattering is surely present (scattering of photons from charged particles that took place at the last scattering surface) and contributes up to $\sim 5\%$ level which in terms of temperature fluctuations corresponds to few μK. Thompson (linear) polarization was indeed experimentally detected.
The Thompson cross section is proportional to the photon polarization direction before ($\hat{\epsilon}$) and after the scattering ($\hat{\epsilon}'$)

$$\frac{d\sigma}{d\Omega} \propto |\hat{\epsilon} \cdot \hat{\epsilon}'|^2 \quad . \qquad (7.11)$$

Pictorially, the incident photon makes the charged particle (e.g. an electron) oscillate in the direction of the polarization. The oscillation creates radiation with polarization mostly parallel to the initial polarization. If the incident radiation has quadrupole anisotropies, this will result in an emitted linearly polarized radiation (this can be seen since the incident orthogonal components are suppressed in Eq. 7.11).
A photon can be polarized only in the two directions orthogonal to its

propagation. The polatization can always be decomposed in two othogonal modes which are both orthogonal to the direction of propagation. The superposition of the two polarization states given in general an elliptical polarization, and linear or circular polarizations are special cases.

Defining the polarization vector $\hat{\epsilon} = \vec{E}/|E|$ where E is the electric field, the **polarization tensor** is defined as the time average (considering E as an oscillating field in complex representation)

$$p_{ij} = \langle \hat{\epsilon}_i \hat{\epsilon}_j^* \rangle \quad . \tag{7.12}$$

The polarization tensor is traceless

$$Trp = p_{ii} = \langle \hat{\epsilon}_i \hat{\epsilon}_i^* \rangle = \langle |\epsilon| \rangle = 1 \tag{7.13}$$

and Hermitian $(p_{ij})^* = p_{ji}$. An orthogonal basis for Hermitian matrices is provided by the three 2×2 Pauli matrices σ_k. The last observation, combined with the fact that Pauli matrices are trace-less but $Trp = 1$ leads to the following decomposition

$$p_{ij} = \frac{1}{2} \left(I + Q\sigma_1 + U\sigma_2 + V\sigma_3 \right) \tag{7.14}$$

where I is the identity matrix and

$$\sigma_1 = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \quad ; \quad \sigma_2 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \quad ; \quad \sigma_3 = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix} \quad . \tag{7.15}$$

The numbers Q,U,V are called *Stokes parameters* and their nice property is that they are measurable. For example, if we take a linear polarization filter and pass polarized light through it and measure the intensity of light F as a function the filter θ (F_θ) we can verify that $Q = F_0 - F_{90}$, $U = F_{45} - F_{135}$. The "chirality" (the direction where the polarization is rotating) is $V = 2F_C - F$ where F_C is the intensity of the light after passig through a filter which passes circularly polarized light in a certain direction and F is the total incident intensity. Stokes parameters are usually defined between -1 and 1, so Q, U, V are normalized to F.

The degree of polarization is sometimes given as $r = \sqrt{Q^2 + U^2 + V^2}$. The Stokes parameters vary on a spherical surface referred as the Poincare' sphere.

The **intensity tensor** which tells how much intensity there is in each polarization mode is analogously defined as

$$\rho_{ij} = \langle E_i E_j^* \rangle = \frac{1}{2} \left(J \cdot I + Q\sigma_1 + U\sigma_2 + V\sigma_3 \right) \quad . \tag{7.16}$$

In this case we did not normalize by the electric field vector length and thus we have the new factor

$$J = \delta_{ij}\rho_{ij} = |E_x|^2 + |E_y|^2 \tag{7.17}$$

for a certain choice of orthogonal axes x, y while z is the propagation direction of the wave. J is obviously a geometric invariant (independent from the coordinate choice). A second invariant is

$$V = \epsilon_{ij}\rho_{ij} \quad , \tag{7.18}$$

while the Stokes parameters Q and U change with the change of coordinates.
Electromagnetic interactions are parity-conserving and this demands that the helicity must vanish: V=0.
Furthermore, there are two differential invariants (independent from the orientation of the axes)

$$\begin{aligned} S &= \nabla^2 P_E = \partial_i \partial_j \rho_{ij} \\ P &= \nabla^2 P_B = \epsilon_{ik} \partial_i \partial_j \rho_{jk} \end{aligned} \tag{7.19}$$

called scalar and pseudo-scalar invariants, respectively. We consider second-derivatives also because we are dealing with a rank-2 tensor. The other notation ($P_{E/B}$) refers to the so-called "E-modes" and "B-modes" respectively, in analogy to the Helmholtz decomposition of a vector V in a curl-free (irrotational) and divergence-free (solenoidal) parts using a scalar function ψ and a vector function A: $\vec{V} = \vec{\nabla}\psi + \vec{\nabla} \times \vec{A}$.
Actually, the polarization tensor can indeed be decomposed using two scalar functions A and B: $\rho_{ij} = (\partial_i \partial_j - \frac{1}{2}\partial^2)A + (\partial_i \partial_k \epsilon_{kj} + \partial_j \partial_k \epsilon_{ki})B$. We defined already the TT power spectrum related to the correlation function of the temperature fluctuations. We can now define also correlation functions for the polarization fluctiations. Using the decomposition of

