Lectures on General Relativity and Related Topics

Book · January 2016

1 author:

Badis Ydri
Badji Mokhtar - Annaba University

Lectures on General Relativity and Related Topics: Differential Geometry, Cosmology, Black Holes, QFT on Curved Backgrounds and Quantum Gravity

Badis Ydri
Department of Physics, Faculty of Sciences, Annaba University,
Annaba, Algeria.

February 14, 2015

Abstract

These notes originated from a formal course of lectures delivered during the academic years $2012 - 2013$, $2014 - 2015$ to Master students of theoretical physics and also from informal lectures given to Master and doctoral students in theoretical physics who were and still are preparing their dissertations under my supervision.

Contents

Chapter 1

Summary of General Relativity Essentials

1.1 Equivalence Principle

The classical (Newtonian) theory of gravity is based on the following two equations. The gravitational potential Φ generated by a mass density ρ is given by Poisson's equations (with G being Newton constant)

$$\nabla^2\Phi = 4\pi G\rho. \tag{1.1}$$

The force exerted by this potential Φ on a particle of mass m is given by

$$\vec{F} = -m\vec{\nabla}\Phi. \tag{1.2}$$

These equations are obviously not compatible with the special theory of relativity. The above first equation will be replaced, in the general relativistic theory of gravity, by Einstein's equations of motion while the second equation will be replaced by the geodesic equation. From the above two equations we see that there are two measures of gravity: $\nabla^2\Phi$ measures the source of gravity while $\vec{\nabla}\Phi$ measure the effect of gravity. Thus $\vec{\nabla}\Phi$, outside a source of gravity where $\rho = \nabla^2\Phi = 0$, need not vanish. The analogues of these two different measures of gravity, in general relativity, are given by the so-called Ricci curvature tensor $R_{\mu\nu}$ and Riemann curvature tensor $R_{\mu\nu\alpha\beta}$ respectively.

The basic postulate of general relativity is simply that gravity is geometry. More precisely gravity will be identified with the curvature of spacetime which is taken to be a pseudo-Riemannian (Lorentzian) manifold. This can be made more precise by employing the two guiding "principles" which led Einstein to his equations. These are:

- The weak equivalence principle: This states that all particles fall the same way in a gravitational field which is equivalent to the fact that the inertial mass is identical to the gravitational mass. In other words, the dynamics of all free particles, falling in a gravitational field, is completely specified by a single worldline. This is to be contrasted with charged particles in an electric field which obviously follow different worldlines depending on their electric charges. Thus, at any point in spacetime, the effect of gravity is fully encoded in the set of all possible worldlines, corresponding to all initial velocities, passing at that point. These worldlines are precisely the so-called geodesics.

In measuring the electromagnetic field we choose "background observers" who are not subject to electromagnetic interactions. These are clearly inertial observers who follow geodesic motion. The worldline of a charged test body can then be measured by observing the deviation from the inertial motion of the observers.

This procedure can not be applied to measure the gravitational field since by the equivalence principle gravity acts the same way on all bodies, i.e. we can not insulate the "background observers" from the effect of gravity so that they provide inertial observers. In fact, any observer will move under the effect of gravity in exactly the same way as the test body.

The central assumption of general relativity is that we can not, even in principle, construct inertial observers who follow geodesic motion and measure the gravitational force. Indeed, we assume that the spacetime metric is curved and that the worldlines of freely falling bodies in a gravitational field are precisely the geodesics of the curved metric. In other words, the "background observers" which are the geodesics of the curved metric coincide exactly with motion in a gravitational field.

Therefore, gravity is not a force since it can not be measured but is a property of spacetime. Gravity is in fact the curvature of spacetime. The gravitational field corresponds thus to a deviation of the spacetime geometry from the flat geometry of special relativity. But infinitesimally each manifold is flat. This leads us to the Einstein's equivalence principle: In small enough regions of spacetime, the non-gravitational laws of physics reduce to special relativity since it is not possible to detect the existence of a gravitational field through local experiments.

- Mach's principle: This states that all matter in the universe must contribute to the local definition of "inertial motion" and "non-rotating motion". Equivalently the concepts of "inertial motion" and "non-rotating motion" are meaningless in an empty universe. In the theory of general relativity the distribution of matter in the universe, indeed, influence the structure of spacetime. In contrast, the theory of special relativity asserts that "inertial motion" and "non-rotating motion" are not influenced by the distribution of matter in the universe.

Therefor, in general relativity the laws of physics must:

1) reduce to the laws of physics in special relativity in the limit where the metric $g_{\mu\nu}$ becomes flat or in a sufficiently small region around a given point in spacetime.

2) be covariant under general coordinate transformations which generalizes the covariance under Poincaré found in special relativity. This means in particular that only the metric $g_{\mu\nu}$ and quantities derived from it can appear in the laws of physics.

In summary, general relativity is the theory of space, time and gravity in which spacetime is a curved manifold M, which is not necessarily R^4, on which a Lorentzian metric $g_{\mu\nu}$ is defined. The curvature of spacetime in this metric is related to the stress-energy-momentum tensor of the matter in the universe, which is the source of gravity, by Einstein's equations which are schematically given by equations of the form

$$\text{curvature} \propto \text{source of gravity}. \tag{1.3}$$

This is the analogue of (1.1). The worldlines of freely falling objects in this gravitational field are precisely given by the geodesics of this curved metric. In small enough regions of spacetime, curvature vanish, i.e. spacetime becomes flat, and the geodesic become straight. Thus, the analogue of (1.2) is given schematically by an equation of the form

$$\text{worldline of freely falling objects} = \text{geodesic}. \tag{1.4}$$

1.2 Relativistic Mechanics

In special relativity spacetime has the manifold structure R^4 with a flat metric of Lorentzian signature defined on it. In special relativity, as in pre-relativity physics, an inertial motion is one in which the observer or the test particle is non-accelerating which obviously corresponds to no external forces acting on the observer or the test particle. An inertial observer at the origin of spacetime can construct a rigid frame where the grid points are labeled by $x^1 = x$, $x^2 = y$ and $x^3 = z$. Furthermore, she/he can equip the grid points with synchronized clocks which give the reading $x^0 = ct$. This provides a global inertial coordinate system or reference frame of spacetime where every point is labeled by (x^0, x^1, x^2, x^3). The labels has no intrinsic meaning but the interval between two events A and B defined by $-(x_A^0 - x_B^0)^2 + (x_A^i - x_B^i)^2$ is an intrinsic property of spacetime since its value is the same in all global inertial reference frames. The metric tensor of spacetime in a global inertial reference frame $\{x^\mu\}$ is a tensor of type $(0,2)$ with components $\eta_{\mu\nu} = (-1, +1, +1, +1)$, i.e. $ds^2 = -(dx^0)^2 + (dx^i)^2$. The derivative operator associated with this metric is the ordinary derivative, and as a consequence the curvature of this metric vanishes. The geodesics are straight lines. The timelike geodesics are precisely the world lines of inertial observables.

Let t^a be the tangent of a given curve in spacetime. The norm $\eta_{\mu\nu} t^\mu t^\nu$ is positive, negative and zero for spacelike, timelike and lightlike(null) curves respectively. Since material objects can not travel faster than light their paths in spacetime must be timelike. The proper time along a timelike curve parameterized by t is defined by

$$c\tau = \int \sqrt{-\eta_{\mu\nu} t^\mu t^\nu}\, dt. \tag{1.5}$$

This proper time is the elapsed time on a clock carried on the timelike curve. The so-called "twin paradox" is the statement that different timelike curves connecting two points have different proper times. The curve with maximum proper time is the geodesic connecting the two points in question. This curve corresponds to inertial motion between the two points.

The 4−vector velocity of a massive particle with a 4−vector position x^μ is $U^\mu = dx^\mu/d\tau$ where τ is the proper time. Clearly we must have $U^\mu U_\mu = -c^2$. In general, the tangent vector U^μ of a timelike curve parameterized by the proper time τ will be called the 4−vector velocity of the curve and it will satisfy

$$U^\mu U_\mu = -c^2. \tag{1.6}$$

A free particle will be in an inertial motion. The trajectory will therefore be given by a timelike geodesic given by the equation

$$U^\mu \partial_\mu U^\nu = 0. \tag{1.7}$$

Indeed, the operator $U^\mu \partial_\mu$ is the directional derivative along the curve. The energy-momentum 4−vector p^μ of a particle with rest mass m is given by

$$p^\mu = mU^\mu. \tag{1.8}$$

This leads to (with $\gamma = 1/\sqrt{1-\vec{u}^2/c^2}$ and $\vec{u} = d\vec{x}/dt$)

$$E = cp^0 = m\gamma c^2 \ , \ \vec{p} = m\gamma\vec{u}. \tag{1.9}$$

We also compute

$$p^\mu p_\mu = -m^2c^2 \Leftrightarrow E = \sqrt{m^2c^4 + \vec{p}^2c^2}. \tag{1.10}$$

The energy of a particle as measured by an observed whose velocity is v^μ is then clearly given by

$$E = -p^\mu v_\mu. \tag{1.11}$$

1.3 Differential Geometry Primer

1.3.1 Metric Manifolds and Vectors

Metric Manifolds: An $n-$dimensional manifold M is a space which is locally flat, i.e. locally looks like R^n, and furthermore can be constructed from pieces of R^n sewn together smoothly. A Lorentzian or pseudo-Riemannian manifold is a manifold with the notion of "distance", equivalently "metric", included. "Lorentzian" refers to the signature of the metric which in general relativity is taken to be $(-1,+1,+1,+1)$ as opposed to the more familiar/natural "Euclidean" signature given by $(+1,+1,+1,+1)$ valid for Riemannian manifolds. The metric is usually denoted by $g_{\mu\nu}$ while the line element (also called metric in many instances) is written as

$$ds^2 = g_{\mu\nu}dx^\mu dx^\nu. \tag{1.12}$$

For example Minkowski spacetime is given by the flat metric

$$g_{\mu\nu} = \eta_{\mu\nu} = (-1, +1, +1, +1). \tag{1.13}$$

Another extremely important example is Schwarzschild spacetime given by the metric

$$ds^2 = -(1 - \frac{R_s}{r})dt^2 + (1 - \frac{R_s}{r})^{-1}dr^2 + r^2 d\Omega^2. \tag{1.14}$$

This is quite different from the flat metric $\eta_{\mu\nu}$ and as a consequence the curvature of Schwarzschild spacetime is non zero. Another important curved space is the surface of the 2−dimensional sphere on which the metric, which appears as a part of the Schwarzschild metric, is given by

$$ds^2 = r^2 d\Omega^2 = r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{1.15}$$

The inverse metric will be denoted by $g^{\mu\nu}$, i.e.

$$g_{\mu\nu}g^{\nu\lambda} = \eta_\mu^\lambda. \tag{1.16}$$

Charts: A coordinate system (a chart) on the manifold M is a subset U of M together with a one-to-one map $\phi : U \longrightarrow R^n$ such that the image $V = \phi(U)$ is an open set in R^n, i.e. a set in which every point $y \in V$ is the center of an open ball which is inside V. We say that U is an open set in M. Hence we can associate with every point $p \in U$ of the manifold M the local coordinates $(x^1, ..., x^n)$ by

$$\phi(p) = (x^1, ..., x^n). \tag{1.17}$$

Vectors: A curved manifold is not necessarily a vector space. For example the sphere is not a vector space because we do not know how to add two points on the sphere to get another point on the sphere. The sphere which is naturally embedded in R^3 admits at each point p a tangent plane. The notion of a "tangent vector space" can be constructed for any manifold which is embedded in R^n. The tangent vector space at a point p of the manifold will be denoted by V_p.

There is a one-to-one correspondence between vectors and directional derivatives in R^n. Indeed, the vector $v = (v^1, ..., v^n)$ in R^n defines the directional derivative $\sum_\mu v^\mu \partial_\mu$ which acts on functions on R^n. These derivatives are clearly linear and satisfy the Leibniz rule. We will therefore define tangent vectors at a given point p on a manifold M as directional derivatives which satisfy linearity and the Leibniz rule. These directional derivatives can also be thought of as differential displacements on the spacetime manifold at the point p.

This can be made more precise as follows. First, we define s smooth curve on the manifold M as a smooth map from R into M, viz $\gamma : R \longrightarrow M$. A tangent vector at a point p can then be thought of as a directional derivative operator along a curve which goes through p. Indeed, a tangent vector T at $p=\gamma(t)\in M$, acting on smooth functions f on the manifold M, can be defined by

$$T(f)=\frac{d}{dt}(f\circ\gamma(t))|_p. \tag{1.18}$$

In a given chart ϕ the point p will be given by $p=\phi^{-1}(x)$ where $x=(x^1,...,x^n)\in R^n$. Hence $\gamma(t)=\phi^{-1}(x)$. In other words, the map γ is mapped into a curve $x(t)$ in R^n. We have immediately

$$T(f)=\frac{d}{dt}(f\circ\phi^{-1}(x))|_p=\sum_{\mu=1}^{n}X_\mu(f)\frac{dx^\mu}{dt}|_p. \tag{1.19}$$

The maps X_μ act on functions f on the Manifold M as

$$X_\mu(f)=\frac{\partial}{\partial x^\mu}(f\circ\phi^{-1}(x)). \tag{1.20}$$

These can be checked to satisfy linearity and the Leibniz rule. They are obviously directional derivatives or differential displacements since we may make the identification $X_\mu=\partial_\mu$. Hence these vectors are tangent vectors to the manifold M at p. The fact that arbitrary tangent vectors can be expressed as linear combinations of the n vectors X_μ shows that these vectors are linearly independent, span the vector space V_p and that the dimension of V_p is exactly n. Equation (1.19) can then be rewritten as

$$T=\sum_{\mu=1}^{n}X_\mu T^\mu. \tag{1.21}$$

The components T^μ of the vector T are therefore given by

$$T^\mu=\frac{dx^\mu}{dt}|_p. \tag{1.22}$$

1.3.2 Geodesics

The length l of a smooth curve C with tangent T^μ on a manifold M with Riemannian metric $g_{\mu\nu}$ is given by

$$l=\int dt\sqrt{g_{\mu\nu}T^\mu T^\nu}. \tag{1.23}$$

The length is parametrization independent. Indeed, we can show that

$$l=\int dt\sqrt{g_{\mu\nu}T^\mu T^\nu}=\int ds\sqrt{g_{\mu\nu}S^\mu S^\nu}\ ,\ S^\mu=T^\mu\frac{dt}{ds}. \tag{1.24}$$

In a Lorentzian manifold, the length of a spacelike curve is also given by this expression. For a timelike curve for which $g_{ab}T^aT^b<0$ the length is replaced with the proper time τ which is given by $\tau=\int dt\sqrt{-g_{ab}T^aT^b}$. For a lightlike (or null) curve for which $g_{ab}T^aT^b=0$ the length is always 0.

We consider the length of a curve C connecting two points $p=C(t_0)$ and $q=C(t_1)$. In a coordinate basis the length is given explicitly by

$$l=\int_{t_0}^{t_1}dt\sqrt{g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}}. \tag{1.25}$$

The variation in l under an arbitrary smooth deformation of the curve C which keeps the two points p and q fixed is given by

$$\begin{aligned}\delta l &= \frac{1}{2}\int_{t_0}^{t_1}dt\big(g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}\big)^{-\frac{1}{2}}\bigg(\frac{1}{2}\delta g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}+g_{\mu\nu}\frac{dx^\mu}{dt}\frac{d\delta x^\nu}{dt}\bigg)\\ &= \frac{1}{2}\int_{t_0}^{t_1}dt\big(g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}\big)^{-\frac{1}{2}}\bigg(\frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\sigma}\delta x^\sigma\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}+g_{\mu\nu}\frac{dx^\mu}{dt}\frac{d\delta x^\nu}{dt}\bigg)\\ &= \frac{1}{2}\int_{t_0}^{t_1}dt\big(g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}\big)^{-\frac{1}{2}}\bigg(\frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\sigma}\delta x^\sigma\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}-\frac{d}{dt}(g_{\mu\nu}\frac{dx^\mu}{dt})\delta x^\nu+\frac{d}{dt}(g_{\mu\nu}\frac{dx^\mu}{dt}\delta x^\nu)\bigg).\end{aligned} \tag{1.26}$$

We can assume without any loss of generality that the parametrization of the curve C satisfies $g_{\mu\nu}(dx^\mu/dt)(dx^\nu/dt)=1$. In other words, we choose dt^2 to be precisely the line element (interval) and thus $T^\mu=dx^\mu/dt$ is the $4-$velocity. The last term in the above equation becomes obviously a total derivative which vanishes by the fact that the considered deformation keeps the two end points p and q fixed. We get then

$$\begin{aligned}\delta l &= \frac{1}{2}\int_{t_0}^{t_1}dt\delta x^\sigma\bigg(\frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\sigma}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}-\frac{d}{dt}(g_{\mu\sigma}\frac{dx^\mu}{dt})\bigg)\\ &= \frac{1}{2}\int_{t_0}^{t_1}dt\delta x^\sigma\bigg(\frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\sigma}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}-\frac{\partial g_{\mu\sigma}}{\partial x^\nu}\frac{dx^\nu}{dt}\frac{dx^\mu}{dt}-g_{\mu\sigma}\frac{d^2x^\mu}{dt^2}\bigg)\\ &= \frac{1}{2}\int_{t_0}^{t_1}dt\delta x^\sigma\bigg(\frac{1}{2}(\frac{\partial g_{\mu\nu}}{\partial x^\sigma}-\frac{\partial g_{\mu\sigma}}{\partial x^\nu}-\frac{\partial g_{\nu\sigma}}{\partial x^\mu})\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}-g_{\mu\sigma}\frac{d^2x^\mu}{dt^2}\bigg)\\ &= \frac{1}{2}\int_{t_0}^{t_1}dt\delta x_\rho\bigg(\frac{1}{2}g^{\rho\sigma}(\frac{\partial g_{\mu\nu}}{\partial x^\sigma}-\frac{\partial g_{\mu\sigma}}{\partial x^\nu}-\frac{\partial g_{\nu\sigma}}{\partial x^\mu})\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}-\frac{d^2x^\rho}{dt^2}\bigg).\end{aligned} \tag{1.27}$$

By definition geodesics are curves which extremize the length l. The curve C extremizes the length between the two points p and q if and only if $\delta l=0$. This leads immediately to the equation

$$\Gamma^\rho{}_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}+\frac{d^2x^\rho}{dt^2}=0. \tag{1.28}$$

This equation is called the geodesic equation. It is the relativistic generalization of Newton's second law of motion (1.2). The Christoffel symbols are defined by

$$\Gamma^\rho{}_{\mu\nu}=-\frac{1}{2}g^{\rho\sigma}(\frac{\partial g_{\mu\nu}}{\partial x^\sigma}-\frac{\partial g_{\mu\sigma}}{\partial x^\nu}-\frac{\partial g_{\nu\sigma}}{\partial x^\mu}). \tag{1.29}$$

In the absence of curvature we will have $g_{\mu\nu} = \eta_{\mu\nu}$ and hence $\Gamma = 0$. In other words, the geodesics are locally straight lines.

Since the length between any two points on a Riemannian manifold (and between any two points which can be connected by a spacelike curve on a Lorentzian manifold) can be arbitrarily long we conclude that the shortest curve connecting the two points must be a geodesic as it is an extremum of length. Hence the shortest curve is the straightest possible curve. The converse is not true: a geodesic connecting two points is not necessarily the shortest path.

Similarly, the proper time between any two points which can be connected by a timelike curve on a Lorentzian manifold can be arbitrarily small and thus the curve with the greatest proper time, if it exists, must be a timelike geodesic as it is an extremum of proper time. On the other hand, a timelike geodesic connecting two points is not necessarily the path with maximum proper time.

1.3.3 Tensors

Tangent (Contravariant) Vectors: Tensors are a generalization of vectors. Let us start then by giving a more precise definition of the tangent vector space V_p. Let $\mathcal{F}$ be the set of all smooth functions f on the manifold M, i.e. $f : M \longrightarrow R$. We define a tangent vector v at the point $p \in M$ as a map $v : \mathcal{F} \longrightarrow R$ which is required to satisfy linearity and the Leibniz rule. In other words,

$$v(af + bg) = av(f) + bv(g)\ ,\ v(fg) = f(p)v(g) + g(p)v(f)\ ,\ a, b \in R\ ,\ f, g \in \mathcal{F}. \tag{1.30}$$

The vector space V_p is simply the set of all tangents vectors v at p. The action of the vector v on the function f is given explicitly by

$$v(f) = \sum_{\mu=1}^{n} v^\mu X_\mu(f)\ ,\ X_\mu(f) = \frac{\partial}{\partial x^\mu}(f \circ \phi^{-1}(x)). \tag{1.31}$$

In a different chart $\phi^{'}$ we will have

$$X^{'}_\mu(f) = \frac{\partial}{\partial x^{'\mu}}(f \circ \phi^{'-1})|_{x^{'}=\phi^{'}(p)}. \tag{1.32}$$

We compute

$$\begin{aligned}
X_\mu(f) &= \frac{\partial}{\partial x^\mu}(f \circ \phi^{-1})|_{x=\phi(p)} \\
&= \frac{\partial}{\partial x^\mu} f \circ \phi^{'-1}(\phi^{'} \circ \phi^{-1})|_{x=\phi(p)} \\
&= \sum_{\nu=1}^{n} \frac{\partial x^{'\nu}}{\partial x^\mu} \frac{\partial}{\partial x^{'\nu}}(f \circ \phi^{'-1}(x^{'}))|_{x^{'}=\phi^{'}(p)} \\
&= \sum_{\nu=1}^{n} \frac{\partial x^{'\nu}}{\partial x^\mu} X^{'}_\nu(f).
\end{aligned} \tag{1.33}$$

This is why the basis elements X_μ may be thought of as the partial derivative operators $\partial/\partial x^\mu$. The tangent vector v can be rewritten as $v=\sum_{\mu=1}^n v^\mu X_\mu=\sum_{\mu=1}^n v^{'\mu}X^{'}_\mu$. We conclude immediately that

$$v^{'\nu}=\sum_{\nu=1}^n\frac{\partial x^{'\nu}}{\partial x^\mu}v^\mu. \tag{1.34}$$

This is the transformation law of tangent vectors under the coordinate transformation $x^\mu\longrightarrow x^{'\mu}$.