the polarization in E and B modes, the only non-vanishing correlation functions (including TT calculated before) are

$$\langle T(\hat{n})T(\hat{n}')\rangle = \frac{1}{4\pi}\sum_{l=0}^{l=\infty}(2l+1)C_l^{TT}P_l(\cos\theta)$$

$$\langle T(\hat{n})E(\hat{n}')\rangle = \frac{1}{4\pi}\sum_{l=0}^{l=\infty}(2l+1)C_l^{TE}P_l(\cos\theta)$$

$$\langle E(\hat{n})E(\hat{n}')\rangle = \frac{1}{4\pi}\sum_{l=0}^{l=\infty}(2l+1)C_l^{EE}P_l(\cos\theta)$$

$$\langle B(\hat{n})B(\hat{n}')\rangle = \frac{1}{4\pi}\sum_{l=0}^{l=\infty}(2l+1)C_l^{BB}P_l(\cos\theta) \quad .$$

(7.20)

Having E and B opposite parity properties, their cross-correlations vanish. The origin of the E/B notation comes from electromagnetism, since an electric (E) field can be written as the gradient of a scalar field, while the magnetic field (B) can be written as the curl of a vector field.
Thompson scattering, being a purely electromagnetic process (parity-conserving), can induce only E-mode polarizations.
B-modes can arise only if $P \neq 0$ and this can happen for example in the case of vector perturbations ($\rho_{ij} = \partial_i V_j - \partial_j V_i \Rightarrow P = \epsilon_{ij}\partial^2\partial_i Vj$: can be caused by magnetized interstellar or intergalactic media), tensor perturbations (e.g. from gravitational waves) or second order scalar perturbations.

7.5 CMB Anisotropies

The spherical harmonic expansion in multipoles of the CMB temperature is formally done from $l = 0$ to $l = \infty$. The $l = 0$ multipole (the *monopole*) is just a constant and its physical interpretation is the average temperature over the whole sky: $T_0 = 2.7255 \pm 0.0006$ K. The temperature can be converted in density of photons n^0, density of mass ρ^0 or density parameter Ω^0_{CMB}:

$$n^0 = 411 \text{ photons/cm}^3$$
$$\rho^0 = 4.64 \times 10^{-34}\text{g/cm}^3 = 2.6 \times 10^{-10}\text{GeV/cm}^3$$
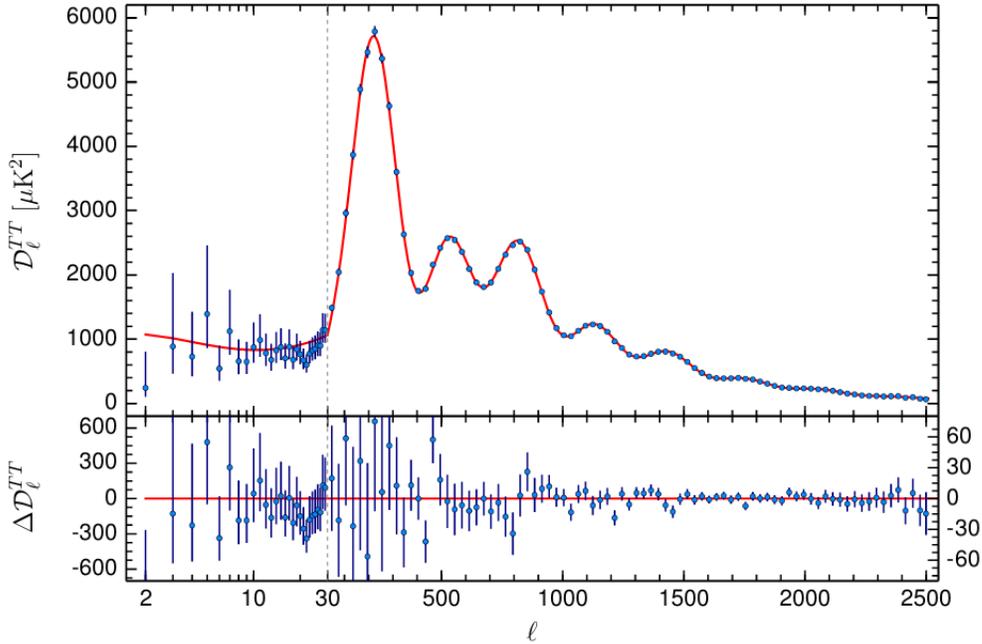$$\Omega^0_{CMB}h_0^2 = 2.47 \times 10^{-5} \quad .$$

(7.21)

Figure 7.1: Planck Spectrum (arXiv:1502.01589).

The $l = 1$ multipole (the *dipole*) represents fluctuations on an angular scale of order π (or 180°): it is like dividing the sky in two halves and look for differences. The dominant contribution to the dipole term comes from the motion of our detector (ultimately of the earth and the sun) with respect to the CMB frame.

The amplitude of the dipole is $T_1 = 3.355 \pm 0.008$ mK, 10^3 times smaller than the monopole: this shows already the rather high uniformity of the CMB, but the error in the measurement tells us that even small anisotropies can be measured with good accuracy. The dipole amplitude leads to the conclusion that the solar system is moving with velocity $v \sim 370$ km/s with respect to the CMB.

Subtracting the $l = 0$ and $l = 1$ terms whose origin is clear, we conclude that the important cosmological information must be encoded into the $l > 1$ multipoles, up to an l_{max} determined by the experimental resolution. Usually, two classes of fluctuations are considered:

Primary Fluctuations: Produced at the last scattering surface or before. These anisotropies carry information about the early universe. In princi-

ple, these anisotropies can be of the scalar, vector or tensor type. Vector modes are stretched out by the expansion of the universe and are therefore expected to be unobservable. Tensor modes decay as they enter the cosmological horizon, so they are suppressed at angular scales smaller than the one of the last scattering surface ($\sim 1°$). The leave a too small imprint into the TT spectrum to be detected but they might be observed in the BB spectrum.

Secondary Fluctuations: These anisotropies araised after the recombination era: they can provide information about the "normal matter" expansion era.

7.5.1 Primary Anisotropies

Looking at Fig. 7.1, the most prominent characteristic is the presence of peaks at $l > 100$. These peaks are the result of the oscillations of the photon-baryon plasma before recombination. Oscillations happen when two opposing forces are at work. In this case, the gravitational force tending to cluster matter (likely around dark matter concentrations) found opposition from the pressure caused by photons. The amplitude of the resulting density fluctuations is quite small ($\delta\rho/\rho \sim 10^{-5}$). This means that we can consider a linear evolution of these perturbations as a good approximantion and in a linear theory every oscillation mode evolves decoupled from the others.

When an inhomogeneity of certain wavelenght entered the cosmological horizon, plasma oscillations started. In the linear approximation, all the inhomogeneity with the same wavelenght entered the horizon at the same time, thus adding in phase.

The first (and highest) peak in Fig. 7.1 was generated by perturbations which entered the horizon at the recombination time. The (smaller) peaks at higher l are caused by perturbations which entered the horizon before recombination.

An analysis of the cosmological perturbations shows that the perturbation spectrum is flat: this means that every mode should have the same amplitude when entering the horizon. Fig. 7.1 shows that the peaks are decreasing in amplitude: this is due to cosmological expansion, since high-l peaks entered the horizon before and were stretched more. After recombination, no peaks can be produced since there is no oscillating

plasma.

The location and height of the peaks depends from the cosmological parameters. The position of the **first peak** is tightly connected to the total amount of matter/enegy present in the universe today (Ω_{tot}^0)

$$l_{1st peak} \sim \frac{220}{\sqrt{\Omega_m}} \quad . \tag{7.22}$$

A quick inspection of the data in Fig. 7.1 shows that $\Omega_{tot}^0 \sim 1$. This means that the Universe today is very close to the critical density and in turn the geometry is very close to the flat one.

The height of the first peak instead, is tightly connected to the amount of barionic matter Ω_B^0.

Why the position of the first peak has to do with the geometry of the Universe? A physical argument is the following. The size corresponding to the first is known, since it is equal to the cosmological horizon (better: the sound horizon, where the speed of light is replaced with the sound speed in the plasma) at the recombination time. The angle[2] under which the first peak is observed today depends on the geometry. This angle in the case of a flat Universe is about $\sim 1°$ which corresponds to $l \sim 220$: exactly where the first peak is.

The plasma oscillations can be seen as the baryons playing the role of the mass while the photons with their pressure are the string. This system oscillates in non-uniform gravitational fields. Following this simple model, odd number peaks (first, third, ...) are associated with how far the plasma "falls" into gravitational potential wells (or how much the plasma compresses). An increase in the amount of baryons enhances these peaks. The even number peaks (second, fourth,...) are associated with how how much the plasma rarefies, therefore with more baryons the odd peaks are enhanced with respect to the even peaks. In summary, baryons make the first peak larger than the second. The ratio between even and odd peaks gives an estimate of the **cosmological baryon to photon ratio** η.

For $l > 1000$, the peaks are highly suppressed by a mechanism called **Silk damping** or diffusion damping. Silk damping describes the diffusion of photons during and after recombination from "hot" to "cold" parts of

---

[2]The angle here is the angle calculated with the triangle given by the distance from us to the last scattering surface and the particle horizon for points at the surfece.

the plasma, realizing effectively a damping of the inhomogeneities resulting in a more uniform distribution. Silk damping exponentially decreases the anisotropies in the CMB on small scales (bounded by the **Silk scale**) which is much smaller than a degree. The recombination (or decoupling) era lasted for a very short time, effectively amplifiying Silk damping.

7.5.2 Secondary Anisotropies

Secondary anisotropies are anisotropies originated in the period at and after recombination. The mechanisms creating these anisotropies are the Sachs-Wolfe effect, reionization and the Sunyaev-Zeldovich effect.
The **Sachs-Wolfe effect** affects photons at the last scattering surface and it is originated by the inhomogeneous gravitational fields present at that time. The **integrated Sachs-Wolfe effect** is the result of photons passing in time-varying gravitational fields. For example a photon can go down a potential well and blue-shift, while the well becomes shallower over time. Since the climb will be less steep, there is an overall non-compensating effect on the photon's wavelenght.
Reionization happened at about $z \sim 10$ after the appearence of the first stars. The presence of free electrons allowed again the possibility of Thomson scattering of CMB photons.
The Sunyaev-Zeldovich effect is a distorsion due to the inverse Compton scattering of CMB photons on electrons accelerated within galaxy clusters. This effect is indeed used for studiying galaxy clusters.