Cotangent Dual (covariant) Vectors or 1-Forms: Let V_p^* be the space of all linear maps ω^* from V_p into R, viz $\omega^*:V_p\longrightarrow R$. The space V_p^* is the so-called dual vector space to V_p where addition and multiplication by scalars are defined in an obvious way. The elements of V_p^* are called dual vectors. The dual vector space V_p^* is also called the cotangent dual vector space at p and the vector space of one-forms at p. The elements of V_p^* are then called cotangent dual vectors. Another nomenclature is to refer to the elements of V_p^* as covariant vectors as opposed to the elements of V_p which are referred to as contravariant vectors.

The basis $\{X^{\mu *}\}$ of V_p^* is called the dual basis to the basis $\{X_\mu\}$ of V_p. The basis elements of V_p^* are given by vectors $X^{\mu *}$ defined by

$$X^{\mu *}(X_\nu)=\delta^\mu_\nu. \tag{1.35}$$

We have the transformation law

$$X^{\mu *} \;=\; \sum_{\nu=1}^n\frac{\partial x^\mu}{\partial x^{'\nu}}X^{\nu *'}. \tag{1.36}$$

From this result we can think of the basis elements $X^{\mu *}$ as the gradients dx^μ, viz

$$X^{\mu *}\equiv dx^\mu. \tag{1.37}$$

Let $v=\sum_\mu v^\mu X_\mu$ be an arbitrary tangent vector in V_p, then the action of the dual basis elements $X^{\mu *}$ on v is given by

$$X^{\mu *}(v)=v^\mu. \tag{1.38}$$

The action of a general element $\omega^*=\sum_\mu\omega_\mu X^{\mu *}$ of V_p^* on v is given by

$$\omega^*(v)=\sum_\mu\omega_\mu v^\mu. \tag{1.39}$$

Again we conclude the transformation law

$$\omega^{'}_\nu=\sum_{\nu=1}^n\frac{\partial x^\mu}{\partial x^{'\nu}}\omega_\mu. \tag{1.40}$$

Generalization: A tensor T of type (k,l) over the tangent vector space V_p is a multilinear map form $(V_p^*\times V_p^*\times ...\times V_p^*)\times(V_p\times V_p\times ...\times V_p)$ into R given by

$$T: V_p^*\times V_p^*\times ...\times V_p^*\times V_p\times V_p\times ...\times V_p\longrightarrow R. \tag{1.41}$$

The domain of this map is the direct product of k cotangent dual vector space V_p^* and l tangent vector space V_p. The space $\mathcal{T}(k,l)$ of all tensors of type (k,l) is a vector space of dimension $n^k.n^l$ since $\mathrm{dim}V_p=\mathrm{dim}V_p^*=n$.

The tangent vectors $v\in V_p$ are therefore tensors of type $(1,0)$ whereas the cotangent dual vectors $v\in V_p^*$ are tensors of type $(0,1)$. The metric g is a tensor of type $(0,2)$, i.e. a linear map from $V_p\times V_p$ into R, which is symmetric and nondegenerate.

1.4 Curvature Tensor

1.4.1 Covariant Derivative

A covariant derivative is a derivative which transforms covariantly under coordinates transformations $x\longrightarrow x^{'}$. In other words, it is an operator ∇ on the manifold M which takes a differentiable tensor of type (k,l) to a differentiable tensor of type $(k,l+1)$. It must clearly satisfy the obvious properties of linearity and Leibniz rule but also satisfies other important rules such as the torsion free condition given by

$$\nabla_{\mu}\nabla_{\nu}f=\nabla_{\nu}\nabla_{\mu}f\ ,\ f\in\mathcal{F}. \tag{1.42}$$

Furthermore, the covariant derivative acting on scalars must be consistent with tangent vectors being directional derivatives. Indeed, for all $f\in\mathcal{F}$ and $t^{\mu}\in V_p$ we must have

$$t^{\mu}\nabla_{\mu}f=t(f)\equiv t^{\mu}\partial_{\mu}f. \tag{1.43}$$

In other words, if ∇ and $\tilde{\nabla}$ be two covariant derivative operators, then their action on scalar functions must coincide, viz

$$t^{\mu}\nabla_{\mu}f=t^{\mu}\tilde{\nabla}_{\mu}f=t(f). \tag{1.44}$$

We compute now the difference $\tilde{\nabla}_{\mu}(f\omega_{\nu})-\nabla_{\mu}(f\omega_{\nu})$ where ω is some cotangent dual vector. We have

$$\begin{aligned}\tilde{\nabla}_{\mu}(f\omega_{\nu})-\nabla_{\mu}(f\omega_{\nu}) &= \tilde{\nabla}_{\mu}f.\omega_{\nu}+f\tilde{\nabla}_{\mu}\omega_{\nu}-\nabla_{\mu}f.\omega_{\nu}-f\nabla_{\mu}\omega_{\nu}\\ &= f(\tilde{\nabla}_{\mu}\omega_{\nu}-\nabla_{\mu}\omega_{\nu}).\end{aligned} \tag{1.45}$$

We use without proof the following result. Let $\omega_{\nu}^{'}$ be the value of the cotangent dual vector ω_{ν} at a nearby point $p^{'}$, i.e. $\omega_{\nu}^{'}-\omega_{\nu}$ is zero at p. Since the cotangent dual vector ω_{ν} is a smooth function on the manifold, then for each $p^{'} \in M$, there must exist smooth functions $f_{(\alpha)}$ which vanish at the point p and cotangent dual vectors $\mu_\nu^{(\alpha)}$ such that

$$\omega_\nu^{'} - \omega_\nu = \sum_\alpha f_{(\alpha)} \mu_\nu^{(\alpha)}. \tag{1.46}$$

We compute immediately

$$\tilde{\nabla}_\mu (\omega_\nu^{'} - \omega_\nu) - \nabla_\mu (\omega_\nu^{'} - \omega_\nu) = \sum_\alpha f_{(\alpha)} (\tilde{\nabla}_\mu \mu_\nu^{(\alpha)} - \nabla_\mu \mu_\nu^{(\alpha)}). \tag{1.47}$$

This is 0 since by assumption $f_{(\alpha)}$ vanishes at p. Hence we get the result

$$\tilde{\nabla}_\mu \omega_\nu^{'} - \nabla_\mu \omega_\nu^{'} = \tilde{\nabla}_\mu \omega_\nu - \nabla_\mu \omega_\nu. \tag{1.48}$$

In other words, the difference $\tilde{\nabla}_\mu \omega_\nu - \nabla_\mu \omega_\nu$ depends only on the value of ω_ν at the point p although both $\tilde{\nabla}_\mu \omega_\nu$ and $\nabla_\mu \omega_\nu$ depend on how ω_ν changes as we go away from the point p since they are derivatives. Putting this differently we say that the operator $\tilde{\nabla}_\mu - \nabla_\mu$ is a linear map which takes cotangent dual vectors at a point p into tensors, of type $(0,2)$, at p and not into tensor fields defined in a neighborhood of p. We write

$$\nabla_\mu \omega_\nu = \tilde{\nabla}_\mu \omega_\nu - C^\gamma{}_{\mu\nu} \omega_\gamma. \tag{1.49}$$

The tensor $C^\gamma{}_{\mu\nu}$ stands for the map $\tilde{\nabla}_\mu - \nabla_\mu$ and it is clearly a tensor of type $(1,2)$. By setting $\omega_\mu = \nabla_\mu f = \tilde{\nabla}_\mu f$ we get $\nabla_\mu \nabla_\nu f = \tilde{\nabla}_\mu \tilde{\nabla}_\nu f - C^\gamma{}_{\mu\nu} \nabla_\gamma f$. By employing now the torsion free condition (1.42) we get immediately

$$C^\gamma{}_{\mu\nu} = C^\gamma{}_{\nu\mu}. \tag{1.50}$$

Let us consider now the difference $\tilde{\nabla}_\mu (\omega_\nu t^\nu) - \nabla_\mu (\omega_\nu t^\nu)$ where t^ν is a tangent vector. Since $\omega_\nu t^\nu$ is a function we have

$$\tilde{\nabla}_\mu (\omega_\nu t^\nu) - \nabla_\mu (\omega_\nu t^\nu) = 0. \tag{1.51}$$

From the other hand, we compute

$$\tilde{\nabla}_\mu (\omega_\nu t^\nu) - \nabla_\mu (\omega_\nu t^\nu) = \omega_\nu (\tilde{\nabla}_\mu t^\nu - \nabla_\mu t^\nu + C^\nu{}_{\mu\gamma} t^\gamma). \tag{1.52}$$

Hence, we must have

$$\nabla_\mu t^\nu = \tilde{\nabla}_\mu t^\nu + C^\nu{}_{\mu\gamma} t^\gamma. \tag{1.53}$$

For a general tensor $T^{\mu_1 \ldots \mu_k}{}_{\nu_1 \ldots \nu_l}$ of type (k,l) the action of the covariant derivative operator will be given by the expression

$$\nabla_\gamma T^{\mu_1 \ldots \mu_k}{}_{\nu_1 \ldots \nu_l} = \tilde{\nabla}_\gamma T^{\mu_1 \ldots \mu_k}{}_{\nu_1 \ldots \nu_l} + \sum_i C^{\mu_i}{}_{\gamma d} T^{\mu_1 \ldots d \ldots \mu_k}{}_{\nu_1 \ldots \nu_l} - \sum_j C^d{}_{\gamma \nu_j} T^{\mu_1 \ldots \mu_k}{}_{\nu_1 \ldots d \ldots \nu_l}. \tag{1.54}$$

1.4.2 Parallel Transport

Let C be a curve with a tangent vector t^μ. Let v^μ be some tangent vector defined at each point on the curve. The vector v^μ is parallelly transported along the curve C if and only if

$$t^\mu \nabla_\mu v^\nu|_{\rm curve} = 0. \tag{1.55}$$

If t is the parameter along the curve C then $t^\mu = dx^\mu/dt$ are the components of the vector t^μ in the coordinate basis. The parallel transport condition reads explicitly

$$\frac{dv^\nu}{dt} + \Gamma^\nu{}_{\mu\lambda} t^\mu v^\lambda = 0. \tag{1.56}$$

By demanding that the inner product of two vectors v^μ and w^μ is invariant under parallel transport we obtain, for all curves and all vectors, the condition

$$t^\mu \nabla_\mu (g_{\alpha\beta} v^\alpha w^\beta) = 0 \Rightarrow \nabla_\mu g_{\alpha\beta} = 0. \tag{1.57}$$

Thus given a metric $g_{\mu\nu}$ on a manifold M the most natural covariant derivative operator is the one under which the metric is covariantly constant.

There exists a unique covariant derivative operator ∇_μ which satisfies $\nabla_\mu g_{\alpha\beta} = 0$. The proof goes as follows. We know that $\nabla_\mu g_{\alpha\beta}$ is given by

$$\nabla_\mu g_{\alpha\beta} = \tilde{\nabla}_\mu g_{\alpha\beta} - C^\gamma{}_{\mu\alpha} g_{\gamma\beta} - C^\gamma{}_{\mu\beta} g_{\alpha\gamma}. \tag{1.58}$$

By imposing $\nabla_\mu g_{\alpha\beta} = 0$ we get

$$\tilde{\nabla}_\mu g_{\alpha\beta} = C^\gamma{}_{\mu\alpha} g_{\gamma\beta} + C^\gamma{}_{\mu\beta} g_{\alpha\gamma}. \tag{1.59}$$

Equivalently

$$\tilde{\nabla}_\alpha g_{\mu\beta} = C^\gamma{}_{\alpha\mu} g_{\gamma\beta} + C^\gamma{}_{\alpha\beta} g_{\mu\gamma}. \tag{1.60}$$

$$\tilde{\nabla}_\beta g_{\mu\alpha} = C^\gamma{}_{\mu\beta} g_{\gamma\alpha} + C^\gamma{}_{\alpha\beta} g_{\mu\gamma}. \tag{1.61}$$

Immediately, we conclude that

$$\tilde{\nabla}_\mu g_{\alpha\beta} + \tilde{\nabla}_\alpha g_{\mu\beta} - \tilde{\nabla}_\beta g_{\mu\alpha} = 2C^\gamma{}_{\mu\alpha} g_{\gamma\beta}. \tag{1.62}$$

In other words,

$$C^\gamma{}_{\mu\alpha} = \frac{1}{2} g^{\gamma\beta} (\tilde{\nabla}_\mu g_{\alpha\beta} + \tilde{\nabla}_\alpha g_{\mu\beta} - \tilde{\nabla}_\beta g_{\mu\alpha}). \tag{1.63}$$

This choice of $C^\gamma{}_{\mu\alpha}$ which solves $\nabla_\mu g_{\alpha\beta} = 0$ is unique. In other words, the corresponding covariant derivative operator is unique. The most important case corresponds to the choice $\tilde{\nabla}_a = \partial_a$ for which case $C^c{}_{ab}$ is denoted $\Gamma^c{}_{ab}$ and is called the Christoffel symbol.

Equation (1.56) is almost the geodesic equation. Recall that geodesics are the straightest possible lines on a curved manifold. Alternatively, a geodesic can be defined as a curve whose tangent vector t^μ is parallelly transported along itself, viz $t^\mu \nabla_\mu t^\nu = 0$. This reads in a coordinate basis as

$$\frac{d^2x^\nu}{dt^2} + \Gamma^\nu{}_{\mu\lambda}\frac{dx^\mu}{dt}\frac{dx^\lambda}{dt} = 0. \tag{1.64}$$

This is precisely (1.28). This is a set of n coupled second order ordinary differential equations with n unknown $x^\mu(t)$. We know, given appropriate initial conditions $x^\mu(t_0)$ and $dx^\mu/dt|_{t=t_0}$, that there exists a unique solution. Conversely, given a tangent vector t^μ at a point p of a manifold M there exists a unique geodesic which goes through p and is tangent to t^μ.

1.4.3 The Riemann Curvature Tensor

Definition: The parallel transport of a vector from point p to point q on the manifold M is actually path-dependent. This path-dependence is directly measured by the so-called Riemann curvature tensor. The Riemann curvature tensor can be defined in terms of the failure of successive operations of differentiation to commute. Let us start with an arbitrary tangent dual vector ω_a and an arbitrary function f. We want to calculate $(\nabla_a\nabla_b - \nabla_b\nabla_a)\omega_c$. First we have

$$\nabla_a\nabla_b(f\omega_c) = \nabla_a\nabla_b f.\omega_c + \nabla_b f\nabla_a\omega_c + \nabla_a f\nabla_b\omega_c + f\nabla_a\nabla_b\omega_c. \tag{1.65}$$

Similarly

$$\nabla_b\nabla_a(f\omega_c) = \nabla_b\nabla_a f.\omega_c + \nabla_a f\nabla_b\omega_c + \nabla_b f\nabla_a\omega_c + f\nabla_b\nabla_a\omega_c. \tag{1.66}$$

Thus

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)(f\omega_c) = f(\nabla_a\nabla_b - \nabla_b\nabla_a)\omega_c. \tag{1.67}$$

We can follow the same set of arguments which led from (A.58) to (A.62) to conclude that the tensor $(\nabla_a\nabla_b - \nabla_b\nabla_a)\omega_c$ depends only on the value of ω_c at the point p. In other words $\nabla_a\nabla_b - \nabla_b\nabla_a$ is a linear map which takes tangent dual vectors into tensors of type $(0,3)$. Equivalently we can say that the action of $\nabla_a\nabla_b - \nabla_b\nabla_a$ on tangent dual vectors is equivalent to the action of a tensor of type $(1,3)$. Thus we can write

$$(\nabla_a\nabla_b - \nabla_b\nabla_a)\omega_c = R_{abc}{}^d\omega_d. \tag{1.68}$$

The tensor $R_{abc}{}^d$ is precisely the Riemann curvature tensor. We compute explicitly

$$\begin{aligned}\nabla_a\nabla_b\omega_c &= \nabla_a(\partial_b\omega_c - \Gamma^d{}_{bc}\omega_d)\\ &= \partial_a(\partial_b\omega_c - \Gamma^d{}_{bc}\omega_d) - \Gamma^e{}_{ab}(\partial_e\omega_c - \Gamma^d{}_{ec}\omega_d) - \Gamma^e{}_{ac}(\partial_b\omega_e - \Gamma^d{}_{be}\omega_d)\\ &= \partial_a\partial_b\omega_c - \partial_a\Gamma^d{}_{bc}.\omega_d - \Gamma^d{}_{bc}\partial_a\omega_d - \Gamma^e{}_{ab}\partial_e\omega_c + \Gamma^e{}_{ab}\Gamma^d{}_{ec}\omega_d - \Gamma^e{}_{ac}\partial_b\omega_e + \Gamma^e{}_{ac}\Gamma^d{}_{be}\omega_d.\end{aligned} \tag{1.69}$$

Thus

$$(\nabla_a\nabla_b-\nabla_b\nabla_a)\omega_c \;=\; \bigg(\partial_b\Gamma^d{}_{ac}-\partial_a\Gamma^d{}_{bc}+\Gamma^e{}_{ac}\Gamma^d{}_{be}-\Gamma^a{}_{bc}\Gamma^d{}_{ae}\bigg)\omega_d. \tag{1.70}$$

We get then the components

$$R_{abc}{}^d \;=\; \partial_b\Gamma^d{}_{ac}-\partial_a\Gamma^d{}_{bc}+\Gamma^e{}_{ac}\Gamma^d{}_{be}-\Gamma^e{}_{bc}\Gamma^d{}_{ae}. \tag{1.71}$$

The action on tangent vectors can be found as follows. Let t^a be an arbitrary tangent vector. The scalar product $t^a\omega_a$ is a function on the manifold and thus

$$(\nabla_a\nabla_b-\nabla_b\nabla_a)(t^c\omega_c)=0. \tag{1.72}$$

This leads immediately to

$$(\nabla_a\nabla_b-\nabla_b\nabla_a)t^d=-R_{abc}{}^dt^c \tag{1.73}$$

Generalization of this result and the previous one to higher order tensors is given by the following equation

$$(\nabla_a\nabla_b-\nabla_b\nabla_a)T^{d_1...d_k}{}_{c_1...c_l}=-\sum_{i=1}^k R_{abe}{}^{d_i}T^{d_1...e...d_k}{}_{c_1...c_l}+\sum_{i=1}^l R_{abc_i}{}^eT^{d_1...d_k}{}_{c_1...e...c_l}. \tag{1.74}$$

Properties: We state without proof the following properties of the curvature tensor:

- Anti-symmetry in the first two indices:

$$R_{abc}{}^d=-R_{bac}{}^d. \tag{1.75}$$

- Anti-symmetrization of the first three indices yields 0:

$$R_{[abc]}{}^d=0\;,\;R_{[abc]}{}^d=\frac{1}{3}(R_{abc}{}^d+R_{cab}{}^d+R_{bca}{}^d). \tag{1.76}$$

- Anti-symmetry in the last two indices:

$$R_{abcd}=-R_{abdc}\;,\;R_{abcd}=R_{abc}{}^eg_{ed}. \tag{1.77}$$

- Symmetry if the pair consisting of the first two indices is exchanged with the pair consisting of the last two indices:

$$R_{abcd}=R_{cdab}. \tag{1.78}$$

- Bianchi identity:

$$\nabla_{[a}R_{bc]d}{}^{e} = 0 \ , \ \nabla_{[a}R_{bc]d}{}^{e} = \frac{1}{3}(\nabla_a R_{bcd}{}^{e} + \nabla_c R_{abd}{}^{e} + \nabla_b R_{cad}{}^{e}). \quad (1.79)$$

- The so-called Ricci tensor R_{ac}, which is the trace part of the Riemann curvature tensor, is symmetric, viz

$$R_{ac} = R_{ca} \ , \ R_{ac} = R_{abc}{}^{b}. \quad (1.80)$$

- The Einstein tensor can be constructed as follows. By contracting the Bianchi identity and using $\nabla_a g_{bc} = 0$ we get

$$g_e{}^{c}(\nabla_a R_{bcd}{}^{e} + \nabla_c R_{abd}{}^{e} + \nabla_b R_{cad}{}^{e}) = 0 \Rightarrow \nabla_a R_{bd} + \nabla_e R_{abd}{}^{e} - \nabla_b R_{ad} = 0. \quad (1.81)$$

  By contracting now the two indices b and d we get

$$g^{bd}(\nabla_a R_{bd} + \nabla_e R_{abd}{}^{e} - \nabla_b R_{ad}) = 0 \Rightarrow \nabla_a R - 2\nabla_b R_a{}^{b} = 0. \quad (1.82)$$

  This can be put in the form

$$\nabla^a G_{ab} = 0. \quad (1.83)$$

  The tensor G_{ab} is called Einstein tensor and is given by

$$G_{ab} = R_{ab} - \frac{1}{2}g_{ab}R. \quad (1.84)$$

  The so-called scalar curvature R is defined by

$$R = R_a{}^{a}. \quad (1.85)$$

1.5 The Stress-Energy-Momentum Tensor

1.5.1 The Stress-Energy-Momentum Tensor

We will mostly be interested in continuous matter distributions which are extended macroscopic systems composed of a large number of individual particles. We will think of such systems as fluids. The energy, momentum and pressure of fluids are encoded in the stress-energy-momentum tensor $T^{\mu\nu}$ which is a symmetric tensor of type $(2,0)$. The component $T^{\mu\nu}$ of the stress-energy-momentum tensor is defined as the flux of the component p^μ of the $4-$vector energy-momentum across a surface of constant x^ν.

Let us consider an infinitesimal element of the fluid in its rest frame. The spatial diagonal component T^{ii} is the flux of the momentum p^i across a surface of constant x^i, i.e. it is the amount of momentum p^i per unit time per unit area traversing the surface of constant x^i. Thus

T^{ii} is the normal stress which we also call pressure when it is independent of direction. We write $T^{ii}=P_i$. The spatial off-diagonal component T^{ij} is the flux of the momentum p^i across a surface of constant x^j, i.e. it is the amount of momentum p^i per unit time per unit area traversing the surface of constant x^j which means that T^{ij} is the shear stress.

The component T^{00} is the flux of the energy p^0 through the surface of constant x^0, i.e. it is the amount of energy per unit volume at a fixed instant of time. Thus T^{00} is the energy density, viz $T^{00}=\rho c^2$ where ρ is the rest-mass density. Similarly, T^{i0} is the flux of the momentum p^i through the surface of constant x^0, i.e. it is the i momentum density times c. The T^{0i} is the energy flux through the surface of constant x^i divided by c. They are equal by virtue of the symmetry of the stress-energy-momentum tensor, viz $T^{0i}=T^{i0}$.