7.5.3 Polarization Anisotropies

The TE, EE and BB anisotropy spectra can provide important additional information about the early Universe. The most important information carried by the TE and EE spectra regards the **reionization era**. After recombination ($z \sim 1100$) and birth of the CMB, stars started to form and the energy produced by them re-ionized the hydrogen gas present in the young Universe (at least before the 1Gyr age, or $z \sim 6 - 20$). The CMB can thus re-scatter on the free electrons inducing polarization. The current best measurement is $z = 11.3 \pm 1.1$ (about 365 million years after the bigbang). Polarization is induced by scattering and it usually summarized

by the **optical depth** by Thompson scattering

$$\tau(z) = \sigma_T n_e^0 \int dz' \frac{cdt}{dz'} x_e(z')(1+z')^3 \quad, \qquad (7.23)$$

which measures the fraction of photons scattered away from the line of sight.

Chapter 8 | Dark Matter

Already in 1932, Jan Hendrik Oort found some discrepancies between the observed rotation curve (the velocity of the stars as a function of the galactic radius) of our own galaxy and the expected one from luminous matter.From this observation, he was not able to exclude that this discrepancy may have been caused by an underestimate of luminous matter due to the presence of absorbing matter. In 1933, Fritz Zwicky's studies of the Coma cluster [?] pointed to a significant discrepancy between the amount of matter deduced from the knowledge of the typical mass-to-light ratio of galaxies, and the gravitation properties of the system. Under the supposition that the Coma system has reached, mechanically, a stationary state, the Virial Theorem implies

$$\langle E_{kin} \rangle = \frac{1}{2} \langle V_g \rangle \tag{8.1}$$

where $\langle E_{kin} \rangle$ and $\langle V_g \rangle$ denote average kinetic and potential energies. Zwicky assumed an uniform mass distribution and a cluster radius R$\sim$ 1 Mly with 800 galaxies with M $\sim$109 solar masses. The total mass estimate was $\sim$ 1.6×10^{45}g. The average gravitational potential energy was therefore $\langle V_g \rangle = (3/5)GM/R$. Using the virial theorem (Eq. 8.1), the average mean squared velocity can be extracted:

$$\sqrt{\langle v^2 \rangle} \approx 80 \frac{\text{km}}{\text{s}} \tag{8.2}$$

This result has to be compared to the observed value of the average Doppler effect of $\sim$1000 km/s. The conclusion was that the average density of the Coma system would have to be at least 400 times larger than that derived from the observations on luminous matter. Zwicky himself commented:

*If this would be confirmed we would get the surprising result that **dark matter** is present in much greater amount than luminous matter.*

8.1 Galaxy Rotation Curves

Until the 1970s, there was not much progress towards the understanding of this discrepancy attributed to some from of non-luminous matter, until Vera Rubin and coworkers published their work on rotation curves of spiral galaxies.
The measurements showed convincingly that the rotational velocities of the stars as a function of the radius R of the galaxies did not follow the expected Kepler's law

$$v(R) = \sqrt{\frac{GM(R)}{R}} \tag{8.3}$$

but they rather stayed about constant out to very large R, as showed in Fig. 8.1. This implied that galaxies were surrounded by a large amount of invisible matter.

8.2 Barionic Mass Estimation with X-ray Halos

Galaxy clusters are composed by abundant barionic matter which usually does not emit radiation. If this matter is present within strong gravitational potentials, bremsstrahlung photons can be emitted (usually in the X-ray band). Measuring these X-rays can lead to an estimation of the amount of barionic matter contained in a cluster, thus providing a tool for measuring its dark matter content by subtraction, if the total gravitational mass could be estimated in another way.
Approximating the cluster as a spherically symmetric system in equilibrium ($\bar{v} = 0$) the hydrodynamical Euler equation

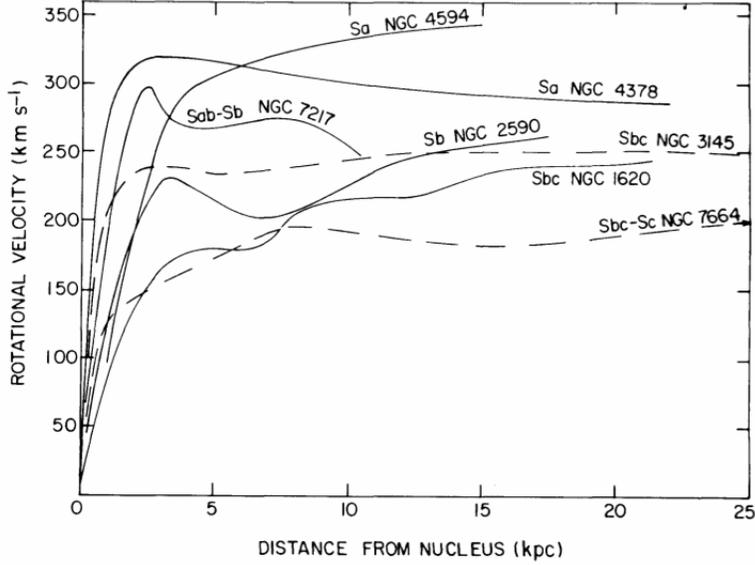$$\rho \frac{d\bar{v}}{dt} = -\nabla P - \rho \nabla \phi \quad , \tag{8.4}$$