1.5.2 Perfect Fluid

We begin with the case of "dust" which is a collection of a large number of particles in spacetime at rest with respect to each other. The particles are assumed to have the same rest mass m. The pressure of the dust is obviously 0 in any direction since there is no motion of the particles, i.e. the dust is a pressureless fluid. The $4-$vector velocity of the dust is the constant $4-$vector velocity U^μ of the individual particles. Let n be the number density of the particles, i.e. the number of particles per unit volume as measured in the rest frame. Clearly $N^i=nU^i=n(\gamma u_i)$ is the flux of the particles, i.e. the number of particles per unit area per unit time in the x^i direction. The $4-$vector number-flux of the dust is defined by

$$N^\mu=nU^\mu. \tag{1.86}$$

The rest-mass density of the dust in the rest frame is clearly given by $\rho=nm$. This rest-mass density times c^2 is the $\mu=0$, $\nu=0$ component of the stress-energy-momentum tensor $T^{\mu\nu}$ in the rest frame. We remark that $\rho c^2=nmc^2$ is also the $\mu=0$, $\nu=0$ component of the tensor $N^\mu p^\nu$ where N^μ is the $4-$vector number-flux and p^μ is the $4-$vector energy-momentum of the dust. We define therefore the stress-energy-momentum tensor of the dust by

$$T^{\mu\nu}=N^\mu p^\nu=(nm)U^\mu U^\nu=\rho U^\mu U^\nu. \tag{1.87}$$

The next fluid of paramount importance is the so-called perfect fluid. This is a fluid determined completely by its energy density ρ and its isotropic pressure P in the rest frame. Hence $T^{00}=\rho c^2$ and $T^{ii}=P$. The shear stresses T^{ij} $(i\neq j)$ are absent for a perfect fluid in its rest frame. It is not difficult to convince ourselves that stress-energy-momentum tensor $T^{\mu\nu}$ is given in this case in the rest frame by

$$T^{\mu\nu}=\rho U^\mu U^\nu+\frac{P}{c^2}(c^2\eta^{\mu\nu}+U^\mu U^\nu)=(\rho+\frac{P}{c^2})U^\mu U^\nu+P\eta^{\mu\nu}. \tag{1.88}$$

This is a covariant equation and thus it must also hold, by the principle of minimal coupling (see below), in any other global inertial reference frame. We give the following examples:

- Dust: $P=0$.

- Gas of Photons: $P=\rho c^2/3$.
- Vacuum Energy: $P=-\rho c^2 \Leftrightarrow T^{ab}=-\rho c^2\eta^{ab}$.

1.5.3 Conservation Law

The stress-energy-momentum tensor $T^{\mu\nu}$ is symmetric, viz $T^{\mu\nu}=T^{\nu\mu}$. It must also be conserved, i.e.

$$\partial_\mu T^{\mu\nu}=0. \tag{1.89}$$

This should be thought of as the equation of motion of the perfect fluid. Explicitly this equation reads

$$\partial_\mu T^{\mu\nu}=\partial_\mu(\rho+\frac{P}{c^2}).U^\mu U^\nu+(\rho+\frac{P}{c^2})(\partial_\mu U^\mu.U^\nu+U^\mu\partial_\mu U^\nu)+\partial^\nu P=0. \tag{1.90}$$

We project this equation along the 4−vector velocity by contracting it with U_ν. We get (using $U_\nu\partial_\mu U^\nu=0$)

$$\partial_\mu(\rho U^\mu)+\frac{P}{c^2}\partial_\mu U^\mu=0. \tag{1.91}$$

We project the above equation along a direction orthogonal to the 4−vector velocity by contracting it with $P^\mu{}_\nu$ given by

$$P^\mu{}_\nu=\delta^\mu_\nu+\frac{U^\mu U_\nu}{c^2}. \tag{1.92}$$

Indeed, we can check that $P^\mu{}_\nu P^\nu{}_\lambda=P^\mu{}_\lambda$ and $P^\mu{}_\nu U^\nu=0$. By contracting equation (1.90) with $P^\lambda{}_\nu$ we obtain

$$(\rho+\frac{P}{c^2})U^\mu\partial_\mu U_\lambda+(\eta_{\nu\lambda}+\frac{U_\nu U_\lambda}{c^2})\partial^\nu P=0. \tag{1.93}$$

We consider now the non-relativistic limit defined by

$$U^\mu=(c,u_i)\ ,\ |u_i|<<1\ ,\ P<<\rho c^2. \tag{1.94}$$

The parallel equation (1.91) becomes the continuity equation given by

$$\partial_t\rho+\vec{\nabla}(\rho\vec{u})=0. \tag{1.95}$$

The orthogonal equation (1.93) becomes Euler's equation of fluid mechanics given by

$$\rho(\partial_t\vec{u}+(\vec{u}\vec{\nabla})\vec{u})=-\vec{\nabla}P. \tag{1.96}$$

1.5.4 Minimal Coupling

The laws of physics in general relativity can be derived from the laws of physics in special relativity by means of the so-called principle of minimal coupling. This consists in writing the laws of physics in special relativity in tensor form and then replacing the flat metric $\eta_{\mu\nu}$ with the curved metric $g_{\mu\nu}$ and the derivative operator ∂_μ with the covariant derivative operator ∇_μ. This recipe works in most cases.

For example take the geodesic equation describing a free particle in special relativity given by $U^\mu\partial_\mu U^\nu = 0$. Geodesic motion in general relativity is given by $U^\mu\nabla_\mu U^\nu = 0$. These are the geodesics of the curved metric $g_{\mu\nu}$ and they describe freely falling bodies in the corresponding gravitational field.

The second example is the equation of motion of a perfect fluid in special relativity which is given by the conservation law $\partial^\nu T_{\nu\lambda} = 0$. In general relativity this conservation law becomes

$$\nabla^\nu T_{\nu\lambda} = 0. \tag{1.97}$$

Also, by applying the principle of minimal coupling, the stress-energy-momentum tensor $T_{\mu\nu}$ of a perfect fluid in general relativity is given by equation (1.88) with the replacement $\eta \longrightarrow g$, viz

$$T_{\mu\nu} = (\rho + \frac{P}{c^2})U_\mu U_\nu + P g_{\mu\nu}. \tag{1.98}$$

1.6 Einstein's Equation

Although local gravitational forces can not be measured by the principle of equivalence, i.e. since the spacetime manifold is locally flat, relative gravitational forces, the so-called tidal gravitational forces, can still be measured by observing the relative acceleration of nearby geodesics. This effect is described by the geodesic deviation equation.

1.6.1 Tidal Gravitational Forces

Let us first start by describing tidal gravitational forces in Newtonian physics. The force of gravity exerted by an object of mass M on a particle of mass m a distance r away is $\vec{F} = -\hat{r}GMm/r^2$ where $\hat{r}$ is the unit vector pointing from M to m and r is the distance between the center of M and m. The corresponding acceleration is $\vec{a} = -\hat{r}GM/r^2 = -\vec{\nabla}\Phi$, $\Phi = -GM/r$. We assume now that the mass m is spherical of radius Δr. The distance between the center of M and the center of m is r. The force of gravity exerted by the mass M on a particle of mass dm a distance $r \pm \Delta r$ away on the line joining the centers of M and m is given by $\vec{F} = -\hat{r}GMdm/(r \pm \Delta r)^2$. The corresponding acceleration is

$$\vec{a} = -\hat{r}GM\frac{1}{(r+\Delta r)^2} = -\hat{r}GM\frac{1}{r^2} + \hat{r}GM\frac{2\Delta r}{r^3} + ... \tag{1.99}$$

The first term is precisely the acceleration experienced at the center of the body m due to M. This term does not affect the observed acceleration of particles on the surface of m. In other words, since m and everything on its surface are in a state of free fall with respect to M, the acceleration of dm with respect to m is precisely the so-called tidal acceleration, and is given by the second term in the above expansion, viz

$$\begin{aligned} \vec{a}_t &= \hat{r}GM\frac{2\Delta r}{r^3}+... \\ &= -(\Delta\vec{r}.\vec{\nabla})(\vec{\nabla}\Phi). \end{aligned} \tag{1.100}$$

1.6.2 Geodesic Deviation Equation

In a flat Euclidean geometry two parallel lines remain always parallel. This is not true in a curved manifold. To see this more carefully we consider a one-parameter family of geodesics $\gamma_s(t)$ which are initially parallel and see what happens to them as we move along these geodesics when we increase the parameter t. The map $(t,s)\longrightarrow\gamma_s(t)$ is smooth, one-to-one, and its inverse is smooth, which means in particular that the geodesics do not cross. These geodesics will then generate a $2-$dimensional surface on the manifold M. The parameters t and s can therefore be chosen to be the coordinates on this surface. This surface is given by the entirety of the points $x^\mu(s,t)\in M$. The tangent vector to the geodesics is defined by

$$T^\mu=\frac{\partial x^\mu}{\partial t}. \tag{1.101}$$

This satisfies therefore the equation $T^\mu\nabla_\mu T^\nu=0$. The so-called deviation vector is defined by

$$S^\mu=\frac{\partial x^\mu}{\partial s}. \tag{1.102}$$

The product $S^\mu ds$ is the displacement vector between two infinitesimally nearby geodesics. The vectors T^μ and S^μ commute because they are basis vectors. Hence we must have $[T,S]^\mu= T^\nu\nabla_\nu S^\mu-S^\nu\nabla_\nu T^\mu=0$ or equivalently

$$T^\nu\nabla_\nu S^\mu=S^\nu\nabla_\nu T^\mu. \tag{1.103}$$

This can be checked directly by using the definition of the covariant derivative and the way it acts on tangent vectors and equations (1.101) and (1.102).

The quantity $V^\mu=T^\nu\nabla_\nu S^\mu$ expresses the rate of change of the deviation vector along a geodesic. We will call V^μ the relative velocity of infinitesimally nearby geodesics. Similarly the relative acceleration of infinitesimally nearby geodesics is defined by $A^\mu=T^\nu\nabla_\nu V^\mu$. We compute

$$\begin{aligned}
A^{\mu} &= T^{\nu}\nabla_{\nu}V^{\mu} \\
&= T^{\nu}\nabla_{\nu}(T^{\lambda}\nabla_{\lambda}S^{\mu}) \\
&= T^{\nu}\nabla_{\nu}(S^{\lambda}\nabla_{\lambda}T^{\mu}) \\
&= (T^{\nu}\nabla_{\nu}S^{\lambda}).\nabla_{\lambda}T^{\mu} + T^{\nu}S^{\lambda}\nabla_{\nu}\nabla_{\lambda}T^{\mu} \\
&= (S^{\nu}\nabla_{\nu}T^{\lambda}).\nabla_{\lambda}T^{\mu} + T^{\nu}S^{\lambda}(\nabla_{\lambda}\nabla_{\nu}T^{\mu} - R_{\nu\lambda\sigma}{}^{\mu}T^{\sigma}) \\
&= S^{\lambda}\nabla_{\lambda}(T^{\nu}\nabla_{\nu}T^{\mu}) - R_{\nu\lambda\sigma}{}^{\mu}T^{\nu}S^{\lambda}T^{\sigma} \\
&= -R_{\nu\lambda\sigma}{}^{\mu}T^{\nu}S^{\lambda}T^{\sigma}. \qquad (1.104)
\end{aligned}$$

This is the geodesic deviation equation. The relative accelaration of infinitesimally nearby geodesics is 0 if and only if $R_{\nu\lambda\sigma}{}^{\mu} = 0$. Geodesics will accelerate towards, or away from, each other if and only if $R_{\nu\lambda\sigma}{}^{\mu} \neq 0$. Thus initially parallel geodesics with $V^{\mu} = 0$ will fail generically to remain parallel.

1.6.3 Einsetin's Equation

We will assume that, in general relativity, the tidal acceleration of two nearby particles is precisely the relative acceleration of infinitesimally nearby geodesics given by equation (1.104), viz

$$\begin{aligned}
A^{\mu} &= -R_{\nu\lambda\sigma}{}^{\mu}T^{\nu}S^{\lambda}T^{\sigma} \\
&= -R_{\nu\lambda\sigma}{}^{\mu}U^{\nu}\Delta x^{\lambda}U^{\sigma}. \qquad (1.105)
\end{aligned}$$

This suggest, by comparing with (1.100), we make the following correspondence

$$R_{\nu\lambda\sigma}{}^{\mu}U^{\nu}U^{\sigma} \leftrightarrow \partial_{\lambda}\partial^{\mu}\Phi. \qquad (1.106)$$

Thus

$$R_{\nu\mu\lambda}{}^{\mu}U^{\nu}U^{\lambda} \leftrightarrow \Delta\Phi. \qquad (1.107)$$

By using the Poisson's equation (1.1) we get then the correspondence

$$R_{\nu\mu\sigma}{}^{\mu}U^{\nu}U^{\sigma} \leftrightarrow 4\pi G\rho. \qquad (1.108)$$

From the other hand, the stress-energy-momentum tensor $T^{\mu\nu}$ provides the correspondence

$$T_{\nu\sigma}U^{\nu}U^{\sigma} \leftrightarrow \rho c^4. \qquad (1.109)$$

We expect therefore an equation of the form

$$\frac{R_{\nu\mu\sigma}{}^{\mu}}{4\pi G} = \frac{T_{\nu\sigma}}{c^4} \Leftarrow R_{\nu\sigma} = \frac{4\pi G}{c^4}T_{\nu\sigma}. \qquad (1.110)$$

This is the original equation proposed by Einstein. However, it has the following problem. From the fact that $\nabla^\nu G_{\nu\sigma}=0$, we get immediately $\nabla^\nu R_{\nu\sigma}=\nabla_\sigma R/2$, and as a consequence $\nabla^\nu T_{\nu\sigma}=c^4\nabla_\sigma R/8\pi G$. This result is in direct conflict with the requirement of the conservation of the stress-energy-momentum tensor given by $\nabla^\nu T_{\nu\sigma}=0$. An immediate solution is to consider instead the equation

$$G_{\nu\sigma}=R_{\nu\sigma}-\frac{1}{2}g_{\nu\sigma}R=\frac{8\pi G}{c^4}T_{\nu\sigma}. \tag{1.111}$$

The conservation of the stress-energy-momentum tensor is now guaranteed. Furthermore, this equation is still in accord with the correspondence $R_{\nu\sigma}U^\nu U^\sigma \leftrightarrow 8\pi G\rho$. Indeed, by using the result $R=-4\pi GT/c^4$ we can rewrite the above equation as

$$R_{\nu\sigma}=\frac{8\pi G}{c^4}(T_{\nu\sigma}-\frac{1}{2}g_{\nu\sigma}T). \tag{1.112}$$

We compute $R_{\mu\nu}U^\mu U^\nu=(8\pi G/c^4)(T_{\mu\nu}U^\mu U^\nu+c^2T/2)$. By keeping only the $\mu=0$, $\nu=0$ component of $T_{\mu\nu}$ and neglecting the other components the right hand side is exactly $4\pi G\rho$ as it should be.

1.6.4 Newtonian Limit

The Newtonian limit of general relativity is defined by the following three requirements:

1) The particles are moving slowly compared with the speed of light.

2) The gravitational field is weak so that the curved metric can be expanded about the flat metric.

3) The gravitational field is static.

Geodesic Equation: We begin with the geodesic equation, with the proper time τ as the parameter of the geodesic, is

$$\Gamma^\rho{}_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}+\frac{d^2x^\rho}{d\tau^2}=0. \tag{1.113}$$

The assumption that particles are moving slowly compared to the speed of light means that

$$|\frac{d\vec{x}}{d\tau}|<<c|\frac{dt}{d\tau}|. \tag{1.114}$$

The geodesic equation becomes

$$c^2\Gamma^\rho{}_{00}(\frac{dt}{d\tau})^2+\frac{d^2x^\rho}{d\tau^2}=0. \tag{1.115}$$

We recall the Christoffel symbols

$$\Gamma^d{}_{ab}=\frac{1}{2}g^{dc}(\partial_a g_{bc}+\partial_b g_{ac}-\partial_c g_{ab}). \tag{1.116}$$

Since the gravitational field is static we have

$$\Gamma^d{}_{00} = -\frac{1}{2}g^{dc}\partial_c g_{00}. \tag{1.117}$$

The second assumption that the gravitational field is weak allows us to decompose the metric as

$$g_{ab} = \eta_{ab} + h_{ab}\ ,\ |h_{ab}| << 1. \tag{1.118}$$

Thus

$$\Gamma^d{}_{00} = -\frac{1}{2}\eta^{dc}\partial_c h_{00}. \tag{1.119}$$

The geodesic equation becomes

$$\frac{d^2x^\rho}{d\tau^2} = \frac{c^2}{2}\eta^{dc}\partial_c h_{00}(\frac{dt}{d\tau})^2. \tag{1.120}$$

In terms of components this reads

$$\frac{d^2x^0}{d\tau^2} = \frac{c^2}{2}\eta^{00}\partial_0 h_{00}(\frac{dt}{d\tau})^2 = 0. \tag{1.121}$$

$$\frac{d^2x^i}{d\tau^2} = \frac{c^2}{2}\eta^{ii}\partial_i h_{00}(\frac{dt}{d\tau})^2 = \frac{c^2}{2}\partial_i h_{00}(\frac{dt}{d\tau})^2. \tag{1.122}$$

The first equation says that $dt/d\tau$ is a constant. The second equation reduces to

$$\frac{d^2x^i}{dt^2} = \frac{c^2}{2}\partial_i h_{00} = -\partial_i\Phi\ ,\ h_{00} = -\frac{2\Phi}{c^2}. \tag{1.123}$$

Einstein's Equations: Now we turn to the Newtonian limit of Einstein's equation $R_{\nu\sigma} = 8\pi G(T_{\nu\sigma} - \frac{1}{2}g_{\nu\sigma}T)/c^4$ with the stress-energy-momentum tensor $T_{\mu\nu}$ of a perfect fluid as a source. The perfect fluid is describing the Earth or the Sun. The stress-energy-momentum tensor is given by $T_{\mu\nu} = (\rho + P/c^2)U_\mu U_\nu + Pg_{\mu\nu}$. In the Newtonian limit this can be approximated by the stress-energy-momentum tensor of dust given by $T_{\mu\nu} = \rho U_\mu U_\nu$ since in this limit pressure can be neglected as it comes from motion which is assumed to be slow. In the rest frame of the perfect fluid we have $U^\mu = (U^0, 0, 0, 0)$ and since $g_{\mu\nu}U^\mu U^\nu = -c^2$ we get $U^0 = c(1 + h_{00}/2)$ and $U_0 = c(-1 + h_{00}/2)$ and as a consequence

$$T^{00} = \rho c^2(1 + h_{00})\ ,\ T_{00} = \rho c^2(1 - h_{00}). \tag{1.124}$$

The inverse metric is obviously given by $g^{00} = -1 - h_{00}$ since $g^{\mu\nu}g_{\nu\rho} = \delta^\mu_\rho$. Hence

$$T = -\rho c^2. \tag{1.125}$$

The $\mu=0$, $\nu=0$ component of Einstein's equation is therefore

$$R_{00}=\frac{4\pi G}{c^2}\rho(1-h_{00}). \tag{1.126}$$

We recall the Riemann curvature tensor and the Ricci tensor

$$R_{\mu\nu\sigma}{}^{\lambda} = \partial_\nu\Gamma^\lambda{}_{\mu\sigma}-\partial_\mu\Gamma^\lambda{}_{\nu\sigma}+\Gamma^\delta{}_{\mu\sigma}\Gamma^\lambda{}_{\nu\delta}-\Gamma^\delta{}_{\nu\sigma}\Gamma^\lambda{}_{\mu\delta}. \tag{1.127}$$

$$R_{\mu\sigma}=R_{\mu\nu\sigma}{}^{\nu}. \tag{1.128}$$

Thus (using in particular $R_{000}{}^0=0$)

$$R_{00}=R_{0i0}{}^i=\partial_i\Gamma^i{}_{00}-\partial_0\Gamma^i{}_{i0}+\Gamma^e{}_{00}\Gamma^i{}_{ie}-\Gamma^e{}_{i0}\Gamma^i{}_{0e}. \tag{1.129}$$

The Christoffel symbols are linear in the metric perturbation and thus one can neglect the third and fourth terms in the above equation. We get then

$$R_{00}=\partial_i\Gamma^i{}_{00}=-\frac{1}{2}\Delta h_{00}. \tag{1.130}$$

Einstein's equation reduces therefore to Newton's equation, viz

$$-\frac{1}{2}\Delta h_{00}=\frac{4\pi G}{c^2}\rho\Rightarrow\Delta\Phi=4\pi G\rho. \tag{1.131}$$

1.7 Killing Vectors and Maximally Symmetric Spaces

A spacetime which is spatially homogeneous and spatially isotropic is a spacetime in which the space is maximally symmetric. A maximally symmetric space is a space with the maximum number of isometries, i.e. the maximum number of symmetries of the metric. These isometries are generated by the so-called Killing vectors.

As an example, if $\partial_\sigma g_{\mu\nu}=0$, for some fixed value of σ, then the translation $x^\sigma\longrightarrow x^\sigma+a^\sigma$ is a symmetry and thus it is an isometry of the curved manifold M with metric $g_{\mu\nu}$. This symmetry will be naturally associated with a conserved quantity. To see this let us first recall that the geodesic equation can be rewritten in terms of the $4-$vector energy-momentum $p^\mu=mU^\mu$ as $p^\mu\nabla_\mu p_\nu=0$. Explicitly

$$\begin{aligned}m\frac{dp_\nu}{dt} &= \Gamma^\lambda{}_{\mu\nu}p^\mu p_\lambda\\ &= \frac{1}{2}\partial_\nu g_{\mu\rho}.p^\mu p^\rho.\end{aligned} \tag{1.132}$$

Thus if the metric is invariant under the translation $x^\sigma\longrightarrow x^\sigma+a^\sigma$ then $\partial_\sigma g_{\mu\nu}=0$ and as a consequence the momentum p_σ is conserved as expected.

For obvious reasons we must rewrite the condition which expresses the symmetry under $x^\sigma \longrightarrow x^\sigma + a^\sigma$ in a covariant fashion. Let us thus introduce the vector $K = \partial_{(\sigma)}$ via its components which are given (in the basis in which $\partial_\sigma g_{\mu\nu} = 0$) by

$$K^\mu = (\partial_{(\sigma)})^\mu = \delta^\mu_\sigma. \tag{1.133}$$

Clearly then $p_\sigma = p_\mu K^\mu$. Since $\partial_\sigma g_{\mu\nu} = 0$ we must have $dp_\sigma/dt = 0$ or equivalently $d(p_\mu K^\mu)/dt = 0$. This means that the directional derivative of the scalar quantity $p_\mu K^\mu$ along the geodesic is 0, viz

$$p^\nu \nabla_\nu (p_\mu K^\mu) = 0. \tag{1.134}$$

We compute

$$p^\nu \nabla_\nu (p_\mu K^\mu) = p^\mu p^\nu \nabla_\mu K_\nu = \frac{1}{2} p^\mu p^\nu (\nabla_\mu K_\nu + \nabla_\nu K_\mu). \tag{1.135}$$

We obtain therefore the so-called Killing equation

$$\nabla_\mu K_\nu + \nabla_\nu K_\mu = 0. \tag{1.136}$$

Thus for any vector K which satisfies the Killing equation $\nabla_\mu K_\nu + \nabla_\nu K_\mu = 0$ the momentum $p_\mu K^\mu$ is conserved along the geodesic with tangent p. The vector K is called a Killing vector. The Killing vector K generates the isometry which is associated with the conservation of $p_\mu K^\mu$. The symmetry transformation under which the metric is invariant is expressed as infinitesimal motion in the direction of K.

Let us check that the vector $K^\mu = \delta^\mu_\sigma$ satisfies the Killing equation. Immediately, we have $K_\mu = g_{\mu\sigma}$ and

$$\begin{aligned} \nabla_\mu K_\nu + \nabla_\nu K_\mu &= \partial_\mu g_{\nu\sigma} + \partial_\nu g_{\mu\sigma} - 2\Gamma^\rho{}_{\mu\nu} g_{\rho\sigma} \\ &= \partial_\sigma g_{\mu\nu} \\ &= 0. \end{aligned} \tag{1.137}$$

Thus if the metric is independent of x^σ then the vector $K^\mu = \delta^\mu_\sigma$ will satisfy the Killing equation. Conversely if a vector satisfies the killing equation then one can always find a basis in which the vector satisfies $K^\mu = \delta^\mu_\sigma$. However, if we have more than one Killing vector we can not find a single basis in which all of them satisfy $K^\mu = \delta^\mu_\sigma$.

Some of the properties of Killing vectors are:

$$\nabla_\mu \nabla_\nu K^\lambda = R_{\nu\mu\rho}{}^{\lambda} K^\rho. \tag{1.138}$$

$$\nabla_\mu \nabla_\nu K^\mu = R_{\nu\mu} K^\mu. \tag{1.139}$$

$$K^\mu \nabla_\mu R = 0. \tag{1.140}$$

The last identity in particular shows explicitly that the geometry does not change under a Killing vector.

The isometries of R^n with flat Euclidean metric are n independent translations and $n(n-1)/2$ independent rotations (which form the group of $SO(n)$ rotations). Hence R^n with flat Euclidean metric has $n + n(n-1)/2 = n(n+1)/2$ isometries. This is the number of Killing vectors on R^n with flat Euclidean metric which is the maximum possible number of isometries in n dimensions. The space R^n is therefore called maximally symmetric space. In general a maximally symmetric space is any space with $n(n+1)/2$ Killing vectors (isometries). These spaces have the maximum degree of symmetry. The only Euclidean maximally symmetric spaces are planes R^n with 0 scalar curvature, spheres S^n with positive scalar curvature and hyperboloids H^n with negative scalar curvature [1].

The curvature of a maximally symmetric space must be the same everywhere (translations) and the same in every direction (rotations). More precisely, a maximally symmetric space must be locally fully characterized by a constant scalar curvature R and furthermore must look like the same in all directions, i.e. it must be invariant under all Lorentz transformations at the point of consideration.

In the neighborhood of a point $p \in M$ we can always choose an inertial reference frame in which $g_{\mu\nu} = \eta_{\mu\nu}$. This is invariant under Lorentz transformations at p. Since the space is maximally symmetric the Riemann curvature tensor $R_{\mu\nu\lambda\rho}$ at p must also be invariant under Lorentz transformations at p. This tensor must therefore be constructed from $\eta_{\mu\nu}$, the Kronecker delta $\delta_{\mu\nu}$ and the Levi-Civita tensor $\epsilon_{\mu\nu\lambda\rho}$ which are the only tensors which are known to be invariant under Lorentz transformations. However, the curvature tensor satisfies $R_{\mu\nu\lambda\gamma} = -R_{\nu\mu\lambda\gamma}$, $R_{\mu\nu\lambda\gamma} = -R_{\mu\nu\gamma\lambda}$, $R_{\mu\nu\lambda\gamma} = R_{\lambda\gamma\mu\nu}$, $R_{[\mu\nu\lambda]\gamma} = 0$ and $\nabla_{[\mu} R_{\nu\lambda]\gamma\rho} = 0$. The only combination formed out of $\eta_{\mu\nu}$, $\delta_{\mu\nu}$ and $\epsilon_{\mu\nu\lambda\rho}$ which satisfies these identities is $R_{\mu\nu\lambda\gamma} = \kappa(\eta_{\mu\lambda}\eta_{\nu\gamma} - \eta_{\mu\gamma}\eta_{\nu\lambda})$ with κ a constant. This tensorial relation must hold in any other coordinate system, viz

$$R_{\mu\nu\lambda\gamma} = \kappa(g_{\mu\lambda}g_{\nu\gamma} - g_{\mu\gamma}g_{\nu\lambda}). \tag{1.141}$$

We compute $R_{\mu\nu\lambda}{}^{\gamma} = \kappa(g_{\mu\lambda}\delta^\gamma_\nu - \delta^\gamma_\mu g_{\nu\lambda})$, $R_{\mu\lambda} = R_{\mu\nu\lambda}{}^{\nu} = \kappa(n-1)g_{\mu\lambda}$ and hence $R = \kappa n(n-1)$. In other words the scalar curvature of a maximally symmetric space is a constant over the manifold. Thus the curvature of a maximally symmetric space must be of the form

$$R_{\mu\nu\lambda\gamma} = \frac{R}{n(n-1)}(g_{\mu\lambda}g_{\nu\gamma} - g_{\mu\gamma}g_{\nu\lambda}). \tag{1.142}$$

Conversely if the curvature tensor is given by this equation with R constant over the manifold then the space is maximally symmetric.

[1]The corresponding maximally symmetric Lorentzian spaces are Minkowski spaces M^n $(R = 0)$, de Sitter spaces dS^n $(R > 0)$ and Anti-de Sitter spaces AdS^n $(R < 0)$.

1.8 The Hilbert-Einstein Action

The Einstein's equations for general relativity reads

$$R_{\mu\nu}-\frac{1}{2}g_{\mu\nu}R=8\pi G T_{\mu\nu}. \tag{1.143}$$

The dynamical variable is obviously the metric $g_{\mu\nu}$. The goal is to construct an action principle from which the Einstein's equations follow as the Euler-Lagrange equations of motion for the metric. This action principle will read as

$$S=\int d^nx\ \mathcal{L}(g). \tag{1.144}$$

The first problem with this way of writing is that both d^nx and $\mathcal{L}$ are tensor densities rather than tensors. We digress briefly to explain this important different.

Let us recall the familiar Levi-Civita symbol in n dimensions defined by

$$\begin{aligned}\tilde{\epsilon}_{\mu_1...\mu_n} &= +1 \text{ even permutation}\\ &= -1 \text{ odd permutation}\\ &= \ \ 0 \text{ otherwise.}\end{aligned} \tag{1.145}$$

This is a symbol and not a tensor since it does not change under coordinate transformations, The determinant of a matrix M can be given by the formula

$$\tilde{\epsilon}_{\nu_1...\nu_n}\mathrm{det}M = \tilde{\epsilon}_{\mu_1...\mu_n}M^{\mu_1}{}_{\nu_1}...M^{\mu_n}{}_{\nu_n}. \tag{1.146}$$

By choosing $M^{\mu}{}_{\nu}=\partial x^{\mu}/\partial y^{\nu}$ we get the transformation law

$$\tilde{\epsilon}_{\nu_1...\nu_n} = \mathrm{det}\frac{\partial y}{\partial x}\ \tilde{\epsilon}_{\mu_1...\mu_n}\frac{\partial x^{\mu_1}}{\partial y^{\nu_1}}...\frac{\partial x^{\mu_n}}{\partial y^{\nu_n}}. \tag{1.147}$$

In other words $\tilde{\epsilon}_{\mu_1...\mu_n}$ is not a tensor because of the determinant appearing in this equation. This is an example of a tensor density. Another example of a tensor density is $\mathrm{det}g$. Indeed from the tensor transformation law of the metric $g^{'}_{\alpha\beta}=g_{\mu\nu}(\partial x^{\mu}/\partial y^{\alpha})(\partial x^{\nu}/\partial y^{\beta})$ we can show in a straightforward way that

$$\mathrm{det}g^{'} = (\mathrm{det}\frac{\partial y}{\partial x})^{-2}\ \mathrm{det}g. \tag{1.148}$$

The actual Levi-Civita tensor can then be defined by

$$\epsilon_{\mu_1...\mu_n}=\sqrt{\mathrm{det}g}\ \tilde{\epsilon}_{\mu_1...\mu_n}. \tag{1.149}$$

Next under a coordinate transformation $x\longrightarrow y$ the volume element transforms as

$$d^nx\longrightarrow d^ny=\mathrm{det}\frac{\partial y}{\partial x}\ d^nx. \tag{1.150}$$

In other words the volume element transforms as a tensor density and not as a tensor. We verify this important point in our language as follows. We write

$$\begin{aligned} d^nx &= dx^0\wedge dx^1\wedge ...\wedge dx^{n-1}\\ &= \frac{1}{n!}\tilde{\epsilon}_{\mu_1...\mu_n}dx^{\mu_1}\wedge ...\wedge dx^{\mu_n}. \end{aligned} \tag{1.151}$$

Recall that a differential $p-$form is a $(0,p)$ tensor which is completely antisymmetric. For example scalars are $0-$forms and dual cotangent vectors are $1-$forms. The Levi-Civita tensor $\epsilon_{\mu_1...\mu_n}$ is a $4-$form. The differentials dx^{μ} appearing in the second line of equation (1.151) are $1-$forms and hence under a coordinate transformation $x\longrightarrow y$ we have $dx^{\mu}\longrightarrow dy^{\mu}=dx^{\nu}\partial y^{\mu}/\partial x^{\nu}$. By using this transformation law we can immediately show that dx^n transforms to d^ny exactly as in equation (1.150).

It is not difficult to see now that an invariant volume element can be given by the $n-$form defined by the equation

$$dV=\sqrt{\det g}\ d^nx. \tag{1.152}$$

We can show that

$$\begin{aligned} dV &= \frac{1}{n!}\sqrt{\det g}\ \tilde{\epsilon}_{\mu_1...\mu_n}dx^{\mu_1}\wedge ...\wedge dx^{\mu_n}\\ &= \frac{1}{n!}\epsilon_{\mu_1...\mu_n}dx^{\mu_1}\wedge ...\wedge dx^{\mu_n}\\ &= \epsilon_{\mu_1...\mu_n}dx^{\mu_1}\otimes ...\otimes dx^{\mu_n}\\ &= \epsilon. \end{aligned} \tag{1.153}$$

In other words the invariant volume element is precisely the Levi-Civita tensor. In the case of Lorentzian signature we replace $\det g$ with $-\det g$.

We go back now to equation (1.144) and rewrite it as

$$\begin{aligned} S &= \int d^nx\ \mathcal{L}(g)\\ &= \int d^nx\sqrt{-\det g}\ \hat{\mathcal{L}}(g). \end{aligned} \tag{1.154}$$

Clearly $\mathcal{L}=\sqrt{-\det g}\ \hat{\mathcal{L}}$. Since the invariant volume element $d^nx\sqrt{-\det g}$ is a scalar the function $\hat{\mathcal{L}}$ must also be a scalar and as such can be identified with the Lagrangian density.

We use the result that the only independent scalar quantity which is constructed from the metric and which is at most second order in its derivatives is the Ricci scalar R. In other words the simplest choice for the Lagrangian density $\hat{\mathcal{L}}$ is

$$\hat{\mathcal{L}}(g)=R. \tag{1.155}$$

The corresponding action is called the Hilbert-Einstein action. We compute

$$\delta S = \int d^nx\delta\sqrt{-\det g}\ g^{\mu\nu}R_{\mu\nu}+\int d^nx\sqrt{-\det g}\ \delta g^{\mu\nu}R_{\mu\nu}+\int d^nx\sqrt{-\det g}\ g^{\mu\nu}\delta R_{\mu\nu}. \tag{1.156}$$

We have

$$\begin{aligned}
\delta R_{\mu\nu} &= \delta R_{\mu\rho\nu}{}^{\rho}\\
&= \partial_\rho \delta\Gamma^\rho{}_{\mu\nu} - \partial_\mu\delta\Gamma^\rho{}_{\rho\nu} + \delta(\Gamma^\lambda{}_{\mu\nu}\Gamma^\rho{}_{\rho\lambda} - \Gamma^\lambda{}_{\rho\nu}\Gamma^\rho{}_{\mu\lambda})\\
&= (\nabla_\rho\delta\Gamma^\rho{}_{\mu\nu} - \Gamma^\rho{}_{\rho\lambda}\delta\Gamma^\lambda{}_{\mu\nu} + \Gamma^\lambda{}_{\rho\mu}\delta\Gamma^\rho{}_{\lambda\nu} + \Gamma^\lambda{}_{\rho\nu}\delta\Gamma^\rho{}_{\lambda\mu}) - (\nabla_\mu\delta\Gamma^\rho{}_{\rho\nu} - \Gamma^\rho{}_{\mu\lambda}\delta\Gamma^\lambda{}_{\rho\nu} + \Gamma^\lambda{}_{\mu\rho}\delta\Gamma^\rho{}_{\lambda\nu}\\
&+ \Gamma^\lambda{}_{\mu\nu}\delta\Gamma^\rho{}_{\rho\lambda}) + \delta(\Gamma^\lambda{}_{\mu\nu}\Gamma^\rho{}_{\rho\lambda} - \Gamma^\lambda{}_{\rho\nu}\Gamma^\rho{}_{\mu\lambda})\\
&= \nabla_\rho\delta\Gamma^\rho{}_{\mu\nu} - \nabla_\mu\delta\Gamma^\rho{}_{\rho\nu}.
\end{aligned} \tag{1.157}$$

In the second line of the above equation we have used the fact that $\delta\Gamma^\rho{}_{\mu\nu}$ is a tensor since it is the difference of two connections. Thus

$$\begin{aligned}
\int d^nx\sqrt{-\det g}\; g^{\mu\nu}\delta R_{\mu\nu} &= \int d^nx\sqrt{-\det g}\; g^{\mu\nu}\bigg(\nabla_\rho\delta\Gamma^\rho{}_{\mu\nu} - \nabla_\mu\delta\Gamma^\rho{}_{\rho\nu}\bigg)\\
&= \int d^nx\sqrt{-\det g}\;\nabla_\rho\bigg(g^{\mu\nu}\delta\Gamma^\rho{}_{\mu\nu} - g^{\rho\nu}\delta\Gamma^\mu{}_{\mu\nu}\bigg).
\end{aligned} \tag{1.158}$$

We compute also (with $\delta g_{\mu\nu} = -g_{\mu\alpha}g_{\nu\beta}\delta g^{\alpha\beta}$)

$$\begin{aligned}
\delta\Gamma^\rho{}_{\mu\nu} &= \frac{1}{2}g^{\rho\lambda}\bigg(\nabla_\mu\delta g_{\nu\lambda} + \nabla_\nu\delta g_{\mu\lambda} - \nabla_\lambda\delta g_{\mu\nu}\bigg)\\
&= -\frac{1}{2}\bigg(g_{\nu\lambda}\nabla_\mu\delta g^{\lambda\rho} + g_{\mu\lambda}\nabla_\nu\delta g^{\lambda\rho} - g_{\mu\alpha}g_{\nu\beta}\nabla^\rho\delta g^{\alpha\beta}\bigg).
\end{aligned} \tag{1.159}$$

Thus

$$\int d^nx\sqrt{-\det g}\; g^{\mu\nu}\delta R_{\mu\nu} = \int d^nx\sqrt{-\det g}\;\nabla_\rho\bigg(g_{\mu\nu}\nabla^\rho\delta g^{\mu\nu} - \nabla_\mu\delta g^{\mu\rho}\bigg). \tag{1.160}$$

By Stokes's theorem this integral is equal to the integral over the boundary of spacetime of the expression $g_{\mu\nu}\nabla^\rho\delta g^{\mu\nu} - \nabla_\mu\delta g^{\mu\rho}$ which is 0 if we assume that the metric and its first derivatives are held fixed on the boundary. The variation of the action reduces to

$$\delta S = \int d^nx\delta\sqrt{-\det g}\; g^{\mu\nu}R_{\mu\nu} + \int d^nx\sqrt{-\det g}\;\delta g^{\mu\nu}R_{\mu\nu}. \tag{1.161}$$

Next we use the result

$$\delta\sqrt{-\det g} = -\frac{1}{2}\sqrt{-\det g}\; g_{\mu\nu}\delta g^{\mu\nu}. \tag{1.162}$$

Hence

$$\delta S = \int d^nx\sqrt{-\det g}\;\delta g^{\mu\nu}(R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R). \tag{1.163}$$

This will obviously lead to Einstein's equations in vacuum which is partially our goal. We want also to include the effect of matter which requires considering the more general actions of the form

$$S = \frac{1}{16\pi G}\int d^nx\;\sqrt{-\det g}\; R + S_M. \tag{1.164}$$

$$S_M = \int d^n x \ \sqrt{-\det g} \ \hat{\mathcal{L}}_M. \tag{1.165}$$

The variation of the action becomes

$$\begin{aligned} \delta S &= \frac{1}{16\pi G} \int d^n x \sqrt{-\det g} \ \delta g^{\mu\nu} (R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R) + \delta S_M \\ &= \int d^n x \sqrt{-\det g} \ \delta g^{\mu\nu} \bigg[\frac{1}{16\pi G} (R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R) + \frac{1}{\sqrt{-\det g}} \frac{\delta S_M}{\delta g^{\mu\nu}} \bigg]. \end{aligned} \tag{1.166}$$

In other words

$$\frac{1}{\sqrt{-\det g}} \frac{\delta S}{\delta g^{\mu\nu}} = \frac{1}{16\pi G} (R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R) + \frac{1}{\sqrt{-\det g}} \frac{\delta S_M}{\delta g^{\mu\nu}}. \tag{1.167}$$

Einstein's equations are therefore given by

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R = 8\pi G T_{\mu\nu}. \tag{1.168}$$

The stress-energy-momentum tensor must therefore be defined by the equation

$$T_{\mu\nu} = -\frac{2}{\sqrt{-\det g}} \frac{\delta S_M}{\delta g^{\mu\nu}}. \tag{1.169}$$

As a first example we consider the action of a scalar field in curved spacetime given by

$$S_\phi = \int d^n x \sqrt{-\det g} \ \bigg[-\frac{1}{2} g^{\mu\nu} \nabla_\mu \phi \nabla_\nu \phi - V(\phi) \bigg]. \tag{1.170}$$

The corresponding stress-energy-momentum tensor is calculated to be given by

$$T_{\mu\nu}^{(\phi)} = \nabla_\mu \phi \nabla_\nu \phi - \frac{1}{2} g_{\mu\nu} g^{\rho\sigma} \nabla_\rho \phi \nabla_\sigma \phi - g_{\mu\nu} V(\phi). \tag{1.171}$$

As a second example we consider the action of the electromagnetic field in curved spacetime given by

$$S_A = \int d^n x \sqrt{-\det g} \ \bigg[-\frac{1}{4} g^{\mu\nu} g^{\alpha\beta} F_{\mu\nu} F_{\alpha\beta} \bigg]. \tag{1.172}$$

In this case the stress-energy-momentum tensor is calculated to be given by

$$T_{\mu\nu}^{(A)} = F^{\mu\lambda} F^{\nu}{}_{\lambda} - \frac{1}{4} g^{\mu\nu} F_{\alpha\beta} F^{\alpha\beta}. \tag{1.173}$$

Chapter 2

Black Holes

2.1 Spherical Star

2.1.1 The Schwarzschild Metric

We consider a matter source which is both static and spherically symmetric. Clearly a static source means that the components of the metric are all independent of time. By requiring also that the physics is invariant under time reversal, i.e. under $t \longrightarrow -t$, the components g_{0i} which provide space-time cross terms in the metric must be absent. We have already found that the most general spherically symmetric metric in $3-$dimension is of the form

$$d\vec{u}^2 = e^{2\beta(r)}dr^2 + r^2 d\Omega^2. \tag{2.1}$$

The most general static and spherically symmetric metric in $4-$dimension is therefore of the form

$$ds^2 = -e^{2\alpha(r)}c^2dt^2 + d\vec{u}^2 = -e^{2\alpha(r)}c^2dt^2 + e^{2\beta(r)}dr^2 + r^2 d\Omega^2. \tag{2.2}$$

We need to determine the functions $\alpha(r)$ and $\beta(r)$ from solving Einstein's equations. First we need to evaluate the Christoffel symbols. We find

$$\begin{aligned}
&\Gamma^0{}_{0r} = \partial_r \alpha \\
&\Gamma^r{}_{00} = \partial_r \alpha e^{2(\alpha-\beta)} \ , \ \Gamma^r{}_{rr} = \partial_r \beta \ , \ \Gamma^r{}_{\theta\theta} = -re^{-2\beta} \ , \ \Gamma^r{}_{\phi\phi} = -re^{-2\beta}\sin^2\theta \\
&\Gamma^\theta{}_{r\theta} = \frac{1}{r} \ , \ \Gamma^\theta{}_{\phi\phi} = -\sin\theta\cos\theta \\
&\Gamma^\phi{}_{r\phi} = \frac{1}{r} \ , \ \Gamma^\phi{}_{\theta\phi} = \frac{\cos\theta}{\sin\theta}.
\end{aligned} \tag{2.3}$$