Figure 8.1: Rotational curves for different galaxies as measured by Rubin *et al* in V.C. Rubin *et al.*, Astrophys. Journal, 255, 107 (1978).

where P is the pressure, ρ the density, and ϕ the gravitational potential, becomes

$$\frac{dP}{dr} = -\frac{GM(r)}{r^2}\rho \quad . \tag{8.5}$$

$M(r)$ is the amount of matter contained within the radius r. Connecting the pressure $P(r)$ with the temperature $T(r)$ through the law of ideal gases $P = \rho k_B T/m$ and considering only protons for simplicity ($m = m_P$), after some algebra we obtain

$$M(r) = \frac{k_B Tr}{Gm_P}\left(-\frac{d\ln\rho}{d\ln r} - \frac{d\ln T}{d\ln r}\right) \quad . \tag{8.6}$$

The previous equation allows the measurement of the mass profile $M(r)$ through the measurement of the temperature and density profiles $T(r)$ and $\rho(r)$.

The temperature is determined via the shape of the frequency spectrum of the X-ray radiation, or through the strength of the emission lines. The

gas density $\rho(r)$ is proportional to the square root of the *luminosity density*, which is another directly measured quantity.

8.3   Gravitational Mass Estimation with Weak Lensing

Galaxies and clusters of galaxies act as gravitational lenses for the light coming toward us. This means that the deflection of light must be affected by the total gravitational mass of the astrophysical object under consideration, including the possible presence of dark matter.
Weak gravitational lensing (WGL) is the deflection of light emitted from sources behind a massive object (like a galaxy or a cluster of galaxies). Since the distortion of light is not very strong (in contrast to *strong gravitational lensing*, where light is so bended to form the characteristic "arcs"), many measurements of different background objects are needed, so that a WGL measurement is an inherently statistical process.
The weak-field approximation of the FLRW metric (see e.g. the discussion in Sec. 3.3) is

$$ds^2 = \left(1 + 2\frac{\phi}{c^2}\right) dt^2 - \left(1 - 2\frac{\phi}{c^2}\right) dr^2 \quad . \tag{8.7}$$

For a light ray, $ds^2 = 0$, and the last equation, evaluated at the first order for $\phi/c \ll 1$ gives the effective speed of light in a weak gravitational field

$$c' = |\frac{dr}{dt}| = c \left(1 + \frac{2\phi}{c^2}\right) \quad . \tag{8.8}$$

We can also introduce the effective index of refraction $n = c/c' = 1 - 2\phi/c^2 > 1$. With reference to Fig.X, we can write the total deviation angle α as

$$\alpha = \frac{dx}{ds}_O - \frac{dx}{dx}_S = \int_S^O ds \frac{\partial n}{\partial x} = \frac{2}{c^2} \int_S^O \nabla_\perp \phi \quad , \tag{8.9}$$

where the gradient is taken only along the two coordinates orthogonal to the light ray (gravity acts only there, while longitudinally along z the contributions sum up to zero) and therefore ϕ in the last integral is viewed as a vector with 2 components instead of three. Considering the simplified case where the deviating mass is point-like, $\phi = -GM/r$ with

$r = \sqrt{x^2 + y^2 + z^2} = \sqrt{b^2 + z^2}$ ($b^2 = x^2 + y^2$ is the impact parameter). Evaluating Eq. 8.9

$$\alpha = \frac{2GM}{c^2}(\hat{x}, \hat{y}) \int_{-\infty}^{+\infty} \frac{dz}{(b^2 + z^2)^{3/2}} = \frac{4GM}{c^2}(x, y) \left[\frac{z}{b^2(b^2 + z^2)^{1/2}} \right]_0^\infty$$

$$= \frac{4GM}{c^2 b}(\cos \varphi, \sin \varphi)$$

(8.10)

Some interesting points about the deviation due to a gravitational field are the following

- α is linear wrt the mass M, so the effects of more masses just add together to the total deviation.

- The same calculation using Newton's gravity would return an angle smaller by a factor of two.

- Introducing the Schwartzschild radius $R_s = 2GM/c^2$, we can rewrite $\alpha = 2R_S/b$.

8.4   Dark Matter from Astrophysical Measurements

Weak gravitational lensing provides a method for estimating the total mass of an astronomical object, while X-ray surveys are sensitive mostly to the baryonic content. If these two mass estimation methods do not agree, the difference among them should be due to some kin of non-baryonic matter. The today's baryonic density parameter is $\Omega_b^0 = \rho_b^0/\rho_c$ and measurements at different red-shifts are related by $\rho_b/\rho_c = \Omega_b^0/a^{-3}$, as we have seen in Ch 4.

Expressing the baryonic density as $\Omega_b h^2$, where h is the Hubble's constant in 100 $km/s/Mpc$, we have the following different measurements

- From X-ray surveys, $\Omega_b h^2 \sim 0.02$.

- From light absorption from far-away quasars (with higher uncertainty), $\Omega_b h^2 \sim 0.02$.