The non-zero components of the Riemann curvature tensor are

$$\begin{aligned}
R_{0rr}{}^0 &= -R_{r0r}{}^0 = \partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\beta\partial_r\alpha \\
R_{0\theta\theta}{}^0 &= -R_{\theta0\theta}{}^0 = re^{-2\beta}\partial_r\alpha \\
R_{0\phi\phi}{}^0 &= -R_{\phi0\phi}{}^0 = re^{-2\beta}\partial_r\alpha\sin^2\theta.
\end{aligned} \tag{2.4}$$

$$\begin{aligned}
R_{0r0}{}^{r} &= -R_{r00}{}^{r} = (\partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\beta\partial_r\alpha)e^{2(\alpha-\beta)}\\
R_{r\theta\theta}{}^{r} &= -R_{\theta r\theta}{}^{r} = -re^{-2\beta}\partial_r\beta\\
R_{r\phi\phi}{}^{r} &= -R_{\phi r\phi}{}^{r} = -re^{-2\beta}\partial_r\beta\sin^2\theta.
\end{aligned}\tag{2.5}$$

$$\begin{aligned}
R_{0\theta 0}{}^{\theta} &= -R_{\theta 00}{}^{\theta} = \frac{1}{r}\partial_r\alpha e^{2(\alpha-\beta)}\\
R_{r\theta r}{}^{\theta} &= -R_{\theta rr}{}^{\theta} = \frac{1}{r}\partial_r\beta\\
R_{\theta\phi\phi}{}^{\theta} &= -R_{\phi\theta\phi}{}^{\theta} = \sin^2\theta(e^{-2\beta}-1).
\end{aligned}\tag{2.6}$$

$$\begin{aligned}
R_{0\phi 0}{}^{\phi} &= -R_{\phi 00}{}^{\phi} = \frac{1}{r}\partial_r\alpha e^{2(\alpha-\beta)}\\
R_{r\phi r}{}^{\phi} &= -R_{\phi rr}{}^{\phi} = \frac{1}{r}\partial_r\beta\\
R_{\theta\phi\theta}{}^{\phi} &= -R_{\phi\theta\theta}{}^{\phi} = 1-e^{-2\beta}.
\end{aligned}\tag{2.7}$$

We compute immediately the non-zero components of the Ricci tensor as follows

$$\begin{aligned}
R_{00} &= R_{0r0}{}^{r} + R_{0\theta 0}{}^{\theta} + R_{0\phi 0}{}^{\phi} = \bigg(\partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\beta\partial_r\alpha + \frac{2}{r}\partial_r\alpha\bigg)e^{2(\alpha-\beta)}\\
R_{rr} &= R_{r0r}{}^{0} + R_{r\theta r}{}^{\theta} + R_{r\phi r}{}^{\phi} = -\partial_r^2\alpha - (\partial_r\alpha)^2 + \partial_r\beta\partial_r\alpha + \frac{2}{r}\partial_r\beta\\
R_{\theta\theta} &= R_{\theta 0\theta}{}^{0} + R_{\theta r\theta}{}^{r} + R_{\theta\phi\theta}{}^{\phi} = e^{-2\beta}\bigg(r\partial_r\beta - r\partial_r\alpha - 1\bigg) + 1\\
R_{\phi\phi} &= R_{\phi 0\phi}{}^{0} + R_{\phi r\phi}{}^{r} + R_{\phi\theta\phi}{}^{\theta} = \sin^2\theta\bigg[e^{-2\beta}\bigg(r\partial_r\beta - r\partial_r\alpha - 1\bigg) + 1\bigg].
\end{aligned}\tag{2.8}$$

We compute also the scalar curvature

$$R = -2e^{-2\beta}\bigg(\partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\beta\partial_r\alpha + \frac{2}{r}(\partial_r\alpha - \partial_r\beta) + \frac{1}{r^2}(1-e^{2\beta})\bigg).\tag{2.9}$$

Now we are in a position to solve Einstein's equations outside the static spherical source (the star). In the absence of any other matter sources in the region outside the star the Einstein's equations read

$$R_{\mu\nu} = 0.\tag{2.10}$$

We have immediately three independent equations

$$\begin{aligned}
&\partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\beta\partial_r\alpha + \frac{2}{r}\partial_r\alpha = 0\\
&\partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\beta\partial_r\alpha - \frac{2}{r}\partial_r\beta = 0\\
&e^{-2\beta}\bigg(r\partial_r\beta - r\partial_r\alpha - 1\bigg) + 1 = 0.
\end{aligned}\tag{2.11}$$

By subtracting the first two conditions we get $\partial_r(\alpha+\beta)=0$ and thus $\alpha=-\beta+c$ where c is some constant. By an appropriate rescaling of the time coordinate we can redefine the value of α as $\alpha+c^{'}$ where $c^{'}$ is an arbitrary constant. We can clearly choose this constant such that $\alpha=-\beta$. The third condition in the above equation (2.11) becomes then

$$e^{2\alpha}\big(2r\partial_r\alpha+1\big)=1. \tag{2.12}$$

Equivalently

$$\partial_r(re^{2\alpha})=1. \tag{2.13}$$

The solution is (with R_s is some constant)

$$e^{2\alpha}=1-\frac{R_s}{r}. \tag{2.14}$$

The first and the second conditions in equation (2.11) take now the form

$$\partial_r^2\alpha+2(\partial_r\alpha)^2+\frac{2}{r}\partial_r\alpha=0 \tag{2.15}$$

We compute

$$\partial_r\alpha=\frac{R_s}{2(r^2-R_sr)}\ ,\ \partial_r^2\alpha=-\frac{R_s(2r-R_s)}{2(r^2-R_sr)^2}. \tag{2.16}$$

In other words the form (2.14) is indeed a solution.

The Schwarzschild metric is the metric corresponding to this solution. This is the most important spacetime after Minkowski spacetime. It reads explicitly

$$ds^2=-(1-\frac{R_s}{r})c^2dt^2+(1-\frac{R_s}{r})^{-1}dr^2+r^2d\Omega^2. \tag{2.17}$$

In the Newtonian limit we know that (with Φ the gravitational potential and M the mass of the spherical star)

$$g_{00}=-(1+2\frac{\Phi}{c^2})=-(1-\frac{2GM}{c^2r}). \tag{2.18}$$

The g_{00} component of the Schwarzschild metric should reduce to this form for very large distances which here means $r>>R_s$. By comparison we obtain

$$R_s=\frac{2GM}{c^2}. \tag{2.19}$$

This is called the Schwarzschild radius. We stress that M can be thought of as the mass of the star only in the weak field limit. In general M will also include gravitational binding energy. In the limit $M\longrightarrow 0$ or $r\longrightarrow\infty$ the Schwarzschild metric reduces to the Minkowski metric. This is called asymptotic flatness.

The powerful Birkhoff's theorem states that the Schwarzschild metric is the unique vacuum solution (static or otherwise) to Einstein's equations which is spherically symmetric [1].

We remark that the Schwarzschild metric is singular at $r=0$ and at $r=R_s$. However only the singularity at $r=0$ is a true singularity of the geometry. For example we can check that the scalar quantity $R^{\mu\nu\alpha\beta}R_{\mu\nu\alpha\beta}$ is divergent at $r=0$ whereas it is perfectly finite at $r=R_s$ [2]. Indeed the divergence of the Ricci scalar or any other higher order scalar such as $R^{\mu\nu\alpha\beta}R_{\mu\nu\alpha\beta}$ at a point is a sufficient condition for that point to be singular. We say that $r=0$ is an essential singularity.

The Schwarzschild radius $r=R_s$ is not a true singularity of the metric and its appearance as such only reflects the fact that the chosen coordinates are behaving badly at $r=R_s$. We say that $r=R_s$ is a coordinate singularity. Indeed it should appear like any other point if we choose a more appropriate coordinates system. It will, on the other hand, specify the so-called event horizon when the spherical sphere becomes a black hole.

2.1.2 Particle Motion in Schwarzschild Spacetime

We start by rewriting the Christoffel symbols (2.3) as

$$\begin{aligned}
&\Gamma^0{}_{0r} = \frac{R_s}{2r(r-R_s)} \\
&\Gamma^r{}_{00} = \frac{R_s(r-R_s)}{2r^3}\ ,\ \Gamma^r{}_{rr} = -\frac{R_s}{2r(r-R_s)}\ ,\ \Gamma^r{}_{\theta\theta} = -r+R_s\ ,\ \Gamma^r{}_{\phi\phi} = (-r+R_s)\sin^2\theta \\
&\Gamma^\theta{}_{r\theta} = \frac{1}{r}\ ,\ \Gamma^\theta{}_{\phi\phi} = -\sin\theta\cos\theta \\
&\Gamma^\phi{}_{r\phi} = \frac{1}{r}\ ,\ \Gamma^\phi{}_{\theta\phi} = \frac{\cos\theta}{\sin\theta}.
\end{aligned} \tag{2.20}$$

The geodesic equation is given by

$$\frac{d^2x^\rho}{d\lambda^2} + \Gamma^\rho{}_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = 0. \tag{2.21}$$

Explicitly we have

$$\frac{d^2x^0}{d\lambda^2} + \frac{R_s}{r(r-R_s)}\frac{dx^0}{d\lambda}\frac{dr}{d\lambda} = 0. \tag{2.22}$$

$$\frac{d^2r}{d\lambda^2} + \frac{R_s(r-R_s)}{2r^3}\Big(\frac{dx^0}{d\lambda}\Big)^2 - \frac{R_s}{2r(r-R_s)}\Big(\frac{dr}{d\lambda}\Big)^2 - (r-R_s)\Big[\Big(\frac{d\theta}{d\lambda}\Big)^2 + \sin^2\theta\Big(\frac{d\phi}{d\lambda}\Big)^2\Big] = 0. \tag{2.23}$$

$$\frac{d^2\theta}{d\lambda^2} + \frac{2}{r}\frac{dr}{d\lambda}\frac{d\theta}{d\lambda} - \sin\theta\cos\theta\Big(\frac{d\phi}{d\lambda}\Big)^2 = 0. \tag{2.24}$$

[1] Exercise: Try to prove this theorem. This is quite difficult so it is better to consult references right away.
[2] Exercise: Verify this explicitly.

$$\frac{d^2\phi}{d\lambda^2}+\frac{2}{r}\frac{dr}{d\lambda}\frac{d\phi}{d\lambda}+2\frac{\cos\theta}{\sin\theta}\frac{d\theta}{d\lambda}\frac{d\phi}{d\lambda}=0. \tag{2.25}$$

The Schwarzschild metric is obviously invariant under time translations and space rotations. There will therefore be 4 corresponding Killing vectors K_μ and 4 conserved quantities given by

$$Q=K_\mu\frac{dx^\mu}{d\lambda}. \tag{2.26}$$

The motion of a particle under a central force of gravity in flat spacetime has invariance under time translation which leads to conservation of energy and invariance under rotations which leads to conservation of angular momentum. The angular momentum is a vector in 3 dimensions with a length (one component) and a direction (two angles). Conservation of the direction means that the motion happens in a plane. In other words we can choose $\theta=\pi/2$.

In Schwarzschild spacetime the same symmetries are still present and therefore the four Killing vectors K_μ must be associated with time translation and rotations and the four conserved quantities Q are the energy and the angular momentum. The two Killing vectors associated with the conservation of the direction of the angular momentum lead precisely, as in the flat case, to a motion in the plane, viz

$$\theta=\frac{\pi}{2}. \tag{2.27}$$

The metric is independent of x^0 and ϕ and hence the corresponding Killing vectors are

$$K^\mu=(\partial_{x^0})^\mu=\delta^\mu_0=(1,0,0,0)\ ,\ K_\mu=g_{\mu 0}=(-(1-\frac{R_s}{r}),0,0,0). \tag{2.28}$$

$$R^\mu=(\partial_\phi)^\mu=\delta^\mu_\phi=(0,0,0,1)\ ,\ R_\mu=g_{\mu\phi}=(0,0,0,r^2\sin^2\theta). \tag{2.29}$$

The corresponding conserved quantities are the energy and the magnitude of the angular momentum given by

$$E=-K_\mu\frac{dx^\mu}{d\lambda}=(1-\frac{R_s}{r})\frac{dx^0}{d\lambda}. \tag{2.30}$$

$$L=R_\mu\frac{dx^\mu}{d\lambda}=r^2\sin^2\theta\frac{d\phi}{d\lambda}. \tag{2.31}$$

The minus sign in the energy is consistent with the definition $p_\sigma=p_\mu K^\mu=cp_\mu(\partial_{(\sigma)})^\mu$. Furthermore E is actually the energy per unit mass for a massive particle whereas for a massless particles it is indeed the energy since the momentum of a massless particle is identified with its $4-$vector velocity. A similar remark applies to the angular momentum. Note that E should be thought of as the total energy including gravitational energy which is the quantity that really needs to be conserved. In other words E is different from the kinetic energy $-p_av^a$ which is the energy measured by an observer whose velocity is v^a. Note also that the conservation of angular momentum is precisely Kepler's 2nd law.

There is an extra conserved quantity along the geodesic given by

$$\epsilon=-g_{\mu\nu}\frac{dx^{\mu}}{d\lambda}\frac{dx^{\nu}}{d\lambda}. \tag{2.32}$$

We compute

$$\begin{aligned}
\frac{d\epsilon}{d\lambda} &= -\frac{dg_{\mu\nu}}{d\lambda}.\frac{dx^{\mu}}{d\lambda}\frac{dx^{\nu}}{d\lambda}-2g_{\mu\nu}\frac{d^2x^{\mu}}{d\lambda^2}\frac{dx^{\nu}}{d\lambda}\\
&= -\frac{dg_{\mu\nu}}{d\lambda}.\frac{dx^{\mu}}{d\lambda}\frac{dx^{\nu}}{d\lambda}+2g_{\mu\nu}\Gamma^{\mu}{}_{\alpha\beta}\frac{dx^{\alpha}}{d\lambda}\frac{dx^{\beta}}{d\lambda}\frac{dx^{\nu}}{d\lambda}\\
&= -\frac{dx^{\alpha}}{d\lambda}\frac{dx^{\beta}}{d\lambda}\frac{dx^{\rho}}{d\lambda}\bigg[\partial_{\rho}g_{\alpha\beta}-2\Gamma^{\mu}{}_{\alpha\beta}g_{\mu\rho}\bigg]\\
&= -\frac{dx^{\alpha}}{d\lambda}\frac{dx^{\beta}}{d\lambda}\frac{dx^{\rho}}{d\lambda}\bigg[\partial_{\rho}g_{\alpha\beta}-\Gamma^{\mu}{}_{\alpha\rho}g_{\mu\beta}-\Gamma^{\mu}{}_{\rho\beta}g_{\mu\alpha}\bigg]\\
&= -\frac{dx^{\alpha}}{d\lambda}\frac{dx^{\beta}}{d\lambda}\frac{dx^{\rho}}{d\lambda}\nabla_{\rho}g_{\alpha\beta}\\
&= 0.
\end{aligned} \tag{2.33}$$

We clearly need to take

$$\epsilon=c^2\ ,\ \text{massive particle}. \tag{2.34}$$

$$\epsilon=0\ ,\ \text{massless particle}. \tag{2.35}$$

The above conserved quantity reads explicitly

$$\epsilon=\frac{E^2}{1-\frac{R_s}{r}}-\frac{1}{1-\frac{R_s}{r}}\Big(\frac{dr}{d\lambda}\Big)^2-\frac{L^2}{r^2}. \tag{2.36}$$

Equivalently

$$E^2-\Big(\frac{dr}{d\lambda}\Big)^2-(1-\frac{R_s}{r})(\frac{L^2}{r^2}+\epsilon)=0. \tag{2.37}$$

We also rewrite this as

$$\frac{1}{2}\Big(\frac{dr}{d\lambda}\Big)^2+V(r)=\mathcal{E}. \tag{2.38}$$

$$\mathcal{E}=\frac{1}{2}(E^2-\epsilon). \tag{2.39}$$

$$\begin{aligned}
V(r) &= \frac{1}{2}(1-\frac{R_s}{r})(\frac{L^2}{r^2}+\epsilon)-\frac{\epsilon}{2}\\
&= -\frac{\epsilon GM}{c^2r}+\frac{L^2}{2r^2}-\frac{GML^2}{c^2r^3}.
\end{aligned} \tag{2.40}$$

This is the equation of a particle with unit mass and energy $\mathcal{E}$ in a potential $V(r)$. In this potential only the last term is new compared to Newtonian gravity. Clearly when $r \longrightarrow 0$ this potential will go to $-\infty$ whereas if the last term is absent (the case of Newtonian gravity) the potential will go to $+\infty$ when $r \longrightarrow 0$. See figure GR1a.

The potential $V(r)$ is different for different values of L. It has one maximum and one minimum if $cL/GM > \sqrt{12}$. Indeed we have

$$\frac{dV(r)}{dr} = 0 \Leftrightarrow \epsilon r^2 - \frac{c^2L^2}{GM}r + 3L^2 = 0. \tag{2.41}$$

For massive particles the stable (minimum) and unstable (maximum) orbits are located at

$$r_{\rm max} = \frac{L^2 - \sqrt{L^4 - \frac{12G^2M^2L^2}{c^2}}}{2GM} \ , \ r_{\rm min} = \frac{L^2 + \sqrt{L^4 - \frac{12G^2M^2L^2}{c^2}}}{2GM}. \tag{2.42}$$

Both orbits are circular. See figure GR1b. In the limit $L \longrightarrow \infty$ we obtain

$$r_{\rm max} = \frac{3GM}{c^2} \ , \ r_{\rm min} = \frac{L^2}{GM}. \tag{2.43}$$

The stable circular orbit becomes farther away whereas the unstable circular orbit approaches $3GM/c^2$.

In the limit of small L, the two orbits coincide when

$$L^4 - \frac{12G^2M^2L^2}{c^2} = 0 \Leftrightarrow L = \sqrt{12}\frac{GM}{c}. \tag{2.44}$$

At which point

$$r_{\rm max} = r_{\rm min} = \frac{L^2}{2GM} = \frac{6GM}{c^2}. \tag{2.45}$$

This is the smallest radius possible of a stable circular orbit in a Schwarzschild spacetime.

For massless particles ($\epsilon = 0$) there is a solution at $r = 3GM/c^2$. This corresponds always to unstable circular orbit. We have then the following criterion

$$\text{stable circular orbits} \ : \ r > \frac{6GM}{c^2}. \tag{2.46}$$

$$\text{unstable circular orbits} \ : \ \frac{3GM}{c^3} < r < \frac{6GM}{c^2}. \tag{2.47}$$

These are of course all geodesics, i.e. orbits corresponding to free fall in a gravitational field. There are also bound non-circular orbits which oscillates around the stable circular orbit. For example if a test particle starts from a point $r_{\rm max} < r_2 < r_{\rm min}$ at which $\mathcal{E} = V(r_2) < 0$ it will move in the potential until it hits the potential at a point $r_1 > r_{\rm min}$ at which $\mathcal{E} = V(r_1)$ where it bounces back. The corresponding bound precessing orbit is shown on figure GR1c.

There exists also scattering orbits. If a test particle comes from infinity with energy $\mathcal{E}>0$ then it will move in the potential and may hit the wall of the potential at $r_{\rm max}<r_2<r_{\rm min}$ for which $\mathcal{E}=V(r_2)>0$. If it does not hit the wall of the potential (the energy $\mathcal{E}$ is sufficiently large) then the particle will plunge into the center of the potential at $r=0$. See figures GR1d and GR1e.

In contrast to Newtonian gravity these orbits do not correspond to conic section as we will show next.

2.1.3 Precession of Perihelia and Gravitational Redshift

Precession of Perihelia The equation for the conservation of angular momentum reads

$$L=r^2\frac{d\phi}{d\lambda}. \tag{2.48}$$

Together with the radial equation

$$\frac{1}{2}\Big(\frac{dr}{d\lambda}\Big)^2+V(r)=\mathcal{E}. \tag{2.49}$$

We have for a massive particle the equation

$$\Big(\frac{dr}{d\phi}\Big)^2+\frac{c^2r^4}{L^2}-\frac{2GMr^3}{L^2}+r^2-\frac{2GMr}{c^2}=\frac{r^4E^2}{L^2}. \tag{2.50}$$

In the case of Newtonian gravity equation (2.41) for a massive particle gives $r=L^2/GM$. This is the radius of a circular orbit in Newtonian gravity. We perform the change of variable

$$x=\frac{L^2}{GMr}. \tag{2.51}$$

The above last differential equation becomes

$$\Big(\frac{dx}{d\phi}\Big)^2+\frac{L^2c^2}{G^2M^2}-2x+x^2-\frac{2G^2M^2x^3}{L^2c^2}=\frac{L^2E^2}{G^2M^2}. \tag{2.52}$$

We differentiate this equation with respect to x to get

$$\frac{d^2x}{d\phi^2}-1+x=\frac{3G^2M^2}{L^2c^2}x^2. \tag{2.53}$$

We solve this equation in perturbation theory as follows. We write

$$x=x_0+x_1. \tag{2.54}$$

The 0th order equation is

$$\frac{d^2x_0}{d\phi^2}-1+x_0=0. \tag{2.55}$$

The 1st order equation is

$$\frac{d^2x_1}{d\phi^2}+x_1=\frac{3G^2M^2}{L^2c^2}x_0^2. \tag{2.56}$$

The solution to the 0th order equation is precisely the Newtonian result

$$x_0=1+e\cos\phi. \tag{2.57}$$

This is an ellipse with eccentricity $e=c/a=\sqrt{1-b^2/a^2}$ with the center of the coordinate system at the focus $(c,0)$ and ϕ is the angle measured from the major axis [3]. The semi-major axis a is the distance to the farthest point whereas the semi-minor axis b is the distance to the closest point. In other words at $\phi=\pi$ we have $x_0=1-e=a(1-e^2)/(a+c)$ and at $\phi=0$ we have $x_0=1+e=a(1-e^2)/(a-c)$. By comparing also the equation of the ellipse $a(1-e^2)/r=1+e\cos\phi$ with the solution for x_0 we obtain the value of the angular momentum

$$L^2=GMa(1-e^2). \tag{2.62}$$

The 1st order equation becomes

$$\begin{aligned}\frac{d^2x_1}{d\phi^2}+x_1 &= \frac{3G^2M^2}{L^2c^2}(1+e\cos\phi)^2\\ &= \frac{3G^2M^2}{L^2c^2}(1+\frac{e^2}{2}+\frac{e^2}{2}\cos 2\phi+2e\cos\phi).\end{aligned} \tag{2.63}$$

[3] The ellipse is the set of points where the sum of the distances r_1 and r_2 from each point on the ellipse to two fixed points (the foci) is a constant equal $2a$. We have then

$$r_1+r_2=2a. \tag{2.58}$$

Let $2c$ be the distance between the two foci F_1 and F_2 and let O be the middle point of the segment $[F_1,F_2]$. The coordinates of each point on the ellipse are x and y with respect to the Cartesian system with O at the origin. Clearly then $r_1=\sqrt{(c+x)^2+y^2}$ and $r_2=\sqrt{(c-x)^2+y^2}$. The equation of the ellipse becomes

$$\frac{x^2}{a^2}+\frac{y^2}{b^2}=1 \tag{2.59}$$

The semi-major axis is a and the semi-minor axis is $b=\sqrt{a^2-c^2}$. We take the focus F_2 as the center of our system of coordinates and we use polar coordinates. Then $x=r\cos\theta-c$ and $y=r\sin\theta$ and hence the equation of the ellipse becomes (with eccentricity $e=c/a$)

$$\frac{a(1-e^2)}{r}=1-e\cos\theta. \tag{2.60}$$

If we had taken the focus F_1 instead as the center of our system of coordinates we would have obtained

$$\frac{a(1-e^2)}{r}=1+e\cos\theta. \tag{2.61}$$

Remark that

$$\frac{d^2}{d\phi^2}(\phi\sin\phi)+\phi\sin\phi=2\cos\phi$$
$$\frac{d^2}{d\phi^2}(\cos 2\phi)+\cos 2\phi=-3\cos 2\phi. \tag{2.64}$$

Then we can write

$$\frac{d^2y_1}{d\phi^2}+y_1 = \frac{3G^2M^2}{L^2c^2}(1+\frac{e^2}{2})\ ,\ y_1=x_1-\frac{3G^2M^2}{L^2c^2}(-\frac{e^2}{6}\cos 2\phi+e\phi\sin\phi). \tag{2.65}$$

Define also

$$z_1=\frac{y_1}{\frac{3G^2M^2}{L^2c^2}(1+\frac{e^2}{2})}. \tag{2.66}$$

The differential equations becomes

$$\frac{d^2z_1}{d\phi^2}-1+z_1=0 \tag{2.67}$$

The solution is immediately given by

$$z_1=1+e\cos\phi\Leftrightarrow x_1=\frac{3G^2M^2}{L^2c^2}(1+\frac{e^2}{2})(1+e\cos\phi)+\frac{3G^2M^2}{L^2c^2}(-\frac{e^2}{6}\cos 2\phi+e\phi\sin\phi). \tag{2.68}$$

The complete solution is

$$x=\bigg[1+\frac{3G^2M^2}{L^2c^2}(1+\frac{e^2}{2})\bigg](1+e\cos\phi)+\frac{3G^2M^2}{L^2c^2}(-\frac{e^2}{6}\cos 2\phi+e\phi\sin\phi). \tag{2.69}$$

We can rewrite this in the form

$$x=\bigg[1+\frac{3G^2M^2}{L^2c^2}(1+\frac{e^2}{2})\bigg](1+e\cos(1-\alpha)\phi)+\frac{3G^2M^2}{L^2c^2}(-\frac{e^2}{6}\cos 2\phi). \tag{2.70}$$

The small number α is given by

$$\alpha=\frac{3G^2M^2}{L^2c^2}. \tag{2.71}$$

The last term in the above solution oscillates around 0 and hence averages to 0 over successive revolutions and as such it is irrelevant to our consideration.

The above result can be interpreted as follows. The orbit is an ellipse but with a period equal $2\pi/(1-\alpha)$ instead of 2π. Thus the perihelion advances in each revolution by the amount

$$\Delta\phi=2\pi\alpha=\frac{6\pi G^2M^2}{L^2c^2}. \tag{2.72}$$

By using now the value of the angular momentum for a perfect ellipse given by equation (2.62) we get

$$\Delta\phi = \frac{6\pi GM}{a(1-e^2)c^2}. \tag{2.73}$$

In the case of the motion of Mercury around the Sun we can use the values

$$\frac{GM}{c^2} = 1.48\times 10^5 \mathrm{cm}\ ,\ a = 5.79\times 10^{12}\mathrm{cm}\ ,\ e = 0.2056. \tag{2.74}$$

We obtain

$$\Delta\phi_{\mathrm{Mercury}} = \frac{6\pi GM}{a(1-e^2)c^2} = 5.03\times 10^{-7}\ \mathrm{rad/orbit}. \tag{2.75}$$

However Mercury completes one orbit each 88 days thus in a century its perihelion will advance by the amount

$$\begin{aligned}\Delta\phi_{\mathrm{Mercury}} &= \frac{100\times 365}{88}\times 5.03\times 10^{-7}\frac{180\times 3600}{3.14}\ \mathrm{arcsecond/century}\\ &= 43.06\ \mathrm{arcsecond/century}.\end{aligned} \tag{2.76}$$

The total precession of Mercury is around 575 arcseconds per century[4] with a 532 arcseconds per century due to other planets and 43 arcseconds per century due to the curvature of spacetime caused by the Sun[5].

Gravitational Redshift We consider a stationary observer ($U^i = 0$) in Schwarzschild spacetime. The 4−vector velocity satisfies $g_{\mu\nu}U^\mu U^\nu = -c^2$ and hence

$$U^0 = \frac{c}{\sqrt{1-\frac{2GM}{c^2r}}}. \tag{2.77}$$

The energy (per unit mass) of a photon as measured by this observer is

$$\begin{aligned}E_\gamma &= -U_\mu\frac{dx^\mu}{d\lambda}\\ &= c^2\sqrt{1-\frac{2GM}{c^2r}}\frac{dt}{d\lambda}\\ &= \frac{cE}{\sqrt{1-\frac{2GM}{c^2r}}}.\end{aligned} \tag{2.78}$$

The E^2 is the conserved energy (per unit mass) of the Schwarzschild metric given by (2.30). Thus a photon emitted with an energy $E_{\gamma 1}$ at a distance r_1 will be observed at a distance $r_2 > r_1$ with an energy $E_{\gamma 2}$ given by

$$\frac{E_{\gamma 2}}{E_{\gamma 1}} = \sqrt{\frac{1-\frac{2GM}{c^2r_1}}{1-\frac{2GM}{c^2r_2}}} < 1. \tag{2.79}$$

[4]There is a huge amount of precession due to the precession of equinoxes which is not discussed here.

[5]There is also a minute contribution due to the oblatness of the Sun

Thus the energy $E_{\gamma 2} < E_{\gamma 1}$, i.e. as the photon climbs out of the gravitational field it gets redshifted. In other words the frequency decreases as the strength of the gravitational field decreases or equivalently as the gravitational potential increases. This is the gravitational redshift. In the limit $r >> 2GM/c^2$ the formula becomes

$$\frac{E_{\gamma 2}}{E_{\gamma 1}} = 1+\frac{\Phi_1}{c^2}-\frac{\Phi_2}{c^2} \ , \ \Phi=-\frac{GM}{r}. \tag{2.80}$$

2.1.4 Free Fall

For a radially (vertically) freely object we have $d\phi/d\lambda=0$ and thus the angular momentum is 0, viz $L=0$. The radial equation of motion becomes

$$(\frac{dr}{d\lambda})^2-\frac{2GM}{r}=E^2-c^2. \tag{2.81}$$

This is essentially the Newtonian equation of motion. The conserved energy is given by

$$E=c(1-\frac{2GM}{c^2r})\frac{dt}{d\lambda}. \tag{2.82}$$

We also consider the situation in which the particle was initially at rest at $r=r_i$, viz

$$\frac{dr}{d\lambda}|_{r=r_i}=0. \tag{2.83}$$

This means in particular that

$$E^2-c^2=-\frac{2GM}{r_i}. \tag{2.84}$$

The equation of motion becomes

$$(\frac{dr}{d\lambda})^2=\frac{2GM}{r}-\frac{2GM}{r_i}. \tag{2.85}$$

We can identify the affine parameter λ with the proper time for a massive particle. The proper time required to reach the point $r=r_f$ is

$$\tau=\int_0^\tau d\lambda=-(2GM)^{-\frac{1}{2}}\int_{r_i}^{r_f}dr\sqrt{\frac{rr_i}{r_i-r}}. \tag{2.86}$$

The minus sign is due to the fact that in a free fall $dr/d\lambda<0$. By performing the change of variables $r=r_i(1+\cos\alpha)/2$ we find the closed result

$$\tau=\sqrt{\frac{r_i^3}{8GM}}(\alpha_f+\sin\alpha_f). \tag{2.87}$$

This is finite when $r\longrightarrow 2GM/c^2$. Thus a freely falling object will cross the Schwarzschild radius in a finite proper time.

We consider now a distant stationary observer hovering at a fixed radial distance r_∞. His proper time is

$$\tau_\infty = \sqrt{1-\frac{2GM}{c^2 r_\infty^2}}t. \tag{2.88}$$

By using equations (2.81) and (2.82) we can find dr/dt. We get

$$\begin{aligned}\frac{dr}{dt} &= -E^{\frac{1}{2}}\frac{d\lambda}{dt}(E-c\frac{d\lambda}{dt})^{\frac{1}{2}}\\ &= -\frac{c}{E}(1-\frac{2GM}{c^2 r})\bigg(E^2-c^2(1-\frac{2GM}{c^2 r})\bigg)^{\frac{1}{2}}.\end{aligned} \tag{2.89}$$

Near $r=2GM/c^2$ we have

$$\frac{dr}{dt} = -\frac{c^3}{2GM}(r-\frac{2GM}{c^2}). \tag{2.90}$$

The solution is

$$r-\frac{2GM}{c^2} = \exp(-\frac{c^3 t}{2GM}). \tag{2.91}$$

Thus when $r\longrightarrow 2GM/c^2$ we have $t\longrightarrow\infty$.

We see that with respect to a stationary distant observer at a fixed radial distance r_∞ the elapsed time τ_∞ goes to infinity as $r\longrightarrow 2GM/c^2$. The correct interpretation of this result is to say that the stationary distant observer can never see the particle actually crossing the Schwarzschild radius $r_s=2GM/c^2$ although the particle does cross the Schwarzschild radius in a finite proper time as seen by an observer falling with the particle.

2.2 Schwarzschild Black Hole

We go back to the Schwarzschild metric (2.17), viz (we use units in which $c=1$)

$$ds^2 = -(1-\frac{2GM}{r})dt^2+(1-\frac{2GM}{r})^{-1}dr^2+r^2 d\Omega^2. \tag{2.92}$$

For a radial null curve, which corresponds to a photon moving radially in Schwarzschild spacetime, the angles θ and ϕ are constants and $ds^2=0$ and thus

$$0 = -(1-\frac{2GM}{r})dt^2+(1-\frac{2GM}{r})^{-1}dr^2. \tag{2.93}$$

In other words

$$\frac{dt}{dr} = \pm\frac{1}{1-\frac{2GM}{r}}. \tag{2.94}$$

This represents the slope of the light cone at a radial distance r on a spacetime diagram of the $t-r$ plane. In the limit $r \longrightarrow \infty$ we get ± 1 which is the flat Minkowski spacetime result whereas as r decreases the slope increases until we get $\pm\infty$ as $r \longrightarrow 2GM$. The light cones close up at $r = 2GM$ (the Schwarzschild radius). See figure GR2.

Thus we reach the conclusion that an infalling observer, as seen by us, never crosses the event horizon $r_s = 2Gm$ in the sense that any fixed interval $\Delta\tau_1$ of its proper time will correspond to a longer and longer interval of our time. In other words the infalling observer will seem us to move slower and slower as it approaches $r_s = 2GM$ but it will never be seen to actually cross the event horizon. This does not mean that the trajectory of the infalling observer will never reach $r_s = 2GM$ because it actually does, however, we need to change the coordinate system to be able to see this.

We integrate the above equation as follows

$$\begin{aligned} t &= \pm \int \frac{dr}{1-\frac{2GM}{r}} \\ &= \pm \int dr \pm 2GM \int \frac{dr}{r-2GM} \\ &= \pm\Big(r + 2GM\log(\frac{r}{2GM}-1)\Big) + \text{constant} \\ &= \pm r_* + \text{constant}. \end{aligned} \tag{2.95}$$

We call r_* the tortoise coordinate which makes sense only for $r > 2GM$. The event horizon $r = 2GM$ corresponds to $r_* \longrightarrow \infty$. We compute $dr_* = rdr/(r-2GM)$ and as a consequence the Schwarzschild metric becomes

$$ds^2 = (1-\frac{2GM}{r})(-dt^2 + dr_*^2) + r^2 d\Omega^2. \tag{2.96}$$

Next we define $v = t + r_*$ and $u = t - r_*$. Then

$$ds^2 = -(1-\frac{2GM}{r})dvdu + r^2 d\Omega^2. \tag{2.97}$$

For infalling radial null geodesics we have $t = -r_*$ or equivalently $v =$ constant whereas for outgoing radial null geodesics we have $t = +r_*$ or equivalently $u =$ constant. We will think of v as our new time coordinate whereas we will change u back to the radial coordinate r via $u = v - 2r_* = v - 2r - 4GM\log(r/(2GM) - 1)$. Thus $du = dv - 2dr/(1-2GM/r)$ and as a consequence

$$ds^2 = -(1-\frac{2GM}{r})dv^2 + 2dvdr + r^2 d\Omega^2. \tag{2.98}$$

These are called the Eddington-Finkelstein coordinates. We remark that the determinant of the metric in this system of coordinates is $g = -r^4\sin^2\theta$ which is regular at $r = 2GM$, i.e. the metric is invertible and the original singularity at $r = 2GM$ is simply a coordinate singularity characterizing the system of coordinates (t, r, θ, ϕ). In the Eddington-Finkelstein coordinates the radial null curves are given by the condition

$$\left[(1 - \frac{2GM}{r})\frac{dv}{dr} - 2\right]\frac{dv}{dr} = 0. \tag{2.99}$$

We have the following solutions:

- $dv/dr = 0$ or equivalently $v =$ constant which corresponds to an infalling observer.

- $dv/dr \neq 0$ or equivalently $dv/dr = 2/(1 - \frac{2GM}{r})$. For $r > 2GM$ we obtain the solution $v = 2r + 4GM \log(r/2GM - 1) +$ constant which corresponds to an outgoing observer since $dv/dr > 0$. This actually corresponds to $u =$ constant.

- $dv/dr \neq 0$ or equivalently $dv/dr = 2/(1 - \frac{2GM}{r})$. For $r < 2GM$ we obtain the solution $v = 2r + 4GM \log(1 - r/2GM) +$ constant which corresponds to an infalling observer since $dv/dr < 0$.

- For $r = 2GM$ the above equation reduces to $dvdr = 0$. This corresponds to the observer trapped at $r = 2GM$.

The above solutions are drawn on figure GR3 in the plane $(v - r) - r$, i.e. the time axis (the perpendicular axis) is $v - r$ and not v. Thus for every point in spacetime we have two solutions:

- The points outside the event horizon such as point 1 on figure GR3: There are two solutions one infalling and one outgoing.

- The points inside the event horizon such as point 3 on figure GR3: There are two solutions both are infalling.

- The points on the event horizon such as point 2 on figure GR3: There are two solutions one infalling and one trapped.

Several other remarks are of order:

- The light cone at each point of spacetime is determined (bounded) by the two solutions at that point. See figure GR3.

- The left side of the light cones is always determined by infalling observers.

- The right side of the light cones for $r > 2GM$ is always determined by outgoing observers.

- The right side of the light cones for $r < 2GM$ is always determined by infalling observers.

- The light cone tilt inward as r decreases. For $r < 2GM$ the light cone is sufficiently tilted that no observer can escape the singularity at $r = 0$.

- The horizon $r = 2GM$ is clearly a null surface which consists of observers who can neither fall into the singularity nor escape to infinity (since it is a solution to a null condition which is trapped at $r = 2GM$).

2.3 The Kruskal-Szekres Diagram: Maximally Extended Schwarzschild Solution

We have shown explicitly that in the (v,r,θ,ϕ) coordinate system we can cross the horizon at $r=2GM$ along future directed paths since from the definition $v=t+r_*$ we see that for a fixed v (infalling null radial geodesics) we must have $t=-r_*+\text{constant}$ and thus as $r\longrightarrow 2GM$ we must have $t\longrightarrow +\infty$. However we have also shown that we can cross the horizon at $r=2GM$ along past directed paths corresponding to $v=2r_*+\text{constant}$ or equivalently $u=\text{constant}$ (outgoing null radial geodesics) and thus as $r\longrightarrow 2GM$ we must have $t\longrightarrow -\infty$. We have also been able to extend the solution to the region $r\leq 2GM$.

In the following we will give a maximal extension of the Schwarzschild solution by constructing a coordinate system valid everywhere in Schwarzschild spacetime. We start by rewriting the Schwarzschild metric in the (u,v,θ,ϕ) coordinate system as

$$ds^2=-(1-\frac{2GM}{r})dvdu+r^2d\Omega^2. \tag{2.100}$$

The radial coordinate r should be given in terms of u and v by solving the equations

$$\frac{1}{2}(v-u)=r+2GM\log(\frac{r}{2GM}-1). \tag{2.101}$$

The event horizon $r=2GM$ is now either at $v=-\infty$ or $u=+\infty$. The coordinates of the event horizon can be pulled to finite values by defining new coordinates $u^{'}$ and $v^{'}$ as

$$\begin{aligned} v^{'} &= \exp(\frac{v}{4GM})\\ &= \sqrt{\frac{r}{2GM}-1}\exp(\frac{r+t}{4GM}). \end{aligned} \tag{2.102}$$

$$\begin{aligned} u^{'} &= -\exp(-\frac{u}{4GM})\\ &= -\sqrt{\frac{r}{2GM}-1}\exp(\frac{r-t}{4GM}). \end{aligned} \tag{2.103}$$

The Schwarzschild metric becomes

$$ds^2=-\frac{32G^3M^3}{r}\exp(-\frac{r}{2GM})dv^{'}du^{'}+r^2d\Omega^2. \tag{2.104}$$

It is clear that the coordinates u and v are null coordinates since the vectors $\partial/\partial u$ and $\partial/\partial v$ are tangent to light cones and hence they are null vectors. As a consequence $u^{'}$ and $v^{'}$ are null coordinates. However, we prefer to work with a single time like coordinate while we prefer the other coordinate to be space like. We introduce therefore new coordinates T and R defined for $r>2GM$ by

$$T=\frac{1}{2}(v^{'}+u^{'})=\sqrt{\frac{r}{2GM}-1}\exp(\frac{r}{4GM})\sinh\frac{t}{4GM}. \tag{2.105}$$

$$R = \frac{1}{2}(v^{'} - u^{'}) = \sqrt{\frac{r}{2GM} - 1}\exp(\frac{r}{4GM})\cosh\frac{t}{4GM}. \tag{2.106}$$

Clearly T is time like while R is space like. This can be confirmed by computing the metric. This is given by

$$ds^2 = \frac{32G^3M^3}{r}\exp(-\frac{r}{2GM})(-dT^2 + dR^2) + r^2 d\Omega^2. \tag{2.107}$$

We see that T is always time like while R is always space like since the sign of the components of the metric never get reversed.

We remark that

$$\begin{aligned} T^2 - R^2 &= v^{'}u^{'} \\ &= -\exp\frac{v-u}{4GM} \\ &= -\exp\frac{r + 2GM\log(\frac{r}{2GM} - 1)}{2GM} \\ &= (1 - \frac{r}{2GM})\exp\frac{r}{2GM}. \end{aligned} \tag{2.108}$$

The radial coordinate r is determined implicitly in terms of T and R from this equation, i.e. equation (2.108). The coordinates (T, R, θ, ϕ) are called Kruskal-Szekres coordinates. Remarks are now in order

- The radial null curves in this system of coordinates are given by

$$T = \pm R + \text{constant}. \tag{2.109}$$

- The horizon defined by $r \longrightarrow 2GM$ is seen to appear at $T^2 - R^2 \longrightarrow 0$, i.e. at (2.109) in the new coordinate system. This shows in an elegant way that the event horizon is a null surface.

- The surfaces of constant r are given from (2.108) by $T^2 - R^2 = \text{constant}$ which are hyperbolae in the $R - T$ plane.

- For $r > 2GM$ the surfaces of constant t are given by $T/R = \tanh t/4GM = \text{constant}$ which are straight lines through the origin. In the limit $t \longrightarrow \pm\infty$ we have $T/R \longrightarrow \pm 1$ which are precisley the horizon $r = 2GM$.

- For $r < 2GM$ we have

$$T = \frac{1}{2}(v^{'} + u^{'}) = \sqrt{1 - \frac{r}{2GM}}\exp(\frac{r}{4GM})\cosh\frac{t}{4GM}. \tag{2.110}$$

$$R = \frac{1}{2}(v^{'} - u^{'}) = \sqrt{1 - \frac{r}{2GM}}\exp(\frac{r}{4GM})\sinh\frac{t}{4GM}. \tag{2.111}$$

  The metric and the condition determining r implicitly in terms of T and R do not change form in the (T, R, θ, ϕ) system of coordinates and thus the radial null curves, the horizon as well as the surfaces of constant r are given by the same equation as before.

- For $r < 2GM$ the surfaces of constant t are given by $T/R = 1/\tanh t/4GM =$ constant which are straight lines through the origin.

- It is clear that the allowed range for R and T is (analytic continuation from the region $T^2 - R^2 < 0$ ($r > 2GM$) to the first singularity which occurs in the region $T^2 - R^2 < 1$ ($r < 2GM$))

$$-\infty \leq R \leq +\infty \ , \ T^2 - R^2 \leq 1. \tag{2.112}$$

A Kruskal-Szekres diagram is shown on figure GR4. Every point in this diagram is actually a 2−dimensional sphere since we are suppressing θ and ϕ and drawing only R and T. The Kruskal-Szekres diagram gives the maximal extension of the Schwarzschild solution. In some sense it represents the entire Schwarzschild spacetime. It can be divided into 4 regions:

- Region 1: Exterior of black hole with $r > 2GM$ ($R > 0$ and $T^2 - R^2 < 0$). Clearly future directed time like (null) worldlines will lead to region 2 whereas past directed time like (null) worldlines can reach it from region 4. Regions 1 and 3 are connected by space like geodesics.