- From the CMB anisotropy measurements, $\Omega_b h^2 = 0.02225 \pm 0.00023$ (Planck satellite). The density of baryonic matter is deduced from the relative height of the of the odd and even acoustic peaks. If Ω_b is enhanced, also the fist peak is enhanced, while the second is suppressed. The enhancement of Ω_b shifts also the peaks to higher l.

- Another method for estimating Ω_b is based on the predictions of the nucleosithesys models based on our knowledge of nuclear physics. The predicted abundances of light elements match quite well the observations.

In summary, the surprising result $\Omega_b h^2 \sim 0.02$ follows from a variety of observations which are consistent with each other.
Weak lensing is instead sensitive to Ω_m, measuring $\Omega_m h^2 \sim 0.3$ as also other methods:

- Measure for different objects of the mass-to-light ratio as a function of the scale. This ratio saturates to a limiting value past the galaxy cluster scale ($\sim Mpc$). From this, the $\Omega_m h^2 \sim 0.3$ result can be deduced.

- Large-scale surveys for mapping the spectrum of the distribution of galaxies lead to $\Omega_m h^2 \sim 0.2$. While this number is quite different from 0.3, it is still much bigger than the baryonic density.

- The mapping of the cosmic velocity field combined with the distribution of galaxies leads also to $\Omega_m h^2 \sim 0.3$.

- The CMB provides directly $\Omega_m h^2 = 0.308 \pm 0.012$ (Planck satellite, Ade et al, Planck Results XIII, Astron.Astroph. 594, A13 (2015)).

The surprising result is that baryons constitute only about 5% of the critical density, while the total matter content about 30%. Both numbers are surprising: we would have expected that matter constitutes the bulk of the Universe's content, while baryons should have constituted the bulk of the matter content. Both these expectations are put into question by many different measurements, which are quite consistent with each other. Neutrinos are very abundant in the Universe, and as non-baryonic matter, they could be solve the puzzle. Unfortunately, their contribution is

estimated to be $\Omega_\nu h^2 \sim 0.0025$ making them irrelevant in the total balance for Ω_m.

In summary, different measurements point to the existence of additional gravitating matter the exact nature of which we do not know yet.

8.5 Dark Matter and Structure Formation

Although indirect, a very strong argument for the existence of dark matter is based on considerations related to structure formation. The structures we observe today (galaxies, clusters of galaxies) should represent inhomogeneities in the early Universe which acted as "seeds" for gravitational instability and aggregation of matter. These density inhomogeneities $\delta\rho/\rho$ can be observed measuring the $\delta T/T$ anisotropies in the CMB. It turns out that $\delta\rho/\rho \sim 10^{-4}$ and $\delta\rho/\rho \sim a$ where a is the scale factor, which has grown by a factor equal to the red-shift since recombination time ($z \sim 1100$).

For structures to form, we need $\delta\rho/\rho \gg 1$ but since recombination, not enough time has passed for going from $\delta\rho/\rho \sim 10^{-4}$ to the needed size of the perturbations.

This tells us that considering only barions, there was not enough time for structures to form and create what we observe today. Therefore, we need some kind of matter which decoupled from the primordial plasma much earlier and started to clump and form in time the required density perturbations.

8.6 Dark Matter Properties

Having estimated by different methods how much Dark Matter (DM) is present in the Universe, we would like now to know what are its properties, in the case DM is really a new kind of particle(s).

- **Mass:** This parameter is not very well constrained and, depending from the model, can vary within tens of orders of magnitude. Simply estimating the de Broglie wavelength for a particle confined on galactic scales (kpc) with a typical escape velocity of 100km/s, we can derive a lower limit of 10^{-22} eV.

- **Interaction:** DM should be indeed "dark", *i.e.* it should not interact electromagnetically. If DM can interact with known particles, it also depends from the specific model. Since DM cannot radiate, it is believed to be rather dissipationless: this would restrict its ability to clump or accrete around compact objects like black holes with respect to barionic matter. Some models of DM based on the existence of a "dark sector" propose an interaction with the Standard Model photon to some level and in this sense some electromagnetic interaction is allowed. Other models predict the possibility for DM to annihilate into Standard Model particles and this might represent a possible astrophysical signal to detect.

- **Self-Interaction:** Limits to the self-interaction of DM allow for cross-sections of the order of the strong ones.

8.7 Dark Matter as a Thermal Relic

The idea of *thermal decoupling* is an appealing framework for the description of DM. Thermal decoupling assumes that DM was in thermodynamical equilibrium in the early Universe. As the Universe expanded and cooled down, DM density dropped to the point that annihilation basically stopped, *freezing out* DM to the density we observe today.
A slightly more quantitative description is the following. As the density dropped via the expansion, the rate

$$\Gamma = n \cdot \sigma \cdot v \qquad (8.11)$$

of the reaction keeping DM in equilibrium becomes smaller. The Hubble time $1/H(T)$ as a function of the temperature T is a measure of the age of the Universe and the inverse of the reaction rate $1/\Gamma$ tells how long does it take for the reaction to happen on average. So, if $\Gamma \ll H(T)$ is, then the reaction keeping the equilibrium is too slow, since less than one reaction happens in one age of the Universe. In other words, the rate of the reaction does not keep up to the expansion rate of the Universe. The **freeze-out** temperature T_{fo} is the temperature at which expansion and reaction rate are equal

$$\Gamma(T_{fo}) = H(T_{fo}) \quad . \qquad (8.12)$$

While the Universe expands, $\Gamma > H$, until T_{fo} is reached. After that, $\Gamma < H$ and the DM density is "frozen" and then it will keep decreasing with the expansion.