- Region 2: Inside of black hole with $r < 2GM$ ($T > 0$, $0 < T^2 - R^2 < 1$). Any future directed path in this region will hit the singularity. In this region r becomes time like (while t becomes space like) and thus we can not stop moving in the direction of decreasing r in the same way that we can not stop time progression in region 1.

- Region 3: Parallel exterior region with $r > 2GM$ ($R < 0$, $T^2 - R^2 < 0$). This is another asymptotically flat region of spacetime which we can not access along future or past directed paths.

- Region 4: Inside of white hole with $r < 2GM$ ($T < 0$, $0 < T^2 - R^2 < 1$). The white hole is the time reverse of the black hole. This corresponds to a singularity in the past at which the universe originated. This is a part of spacetime from which observers can escape to reach us while we can not go there.

2.4 Various Theorems and Results

The various theorems and results quoted in this section requires a much more careful and detailed analysis much more than what we are able to do at this stage.

- **Birkhoff's Theorem:** The Schwarzschild solution is the only spherically symmetric solution of general relativity in *vacuum.*

  This is to be compared with Coulomb potential which is the only spherically symmetric solution of Maxwell's equations in vacuum.

- **No-Hair Theorem (Example):** General relativity coupled to Maxwell's equations admits a small number of *stationary asymptotically flat* black hole solutions which are *non-singular outside the event horizon* and which are characterized by a limited number of parameters given by *the mass, the charge (electric and magnetic) and the angular momentum*.

  In contrast with the above result there exists in general relativity an infinite number of planet solutions and each solution is generically characterized by an infinite number of parameters.

- **Event Horizon:** Black holes are characterized by their event horizons. A horizon is a boundary line between two regions of spacetime. Region I consists of all points of spacetime which are connected to infinity by time like geodesics whereas region II consists of all spacetime points which are not connected to infinity by time like geodesics, i.e. observers can not reach infinity starting from these points. The boundary between regions I and II, which is the event horizon, is a light like (null) hyper surface.

  The event horizon can be defined as the set of points where the light cones are tilted over (in an appropriate coordinate system). In the Schwarzschild solution the event horizon occurs at $r = 2GM$ which is a null surface although $r =$ constant is time like surface for large r.

  In a general stationary metric we can choose a coordinate system where $\partial_t g_{\mu\nu} = 0$ and on hypersurfaces $t =$ constant the coordinates will resemble spherical polar coordinates (r, θ, ϕ) sufficiently far away. Thus hypersurfaces $r =$ constant are time like with the topology $S^2 \times R$ as $r \longrightarrow \infty$. It is obvious that $\partial_\mu r$ is a normal one-form to these hypersurfaces with norm

$$g^{rr} = g^{\mu\nu}\partial_\mu r \partial_\nu r. \tag{2.113}$$

  If the time like hypersurfaces $r =$ constant become null at some $r = r_H$ then we will get an event horizon at $r = r_H$ since any time like geodesic crossing to the region $r < r_H$ will not be able to escape back to infinity. For $r > r_H$ we have clearly $g^{rr} > 0$ whereas for $r < r_H$ we have $g^{rr} < 0$. The event horizon is defined by the condition

$$g^{rr}(r_H) = 0. \tag{2.114}$$

- **Trapped Surfaces:**In general relativity singularities are generic and they are hidden behind event horizons. As shown by Hawking and Penrose singularities are inevitable if gravitational collapse reach a point of no return, i.e. the appearance of trapped surface.

  Let us consider a 2−sphere in Minkowski spacetime. We consider then null rays emanating from the sphere inward or outward. The rays emanating outward describe growing spheres whereas the rays emanating inward describe shrinking spheres. Consider now a 2−sphere in Schwarzschild spacetime with $r < 2GM$. In this case the rays emanating outward and inward will correspond to shrinking spheres (r is time like). This is called a trapped surface.

A trapped surface is a compact space like $2-$dimensional surface with the property that outward light rays are in fact moving inward.

- **Singularity Theorem (Example):** A trapped surface in a manifold M with a generic metric $g_{\mu\nu}$ (which is a solution of Einstein's equation satisfying the strong energy condition [6]) can only be a closed time like curve or a singularity.

- **Cosmic Censorship Conjecture:** In general relativity singularities are hidden behind event horizons. More precisely, naked singularities can not appear in the gravitational collapse of a non singular state in an asymptotically flat spacetime which fulfills the dominant energy condition [7].

- **Hawking's Area Theorem:** In general relativity black holes can not shrink but they can grow in size. Clearly the size of the black hole is measured by the area of the event horizon.

  Hawking's area theorem can be stated as follows. The area of a future event horizon in an asymptotically flat spacetime is always increasing provided the cosmic censorship conjecture and the weak energy condition hold [8] [9].

- **Stokes's Theorem :** Next we recall stokes's theorem

$$\int_{\Sigma} d\omega = \int_{\partial\Sigma} \omega. \tag{2.115}$$

  Explicitly this reads

$$\int_{\Sigma} d^n x\sqrt{|g|}\ \nabla_{\mu} V^{\mu} = \int_{\partial\Sigma} d^{n-1}y\sqrt{|\gamma|}\ \sigma_{\mu} V^{\mu}. \tag{2.116}$$

  The unit vector σ^{μ} is normal to the boundary $\partial\Sigma$. In the case that Σ is the whole space, the boundary $\partial\Sigma$ is the $2-$sphere at infinity and thus σ^{μ} is given, in an appropriate system of coordinates, by the components $(0,1,0,0)$.

- **Energy in GR:** The concept of conserved total energy in general relativity is not straightforward.

[6]Exercise: The strong energy condition is given by $T_{\mu\nu}t^{\mu}t^{\nu} \geq T^{\lambda}{}_{\lambda}t^{\sigma}t_{\sigma}/2$ for any time like vector t^{μ}. Show that this is equivalent to $\rho + P \geq 0$ and $\rho + 3P \geq 0$.

[7]Exercise: The dominant energy condition is given by $T_{\mu\nu}t^{\mu}t^{\nu} \geq 0$ and $T_{\mu\nu}T^{\nu}{}_{\lambda}t^{\mu}t^{\lambda} \leq 0$ for any time like vector t^{μ}. Show that these are equivalent to $\rho \geq |P|$.

[8]Exercise: The weak energy condition is given by $T_{\mu\nu}t^{\mu}t^{\nu} \geq 0$ for any time like vector t^{μ}. Show that these are equivalent to $\rho \geq 0$ and $\rho + P \geq 0$.

[9]Exercise: Show that for a Schwarzschild black hole this theorem implies that the mass of the black hole can only increase.

For a stationary asymptotically flat spacetime with a time like Killing vector field K^μ we can define a conserved energy-momentum current J_T^μ by [10]

$$J_T^\mu = K_\nu T^{\mu\nu}. \tag{2.117}$$

Let Σ by a space like hypersurface with a unit normal vector n^μ and an induced metric γ_{ij}. By integrating the component of J_T^μ along the normal n^μ over the surface Σ we get an energy, viz

$$E_T = \int_\Sigma d^3x\sqrt{\gamma}\; n_\mu J_T^\mu. \tag{2.118}$$

This definition is however inadequate since it gives zero energy in the case of Schwarzschild spacetime.

Let us consider instead the following current

$$\begin{aligned} J_R^\mu &= K_\nu R^{\mu\nu} \\ &= 8\pi G K_\nu\big(T^{\mu\nu} - \frac{1}{2}g^{\mu\nu}T\big). \end{aligned} \tag{2.119}$$

We compute now

$$\nabla_\mu J_R^\mu \;=\; K_\nu \nabla_\mu R^{\mu\nu}. \tag{2.120}$$

By using now the contracted Bianchi identity $\nabla_\mu G^{\mu\nu} = \nabla_\mu(R^{\mu\nu} - g^{\mu\nu}R/2) = 0$ or equivalently $\nabla_\mu R^{\mu\nu} = \nabla^\nu R/2$ we get

$$\nabla_\mu J_R^\mu \;=\; \frac{1}{2}K_\nu \nabla^\nu R. \tag{2.121}$$

The derivative of the scalar curvature along a Killing vector must vanish [11] and as a consequence J_R^μ is conserved. The corresponding energy is defined by

$$E_R = \frac{1}{4\pi G}\int_\Sigma d^3x\sqrt{\gamma}\; n_\mu J_R^\mu. \tag{2.122}$$

The normalization is chosen for later convenience. The Killing vector K^μ satisfies among other things $\nabla_\nu\nabla^\mu K^\nu = R^{\mu\nu}K_\nu$ [12] and hence the vector J_R^μ is actually a total derivative, viz

$$J_R^\mu \;=\; \nabla_\nu\nabla^\mu K^\nu. \tag{2.123}$$

[10] Exercise: Verify that J_T^μ is conserved by using the fact that the energy-momentum tensor is conserved ($\nabla_\mu T^{\mu\nu} = 0$) and the fact that K^μ is a Killing vector ($\nabla_\mu K_\nu + \nabla_\nu K_\mu = 0$).

[11] Exercise: Show this explicitly.

[12] Exercise: Show this explicitly. This is one of the formula which might be used in the previous exercise.

The energy E_R becomes

$$\begin{aligned} E_R &= \frac{1}{4\pi G}\int_\Sigma d^3x\sqrt{\gamma}\; n_\mu\nabla_\nu\nabla^\mu K^\nu \\ &= \frac{1}{4\pi G}\int_\Sigma d^3x\sqrt{\gamma}\; \nabla_\nu\big(n_\mu\nabla^\mu K^\nu\big) - \frac{1}{4\pi G}\int_\Sigma d^3x\sqrt{\gamma}\; \nabla_\nu n_\mu.\nabla^\mu K^\nu. \end{aligned} \tag{2.124}$$

In the second term we can clearly replace $\nabla_\nu n_\mu$ with $(\nabla_\nu n_\mu - \nabla_\mu n_\nu)/2 = (\partial_\nu n_\mu - \partial_\mu n_\nu)/2$. The surface Σ is space like and thus the unit vector n^μ is time like. For example Σ can be the whole of space and thus n^μ must be given, in an appropriate system of coordinates, by the components $(1,0,0,0)$. In this system of coordinates the second term vanishes. The above equation reduces to

$$E_R = \frac{1}{4\pi G}\int_\Sigma d^3x\sqrt{\gamma}\; \nabla_\nu\big(n_\mu\nabla^\mu K^\nu\big). \tag{2.125}$$

By using stokes's theorem we get the result

$$E_R = \frac{1}{4\pi G}\int_{\partial\Sigma} d^2x\sqrt{\gamma^{(2)}}\; \sigma_\nu\big(n_\mu\nabla^\mu K^\nu\big). \tag{2.126}$$

This is Komar integral which defines the total energy of the stationary spacetime. For Schwarzschild spacetime we can check that $E_R = M$ [13]. The Komar energy agrees with the ADM (Arnowitt, Deser, Misner) energy which is obtained from a Hamiltonian formulation of general relativity and which is associated with invariance under time translations.

2.5 Reissner-Nordström (Charged) Black Hole

2.5.1 Maxwell's Equations and Charges in GR

Maxwell's equations in flat spacetime are given by

$$\partial_\mu F^{\mu\nu} = -J^\nu. \tag{2.127}$$

$$\partial_\mu F_{\nu\lambda} + \partial_\lambda F_{\mu\nu} + \partial_\nu F_{\lambda\mu} = 0. \tag{2.128}$$

Maxwell's equations in curved spacetime can be obtained from the above equations using the principle of minimal coupling which consists in making the replacements $\eta_{\mu\nu} \longrightarrow g_{\mu\nu}$ and $\partial_\mu \longrightarrow D_\mu$ where D_ν is the covariant derivative associated with the metric $g_{\mu\nu}$. The homogeneous equation does not change under these substitutions since the extra corrections coming from the Christoffel symbols cancel by virtue of the antisymmetry under permutations of μ, ν and λ [14].

[13] Exercise: Show this explicitly.
[14] Exercise: Show this explicitly.

This also means that the field strength tensor $F_{\mu\nu}$ in curved spacetime is still given by the same formula as in the flat case, viz

$$F_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu. \tag{2.129}$$

The inhomogeneous Maxwell's equation in curved spacetime is given by

$$D_\mu F^{\mu\nu} = -J^\nu. \tag{2.130}$$

We compute

$$\begin{aligned} D_\mu F^{\mu\nu} &= \partial_\mu F^{\mu\nu} + \Gamma^\mu{}_{\mu\alpha} F^{\alpha\nu} \\ &= \partial_\mu F^{\mu\nu} + \frac{1}{2} g^{\mu\rho}\partial_\alpha g_{\mu\rho} F^{\alpha\nu}. \end{aligned} \tag{2.131}$$

Let $g = \det g_{\mu\nu}$ and let e_i be the eigenvalues of the matrix $g_{\mu\nu}$. We have the result

$$\frac{\partial\sqrt{-g}}{\sqrt{-g}} = \frac{1}{2}\frac{\partial g}{g} = \frac{1}{2}\sum_i \frac{\partial e_i}{e_i} = \frac{1}{2} g^{\mu\rho}\partial g_{\mu\rho}. \tag{2.132}$$

Thus

$$D_\mu F^{\mu\nu} = \partial_\mu F^{\mu\nu} + \frac{\partial_\alpha\sqrt{-g}}{\sqrt{-g}} F^{\alpha\nu}. \tag{2.133}$$

Using this result we can put the inhomogeneous Maxwell's equation in the equivalent form

$$\partial_\mu(\sqrt{-g} F^{\mu\nu}) = -\sqrt{-g} J^\nu. \tag{2.134}$$

The law of conservation of charge in curved spacetime is now obvious given by

$$\partial_\mu(\sqrt{-g} J^\mu) = 0. \tag{2.135}$$

This is equivalent to the form

$$D_\mu J^\mu = 0. \tag{2.136}$$

The energy-momentum tensor of electromagnetism is given by [15]

$$T_{\mu\nu} = F_{\mu\alpha} F_\nu{}^\alpha - \frac{1}{4} g_{\mu\nu} F_{\alpha\beta} F^{\alpha\beta} + g_{\mu\nu} J_\alpha A^\alpha. \tag{2.137}$$

We define the electric and magnetic fields by $F_{0i} = E_i$ and $F_{ij} = \epsilon_{ijk} B_k$ with $\epsilon_{123} = -1$.

The amount of electric charge passing through a space like hypersurface Σ with unit normal vector n^μ is given by the integral

$$\begin{aligned} Q &= -\int_\Sigma d^3x \sqrt{\gamma} n_\mu J^\mu \\ &= -\int_\Sigma d^3x \sqrt{\gamma} n_\mu D_\nu F^{\mu\nu}. \end{aligned} \tag{2.138}$$

[15] Exercise: Construct a derivation of this result.

The metric γ_{ij} is the induced metric on the surface Σ. By using Stokes's theorem we obtain

$$Q = -\int_{\partial\Sigma} d^2x\sqrt{\gamma^{(2)}}n_\mu\sigma_\nu F^{\mu\nu}. \tag{2.139}$$

The unit vector σ^μ is normal to the boundary $\partial\Sigma$.

The magnetic charge P can be defined similarly by considering instead the dual field strength tensor $*F^{\mu\nu} = \epsilon^{\mu\nu\alpha\beta}F_{\alpha\beta}/2$.

2.5.2 Reissner-Nordström Solution

We are interested in finding a spherically symmetric solution of Einsetin-Maxwell equations with some mass M, some electric charge Q and some magnetic charge P, i.e. we want to find the gravitational field around a star of mass M, electric charge Q and magnetic charge P.

We start from the metric

$$ds^2 = -A(r)dt^2 + B(r)dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{2.140}$$

We compute immediately $\sqrt{-g} = \sqrt{AB}r^2\sin^2\theta$. The components of the Ricci tensor in this metric are given by (with $A = e^{2\alpha}$, $B = e^{2\beta}$)

$$\begin{aligned}
R_{00} &= \big(\partial_r^2\alpha + (\partial_r\alpha)^2 - \partial_r\beta\partial_r\alpha + \frac{2}{r}\partial_r\alpha\big)e^{2(\alpha-\beta)}\\
R_{rr} &= -\partial_r^2\alpha - (\partial_r\alpha)^2 + \partial_r\beta\partial_r\alpha + \frac{2}{r}\partial_r\beta\\
R_{\theta\theta} &= e^{-2\beta}\big(r\partial_r\beta - r\partial_r\alpha - 1\big) + 1\\
R_{\phi\phi} &= \sin^2\theta\big[e^{-2\beta}\big(r\partial_r\beta - r\partial_r\alpha - 1\big) + 1\big].
\end{aligned} \tag{2.141}$$

We also need to provide an ansatz for the electromagnetic field. By spherical symmetry the most general electromagnetic field configuration corresponds to a radial electric field and a radial magnetic field. For simplicity we will only consider a radial electric field which is also static, viz

$$E_r = f(r)\ ,\ E_\theta = E_\phi = 0\ ,\ B_r = B_\theta = B_\phi = 0. \tag{2.142}$$

We will also choose the current J^μ to be zero outside the star where we are interested in finding a solution. We compute $F^{0r} = -f(r)/AB$ while all other components are 0. The only non-trivial component of the inhomogeneous Maxwell's equation is $\partial_r(\sqrt{-g}F^{r0}) = 0$ and hence

$$\partial_r\big(\frac{r^2 f(r)}{\sqrt{AB}}\big) = 0 \Leftrightarrow f(r) = \frac{Q\sqrt{AB}}{4\pi r^2}. \tag{2.143}$$

The constant of integration Q will play the role of the electric charge since it is expected that A and B approach 1 when $r \longrightarrow \infty$. The homogeneous Maxwell's equation is satisfied since the only non-zero component of $F^{\mu\nu}$, i.e. F^{0r} is clearly of the form $-\partial^r A^0$ for some potential A^0 while the other components of the vector potential (A^r, A^θ and A^ϕ) are 0.

We have therefore shown that the above electrostatic ansatz solves Maxwell's equations. We are now ready to compute the energy-momentum tensor in this configuration. We compute

$$\begin{aligned}T_{\mu\nu} &= \frac{f^2(r)}{AB}\big(\frac{1}{A}g_{\mu 0}g_{\nu 0}-\frac{1}{B}g_{\mu r}g_{\nu r}+\frac{1}{2}g_{\mu\nu}\big)\\ &= \frac{f^2(r)}{2AB}\mathrm{diag}(A,-B,r^2,r^2\sin^2\theta).\end{aligned}\tag{2.144}$$

Also

$$\begin{aligned}T_{\mu}{}^{\nu} &= g^{\nu\lambda}T_{\mu\lambda}\\ &= \frac{f^2(r)}{2AB}\mathrm{diag}(-1,-1,+1,+1).\end{aligned}\tag{2.145}$$

The trace of the energy-momentum is therefore traceless as it should be for the electromagnetic field. Thus Einstein's equation takes the form

$$R_{\mu\nu}=8\pi GT_{\mu\nu}.\tag{2.146}$$

We find three independent equations given by

$$\big(\partial_r^2\alpha+(\partial_r\alpha)^2-\partial_r\beta\partial_r\alpha+\frac{2}{r}\partial_r\alpha\big)A=4\pi Gf^2.\tag{2.147}$$

$$\big(-\partial_r^2\alpha-(\partial_r\alpha)^2+\partial_r\beta\partial_r\alpha+\frac{2}{r}\partial_r\beta\big)A=-4\pi Gf^2.\tag{2.148}$$

$$e^{-2\beta}\big(r\partial_r\beta-r\partial_r\alpha-1\big)+1=4\pi Gf^2\frac{r^2}{AB}.\tag{2.149}$$

From the first two equations (2.147) and (2.148) we deduce

$$\partial_r(\alpha+\beta)=0.\tag{2.150}$$

In other words

$$\alpha=-\beta+c\Leftrightarrow B=\frac{c^{'}}{A}.\tag{2.151}$$

c and $c^{'}$ are constants of integration. By substituting this solution in the third equation (2.149) we obtain

$$\partial_r(\frac{r}{B})=1-GQ^2\frac{1}{4\pi r^2}\Leftrightarrow\frac{1}{B}=1+\frac{GQ^2}{4\pi r^2}+\frac{b}{r}.\tag{2.152}$$

In other words

$$A=c^{'}+\frac{GQ^2c^{'}}{4\pi r^2}+\frac{bc^{'}}{r}.\tag{2.153}$$

The first equation (2.147) is equivalent to

$$\partial_r^2 A + \frac{2}{r}\partial_r A = 8\pi G f^2. \tag{2.154}$$

By substituting the solution (2.153) back in (2.154) we get $c^{'} = 1$. In other words we must have

$$B = \frac{1}{A}\ ,\ A = 1 + \frac{GQ^2}{4\pi r^2} + \frac{bc^{'}}{r}. \tag{2.155}$$

Similarly to the Schwarzschild solution we can now invoke the Newtonian limit to set $bc^{'} = -2GM$. We get then the solution

$$A = 1 - \frac{2GM}{r} + \frac{GQ^2}{4\pi r^2}. \tag{2.156}$$

If we also assume a radial magnetic field generated by a magnetic charge P inside the star we obtain the more general metric [16]

$$ds^2 = -\Delta(r)dt^2 + \Delta^{-1}(r)dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{2.157}$$

$$\Delta = 1 - \frac{2GM}{r} + \frac{G(Q^2+P^2)}{4\pi r^2}. \tag{2.158}$$

This is the Reissner-Nordström solution. The event horizon is located at $r = r_H$ where

$$\Delta(r_H) = 0 \Leftrightarrow r^2 - 2GMr + \frac{G(Q^2+P^2)}{4\pi} = 0. \tag{2.159}$$

We should then consider the discriminant

$$\delta = 4G^2M^2 - \frac{G(Q^2+P^2)}{\pi}. \tag{2.160}$$

There are three possible cases:

- The case $GM^2 < (Q^2+P^2)/4\pi$. There is a naked singularity at $r = 0$. The coordinate r is always space like while the coordinate t is always time like. There is no event horizon. An observer can therefore travel to the singularity and return back. However the singularity is repulsive. More precisely a time like geodesic does not intersect the singularity. Instead it approaches $r = 0$ then it reverses its motion and drives away.