8.8 Hot Thermal Relics and the Example of Neutrinos

If thermal relics are relativistic at the decoupling time, they are called *hot thermal relics*. Neutrinos are an example of such particles: given their almost vanishing mass they move at almost the speed of light at the decoupling. If they have to be thermal relics, they should have been in thermodynamical equilibrium, for example through a reaction like

$$\nu + \bar{\nu} \longleftrightarrow f + \bar{f} \quad , \tag{8.13}$$

where $\nu(\bar{\nu})$ is a neutrino (antineutrino) and $f(\bar{f})$ is a fermion (antifermion). Taking $E \sim T_\nu$ and for the cross-section the Fermi approximation [1] $\sigma \sim G_F^2 T_\nu^2$, at the freeze-out temperature T_ν we require ($v = c = 1$)

$$n(T_\nu) \cdot \sigma(T_\nu) = H(T_\nu) \Rightarrow T_\nu^3 G_F^2 T_\nu^2 = \frac{T_\nu^2}{M_P} \quad . \tag{8.14}$$

where we used the Friedmann equation $H^2 = \frac{8\pi G}{3}\rho$ and $\rho \sim T^4$ for relativistic particles. Solving for the freeze-out temperature

$$T_\nu = (G_F^2 M_P)^{-1/3} \approx 1 \text{ MeV} \quad . \tag{8.15}$$

This result is consistent with the relativistic condition $m \ll T$ assumed at the beginning, so neutrinos are really an example of hot relics.

8.9 Cold Thermal Relics and WIMPs

Cold thermal relics are non-relativistic at freeze-out, so the appropriate approximation for the density is

$$n \sim (mT)^{3/2} e^{-\frac{m}{T}} \quad , \tag{8.16}$$

---

[1] $G_F \sim 10^{-5} \text{GeV}^{-2}$ and we assume $E \ll m_W$.
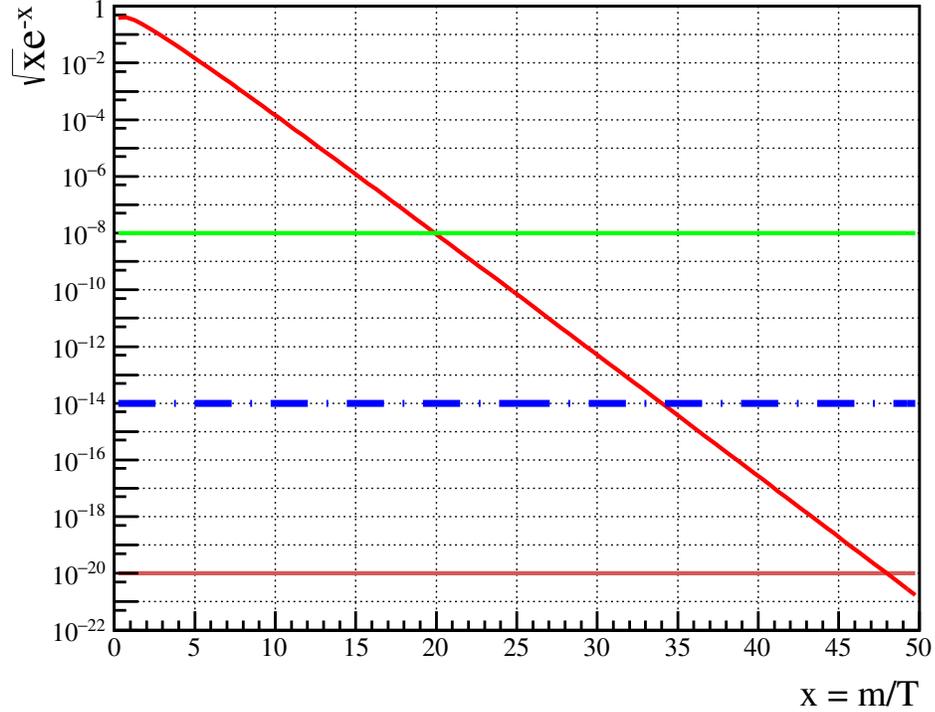
Figure 8.2: Plot of the two sides of Eq. 8.17. The blue dashed line corresponds to the WIMP case $1/(m \cdot M_P \cdot \sigma) = 10^{-14}$. Other two horizontal lines at 10^{-8} and 10^{-20} are added for reference.

The freeze-out condition $n\sigma \sim H$ (we still consider $v \sim c$ up to some factor) in the radiation-dominated phase of the Universe implies $n_{fo} \sim T_{fo}^2/(\sigma M_P)$.

Defining $x = m/T$ ($x \gg 1$ then defines the non-relativistic "cold" regime), the freeze-out condition becomes

$$\sqrt{x}e^{-x} = \frac{1}{m \cdot M_P \cdot \sigma} \quad . \tag{8.17}$$

The last equation does not have analytical solutions and must be solved numerically. A graphical representation of the solution is given in Fig 8.8, where $\sqrt{x}e^{-x}$ is reported together with three cases for $1/(m \cdot M_P \cdot \sigma)$: the solutions are at the intersection points.

Let's try now to calculate the density parameter associated to a cold relic particle with mass m_χ

$$\Omega_\chi = \frac{m_\chi n_\chi(T_0)}{\rho_c} \quad . \tag{8.18}$$

Today, $T_0 = 2.7K \sim 10^{-4}eV$. In an isoentropic FLRW Universe, for relativistic particles we have $T \sim 1/a$ and $n \sim 1/a^3$, so

$$\frac{n_0}{T_0^3} = \frac{n_{fo}}{T_{fo}^3} \quad . \tag{8.19}$$

Substituting n_0 from the last equation into the density parameter equation and using again the freeze-out condition

$$\Omega_\chi = \frac{T_0^3}{\rho_c M_P} \frac{x_{fo}}{\sigma} \quad . \tag{8.20}$$

The dark matter abundance is estimated to be about $\Omega_{DM} \sim 0.2$, so the last equation can be recast in the more suggestive form

$$\frac{\Omega_\chi}{0.2} \simeq \frac{x_{fo}}{20} \left(\frac{10^{-8} \text{GeV}^{-2}}{\sigma} \right) \quad , \tag{8.21}$$

where appropriate numerical values normalize each member to $\mathcal{O}(1)$. In a more exact treatment of the problem, the cross-section of the last equation should be the thermally-averaged cross section $\langle \sigma v \rangle$ for reasons connected to the Boltzmann equation.
Using the equipartition theorem $(3/2)T = (1/2)mv^2$, we can estimate that $v \sim c/3$ for $x \sim 20$ and this leads to the estimate

$$\langle v\sigma \rangle \sim 3 \times 10^{-26} \frac{\text{cm}^3}{\text{s}} \quad . \tag{8.22}$$

This result is often associated to the so-called **WIMP miracle**, which consists in the following coincidence. For various reasons, new physics is expected at the electroweak scale $m \sim E_{EW} \sim 200$ GeV. If we calculate the electroweak pair-annihilation cross-section at freeze-out temperature

$$\sigma_{EW} \sim G_F^2 T_{fo}^2 \sim \left(\frac{E_{EW}}{20} \right)^2 \sim 10^{-8} \text{GeV}^{-2} \quad , \tag{8.23}$$

we obtain the right cross-section which is able to explain the DM abundance. This result is often quoted as an indication that new physics at the electroweak scale might also explain DM in the form of a cold relic from the early Universe. Looking at Fig. 8.8, the dashed line describes about this case with $\sigma_{EW} = G_F^2 m_\chi^2$ and $m_\chi = 100$ GeV, corresponding to $x \sim 35$. Is this really a "miracle"? The previous result was obtained under the assumption of electroweak cross-sections and the cold relic condition $x \gg 1$. In general, following a dimensional argument, a DM annihilation cross-section can be written as $\sigma \sim g^4/m_\chi^2$, where g is some coupling constant. Using Eq. 8.17, $x \gg 1 \Rightarrow m_\chi M_P \sigma \gg 1$, and therefore $m_\chi \gg 0.1$ eV if $\sigma \sim 10^{-8}$ GeV^2. This means that as long as the cross-section is the right one for explaining the DM abundance, the cold relic mass can be very small. The conclusion is that the supposed "miracle" can be realized also without appealing to the electroweak scale.

The argument for understanding the WIMP paradigm can also be restated as following.

As we have seen,

$$\Omega_\chi \propto \frac{1}{\langle v\sigma \rangle} \sim \frac{m_\chi^2}{g_\chi^4} \quad . \tag{8.24}$$

The WIMP miracle states that if we use weak-scale masses and coupling constants, we can roughly reproduce the observed DM abundance. The last equation though fixes only the ratio between couplings and masses and therefore also other combinations might in principle obtain the correct abundance.

8.10 Mass Ranges for Cold Thermal Relics

General limits can be imposed to the allowed mass range of cold thermal dark matter. The requirement of unitarity in the calculation of cross-sections places the approximate bound

$$\sigma < \frac{4\pi}{m_\chi^2} \quad , \tag{8.25}$$

and this, together with Eq. 8.21 approximately implies

$$\frac{\Omega_\chi}{0.2} > 10^{-8} \text{GeV}^{-2} \times \frac{m_\chi^2}{4\pi} \quad . \tag{8.26}$$

Since $\Omega_\chi < 0.2$ we have

$$\left(\frac{m_\chi}{120\ \text{TeV}}\right)^2 < 1 \quad .$$

(8.27)

For a lower limit for WIMPs ($\sigma \sim G_F^2 m_\chi^2$), choosing $x_{fo} \sim 20$ we have

$$\Omega_\chi h^2 \sim 0.1 \frac{10^{-8}\text{GeV}^{-2}}{G_F^2 m_\chi^2} \sim 0.1 \left(\frac{10\ \text{GeV}}{m_\chi}\right)^2 \quad .$$

(8.28)

This lower limit is known as the **Lee-Winberg** limit. The overall mass range allowed for WIMPs goes therefore from few GeVs to many TeVs.

Bibliography

List of Figures

List of Tables