  This solution is in fact unphysical since the condition $GM^2 < (Q^2+P^2)/4\pi$ means that the total energy is less than the sum of two of its components which is impossible.

[16] Exercise: Verify this explicitly.

- The case $GM^2 > (Q^2+P^2)/4\pi$. There are two horizons at

$$r_{\pm} = GM \pm \sqrt{G^2M^2 - \frac{G(Q^2+P^2)}{4\pi}}. \tag{2.161}$$

  These are of course null surfaces. The horizon at $r = r_+$ is similar to the horizon of the Schwarzschild solution. At this point the coordinate r becomes time like ($\Delta < 0$) and a falling observer will keep going in the direction of decreasing r. At $r = r_-$ the coordinate r becomes space like again ($\Delta > 0$). Thus the motion in the direction of decreasing r can be reverses, i.e. the singularity at $r = 0$ can be avoided.

  The fact that the singularity can be avoided is consistent with the fact that $r = 0$ is a time like line in the Reissner-Nordström solution as opposed to the singularity $r = 0$ in the Schwarzschild solution which is a space like surface.

  The observer in the region $r < r_-$ can therefore move either towards the singularity at $r = 0$ or towards the null surface $r = r_-$. After passing $r = r_-$ the coordinate r becomes time like once more and the observer in this case can only move in the direction of increasing r until it emerges from the black hole at $r = r_+$.

- The case $GM^2 = (Q^2+P^2)/4\pi$ (Extremal RN Black Holes). There is a single horizon at $r = GM$. In this case the coordinate r is always space like except at $r = GM$ where it is null. Thus the singularity can also be avoided in this case.

2.5.3 Extremal Reissner-Nordström Black Hole

The metric at $GM^2 = (Q^2+P^2)/4\pi$ takes the form

$$ds^2 = -(1-\frac{GM}{r})^2dt^2 + (1-\frac{GM}{r})^{-2}dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{2.162}$$

We define the new coordinate $\rho = r - GM$ and the function $H(\rho) = 1 + GM/\rho$. The metric becomes

$$ds^2 = -H^{-2}(\rho)dt^2 + H^2(\rho)\big(d\rho^2 + \rho^2(d\theta^2 + \sin^2\theta d\phi^2)\big). \tag{2.163}$$

Equivalently

$$ds^2 = -H^{-2}(\vec{x})dt^2 + H^2(\vec{x})d\vec{x}^2\ ,\ H(\vec{x}) = 1 + \frac{GM}{|\vec{x}|}. \tag{2.164}$$

For simplicity let us consider only a static electric field which is given by $E_r = F_{0r} = Q/4\pi r^2$. From the extremal condition we have $Q^2 = 4\pi GM^2$. For electrostatic fields we have $F_{0r} = -\partial_r A_0$ and the rest are zero. Then it is not difficult to show that

$$A_0 = \frac{Q}{4\pi r} = \frac{1}{\sqrt{4\pi G}}\frac{GM}{\rho + GM}\ ,\ A_i = 0. \tag{2.165}$$

Equivalently

$$\sqrt{4\pi G}A_0 = 1 - \frac{1}{H(\rho)} \ , \ A_i = 0. \tag{2.166}$$

The metric (2.164) together with the gauge field configuration (2.166) with an arbitrary function $H(\vec{x})$ still solves the Einstein-Maxwell's equations provided $H(\vec{x})$ satisfies the Laplace equation [17]

$$\vec{\nabla}^2 H = 0. \tag{2.167}$$

The general solution is given by

$$H(\vec{x}) = 1 + \sum_{i=1}^{N} \frac{GM_i}{|\vec{x} - \vec{x}_i|}. \tag{2.168}$$

This describes a system of N extremal RN black holes located at $\vec{x}_i$ with masses M_i and charges $Q_i^2 = 4\pi G M_i^2$.

2.6 Kerr Spacetime

2.6.1 Kerr (Rotating) and Kerr-Newman (Rotating and Charged) Black Holes

- The Schwarzschild black hols and the Reissner-Nordström black holes are spherically symmetric. Any spherically symmetric vacuum solution of Einstein's equations possess a time like Killing vector and thus is stationary.

  In a stationary metric we can choose coordinates (t, x^1, x^2, x^3) where the killing vector is ∂_t, the metric components are all independent of the time coordinate t and the metric is of the form

  $$ds^2 = g_{00}(x)dt^2 + 2g_{0i}(x)dtdx^i + g_{ij}(x)dx^i dx^j. \tag{2.169}$$

  This stationary metric becomes static if the time like Killing vector ∂_t is also orthogonal to a family of hypersurfaces. In the coordinates (t, x^1, x^2, x^3) the Killing vector ∂_t is orthogonal to the hypersurfaces $t =$ constant and equivalently a stationary metric becomes static if $g_{0i} = 0$.

- In contrast the Kerr and the Kerr-Newman black holes are not spherically symmetric and are not static but they are stationary. A Kerr black hole is a vacuum solution of Einstein's equations which describes a rotating black hole and thus is characterized by mass and angular momentum whereas the Kerr-Newman black hole is a charged Kerr black hole

[17] Exercise: Derive explicitly this result.

and thus is characterized by mass, angular momentum and electric and magnetic charges. The rotation clearly breaks spherical symmetry and makes the black holes not static. However since the black hole rotates in the same way at all times it is still stationary. The Kerr and Kerr-Newman metrics must therefore be of the form

$$ds^2 = g_{00}(x)dt^2 + 2g_{0i}(x)dtdx^i + g_{ij}(x)dx^i dx^j. \tag{2.170}$$

- The Kerr metric must be clearly axial symmetric around the axis fixed by the angular momentum. This will correspond to a second Killing vector ∂_ϕ.

- In summary the metric components, in a properly adapted system of coordinates, will not depend on the time coordinate t (stationary solution) but also it will not depend on the angle ϕ (axial symmetry). Furthermore if we denote the two coordinates t and ϕ by x^a and the other two coordinates by y^i the metric takes then the form

$$ds^2 = g_{ab}(y)dx^a dx^b + g_{ij}(y)dx^i dx^j. \tag{2.171}$$

- In the so-called Boyer-Lindquist coordinates (t, r, θ, ϕ) the components of the Kerr metric are found (Kerr (1963)) to be given by

$$g_{tt} = -(1 - \frac{2GMr}{\rho^2}) \ , \ \rho^2 = r^2 + a^2 \cos^2\theta. \tag{2.172}$$

$$g_{t\phi} = -\frac{2GMar\sin^2\theta}{\rho^2}. \tag{2.173}$$

$$g_{rr} = \frac{\rho^2}{\Delta} \ , \ \Delta = r^2 - 2GMr + a^2. \tag{2.174}$$

$$g_{\theta\theta} = \rho^2 \ , \ g_{\phi\phi} = \frac{\sin^2\theta}{\rho^2}\big[(r^2 + a^2)^2 - a^2\Delta\sin^2\theta\big]. \tag{2.175}$$

This solution is characterized by the two numbers M and a. The mass of the Kerr black hole is precisely M whereas the angular momentum of the black hole is $J = aM$.

- In the limit $a \longrightarrow 0$ (no rotation) we obtain the Schwarzschild solution

$$g_{tt} = -(1 - \frac{2GM}{r}) \ , \ g_{rr} = (1 - \frac{2GM}{r})^{-1} \ , \ g_{\theta\theta} = r^2 \ , \ g_{\phi\phi} = r^2\sin^2\theta. \tag{2.176}$$

- In the limit $M \longrightarrow 0$ we obtain the solution

$$g_{tt} = -1 \ , \ g_{rr} = \frac{r^2 + a^2\cos^2\theta}{r^2 + a^2} \ , \ g_{\theta\theta} = r^2 + a^2\cos^2\theta \ , \ g_{\phi\phi} = (r^2 + a^2)\sin^2\theta. \tag{2.177}$$

A solution with no mass and no rotation must correspond to flat Minkowski spacetime. Indeed the coordinates r, θ and ϕ are nothing but ellipsoidal coordinates in flat space. The corresponding Cartesian coordinates are [18]

$$x = \sqrt{r^2 + a^2} \sin\theta \cos\phi \ , \ y = \sqrt{r^2 + a^2} \sin\theta \sin\phi \ , \ z = r \cos\theta. \tag{2.178}$$

- The Kerr-Newman black hole is a generalization of the Kerr black hole which includes also electric and magnetic charges and an electromagnetic field. The electric and magnetic charges can be included via the replacement

$$2GMr \longrightarrow 2GMr - G(Q^2 + P^2). \tag{2.179}$$

The electromagnetic field is given by

$$A_t = \frac{Qr - Pa\cos\theta}{\rho^2} \ , \ A_\phi = \frac{-Qar\sin^2\theta + P(r^2 + a^2)\cos\theta}{\rho^2}. \tag{2.180}$$

2.6.2 Killing Horizons

In Schwarzschild spacetime the Killing vector $K = \partial_t$ becomes null at the event horizon. We say that the event horizon (which is a null surface) is the Killing horizon of the Killing vector $K = \partial_t$. In general the Killing horizon of a Killing vector χ^μ is a null hypersurface Σ along which the Killing vector χ^μ becomes null. Some important results concerning Killing horizons are as follows:

- Every event horizon in a stationary, asymptotically flat spacetime is a Killing horizon for some Killing vector χ^μ.

In the case that the spacetime is stationary and static the Killing vector is precisely $K = \partial_\mu$. In the case that the spacetime is stationary and axial symmetric then the event horizon is a Killing horizon where the Killing vector is a combination of the Killing vector $R = \partial_t$ and the Killing vector $R = \partial_\phi$ associated with axial symmetry. These results are purely geometrical. In the general case of a stationary spacetime then Einstein's equations together with appropriate assumptions on the matter content will also yield the result that every event horizon is a Killing horizon for some Killing vector which is either stationary or axial symmetric.

2.6.3 Surface Gravity

Every Killing horizon is associated with an acceleration called the surface gravity. Let Σ be a killing horizon for the Killing vector χ^μ. We know that $\chi^\mu \chi_\mu$ is zero on the Killing horizon and thus $\nabla_\nu(\chi^\mu \chi_\mu) = 2\chi_\mu \nabla_\nu \chi^\mu$ must be normal to the Killing horizon in the sense that it is

[18] Exercise: Show this explicitly.

orthogonal to any vector tangent to the horizon. The normal to the Killing horizon is however unique given by χ^μ and as a consequence we must have

$$\chi_\mu \nabla_\nu \chi^\mu = -\kappa \chi_\nu. \tag{2.181}$$

This means in particular that the Killing vector χ^μ is a non-affinely parametrized geodesic on the Killing horizon. The coefficient κ is precisely the surface gravity. Since the Killing vector ξ^μ is hypersurface orthogonal we have by the Frobenius's theorem the result [19]

$$\chi_{[\mu} \nabla_\nu \chi_{\sigma]} = -\kappa \chi_\nu. \tag{2.182}$$

We compute

$$\begin{aligned}
\nabla^\mu \chi^\nu \chi_{[\mu} \nabla_\nu \chi_{\sigma]} &= 2\kappa^2 \chi_\sigma + 2\chi_\sigma \nabla^\mu \chi^\nu \nabla_\mu \chi_\nu + 2\nabla^\mu \chi^\nu \big(\nabla_\sigma(\chi_\mu \chi_\nu) - \chi_\mu \nabla_\sigma \chi_\nu\big) \\
&= 4\kappa^2 \chi_\sigma + 2\chi_\sigma \nabla^\mu \chi^\nu \nabla_\mu \chi_\nu.
\end{aligned} \tag{2.183}$$

We get immediately the surface gravity

$$\kappa^2 = -\frac{1}{2} \nabla^\mu \chi^\nu \nabla_\mu \chi_\nu. \tag{2.184}$$

In a static and asymptotically flat spacetime we have $\chi = K$ where $K = \partial_t$ whereas in a stationary and asymptotically flat spacetime we have $\chi = K + \Omega_H R$ where $R = \partial_\phi$. In both cases fixing the normalization of K as $K^\mu K_\mu = -1$ at infinity will fix the normalization of χ and as a consequence fixes the surface gravity of any Killing horizon uniquely.

In a static and asymptotically flat spacetime a more physical definition of surface gravity can be given. The surface gravity is the acceleration of a static observer on the horizon as seen by a static observer at infinity. A static observer is an observer whose $4-$vector velocity U^μ is proportional to the Killing vector K^μ. By normalizing U^μ as $U^\mu U_\mu = -1$ we have

$$U^\mu = \frac{K^\mu}{\sqrt{-K^\mu K_\mu}}. \tag{2.185}$$

A static observer does not necessarily follow a geodesic. Its acceleration is defined by

$$A^\mu = U^\nu \nabla_\nu U^\mu. \tag{2.186}$$

We define the redshift factor V by

$$V = \sqrt{-K^\mu K_\mu}. \tag{2.187}$$

[19]Exercise: Show this result explicitly.

We compute

$$\begin{aligned} A^\mu &= -\frac{1}{V^3}K^\sigma K^\mu \nabla_\sigma V + \frac{U_\sigma}{V}\nabla^\sigma K^\mu \\ &= \frac{1}{V^4}K^\sigma K^\mu K^\alpha \nabla_\sigma K_\alpha - \frac{U_\sigma}{V}\nabla^\mu K^\sigma \\ &= -\frac{U_\sigma}{V}\nabla^\mu K^\sigma \\ &= -\nabla^\mu\big(\frac{U_\sigma}{V}K^\sigma\big) + \nabla^\mu\big(\frac{U_\sigma}{V}\big)K^\sigma \\ &= \frac{1}{V}\nabla^\mu U_\sigma K^\sigma + \nabla^\mu(\frac{1}{V})U_\sigma K^\sigma \\ &= \nabla^\mu \ln V. \end{aligned} \tag{2.188}$$

The magnitude of the acceleration is

$$A = \frac{\sqrt{\nabla_\mu V \nabla^\mu V}}{V}. \tag{2.189}$$

The redshift factor V goes obviously to 0 at the Killing Horizon and hence A goes to infinity. The surface gravity is given precisely by the product VA, viz

$$\kappa = VA = \sqrt{\nabla_\mu V \nabla^\mu V}. \tag{2.190}$$

This agrees with the original definition (2.184) as one can explicitly check [20]. For a Schwarzschild black hole we compute [21]

$$\kappa = \frac{1}{4GM}. \tag{2.191}$$

2.6.4 Event Horizons, Ergosphere and Singularity

- The event horizons occur at $r = r_H$ where $g^{rr}(r_H) = 0$. Since $g^{rr} = \Delta/\rho^2$ we obtain the equation

  $$r^2 - 2GMr + a^2 = 0. \tag{2.192}$$

  The discriminant is $\delta = 4(G^2M^2 - a^2)$. As in the case of Reissner-Nordström solution there are three possibilities. We focus only on the more physically interesting case of $G^2M^2 > a^2$. In this case there are two solutions

  $$r_\pm = GM \pm \sqrt{G^2M^2 - a^2}. \tag{2.193}$$

  These two solutions correspond to two event horizons which are both null surfaces. Since the Kerr solution is stationary and not static the event horizons are not Killing horizons

[20] Exercise: Verify this statement.
[21] Exercise: Derive this result.

for the Killing vector $K = \partial_t$. In fact the event horizons for the Kerr solutions are Killing horizons for the linear combination of the time translation Killing vector $K = \partial_t$ and the rotational Killing vector $R = \partial_\phi$ which is given by

$$\chi^\mu = K^\mu + \Omega_H R^\mu. \tag{2.194}$$

We can check that this vector becomes null at the outer event horizon r_+. We check this explicitly as follows. First we compute

$$K^\mu = \partial_t^\mu = \delta_t^\mu = (1,0,0,0) \Leftrightarrow K_\mu = g_{\mu t} = (-(1-\frac{2GMr}{\rho^2}),0,0,0). \tag{2.195}$$

$$R^\mu = \partial_\phi^\mu = \delta_\phi^\mu = (0,0,0,1) \Leftrightarrow R_\mu = g_{\mu\phi} = (0,0,0,\frac{\sin^2\theta}{\rho^2}\big[(r^2+a^2)^2 - a^2\Delta\sin^2\theta\big]). \tag{2.196}$$

Then

$$K^\mu K_\mu = -\frac{1}{\rho^2}(\Delta - a^2\sin^2\theta). \tag{2.197}$$

$$R^\mu R_\mu = \frac{\sin^2\theta}{\rho^2}\big[(r^2+a^2)^2 - a^2\Delta\sin^2\theta\big]. \tag{2.198}$$

$$R^\mu K_\mu = g_{\phi t} = -\frac{2GMar\sin^2\theta}{\rho^2}. \tag{2.199}$$

Thus

$$\chi^\mu\chi_\mu = -\frac{1}{\rho^2}(\Delta - a^2\sin^2\theta) + \Omega_H^2\frac{\sin^2\theta}{\rho^2}\big[(r^2+a^2)^2 - a^2\Delta\sin^2\theta\big] - \Omega_H\frac{4GMar\sin^2\theta}{\rho^2}. \tag{2.200}$$

At the outer event horizon $r = r_+$ we have $\Delta = 0$ and thus

$$\chi^\mu\chi_\mu = \frac{\sin^2\theta}{\rho^2}\big[(r_+^2+a^2)\Omega_H - a\big]^2. \tag{2.201}$$

This is zero for

$$\Omega_H = \frac{a}{r_+^2+a^2}. \tag{2.202}$$

As it turns out Ω_H is the angular velocity of the event horizon $r = r_+$ which is defined as the angular velocity of a particle at the event horizon $r = r_+$ [22].

[22]Exercise: Compute this velocity directly by computing the angular velocity of a photon emitted in the ϕ direction at some r in the equatorial plane $\theta = \pi/2$ in a Kerr black hole.

- Let us consider again the Killing vector $K=\partial_t$. We have

$$K^\mu K_\mu = -\frac{1}{\rho^2}(\Delta - a^2\sin^2\theta). \tag{2.203}$$

At $r=r_+$ we have $K^\mu K_\mu = a^2\sin^2\theta/\rho^2 \geq 0$ and hence this vector is space like at the outer horizon except at $\theta=0$ (north pole) and $\theta=\pi$ (south pole) where it becomes null.

The so-called stationary limit surface or ergosurface is defined as the set of points where $K^\mu K_\mu = 0$. This is given by

$$\Delta = a^2\sin^2\theta \Leftrightarrow (r-GM)^2 = G^2M^2 - a^2\cos^2\theta. \tag{2.204}$$

The outer event horizon is given by

$$\Delta = 0 \Leftrightarrow (r_+-GM)^2 = G^2M^2 - a^2. \tag{2.205}$$

The region between the stationary limit surface and the outer event horizon is called the ergosphere. Inside the ergosphere the Killing vector K^μ is spacelike and thus observers can not remain stationary. In fact they must move in the direction of the rotation of the black hole but they can still move towards the event horizon or away from it.

- The naked singularity in Kerr spacetime occurs at $\rho=0$. Since $\rho^2=r^2+a^2\cos^2\theta$ we get the conditions

$$r=0\ ,\ \theta=\frac{\pi}{2}. \tag{2.206}$$

To exhibit what these conditions correspond to we substitute them in equation (2.178) which is valid in the limit $M\longrightarrow 0$. We obtain immediately $x^2+y^2=a^2$ which is a ring. This ring singularity is, of course, only a coordinate singularity in the limit $M\longrightarrow 0$. For $M\neq 0$ the ring singularity is indeed a true or naked singularity as one can explicitly check [23]. The rotation has therefore softened the naked singularity at $r=0$ of the Schwarzschild solution but spreading it over a ring.

- A sketch of the Kerr black hole is shown on figure GR5.

2.6.5 Penrose Process

The conserved energy of a massive particle with mass m in a Kerr spacetime is given by

$$\begin{aligned} E=-K_\mu p^\mu &= -g_{tt}K^tp^t - g_{t\phi}K^tp^\phi \\ &= m(1-\frac{2GMr}{\rho^2})\frac{dt}{d\tau} + \frac{2GmMar\sin^2\theta}{\rho^2}\frac{d\phi}{d\tau}. \end{aligned} \tag{2.207}$$

[23]Exercise: Show that $R_{\mu\nu\alpha\beta}R^{\mu\nu\alpha\beta}$ diverges at $\rho=0$.

The angular momentum of the particle is given by

$$
\begin{aligned}
L = R_\mu p^\mu &= g_{\phi\phi} R^\phi p^\phi + g_{\phi t} R^\phi p^t \\
&= \frac{m \sin^2\theta}{\rho^2}\big[(r^2+a^2)^2 - a^2\Delta\sin^2\theta\big]\frac{d\phi}{d\tau} - \frac{2GmMar\sin^2\theta}{\rho^2}\frac{dt}{d\tau}.
\end{aligned} \tag{2.208}
$$

The minus sign in the definition of the energy guarantees positivity since both K^μ and p^μ are time like vectors at infinity and as such their scalar product is negative. Inside the ergosphere the Killing vector K^μ becomes space like and thus it is possible to have particles for which $E = -K_\mu p^\mu < 0$.

We imagine an object starting outside the ergosphere with energy $E^{(0)}$ and momentum $p^{(0)}$ and falling into the black hole. The energy $E^{(0)} = -K^\mu p_\mu^{(0)}$ is positive and conserved along the geodesic. Once the object enters the ergosphere it splits into two with momenta $p^{(1)}$ and $p^{(2)}$. The object with momentum $p^{(1)}$ is allowed to escape back to infinity while the object with momentum $p^{(2)}$ falls into the black hole.We have the momentum and energy conservations $p^{(0)} = p^{(1)} + p^{(2)}$ and $E^{(0)} = E^{(1)} + E^{(2)}$. It is possible that the infalling object with momentum $p^{(2)}$ have negative energy $E^{(2)}$ and as a consequence $E^{(0)}$ will be less than $E^{(1)}$. In other words the escaping object can have more energy than the original infalling object. This so-called Penrose process allows us therefore to extract energy from the black hole which actually happens by decreasing its angular momentum. This process can be made more explicit as follows.

The outer event horizon of a Kerr black hole is a Killing horizon for the Killing vector $\chi^\mu = K^\mu + \Omega_H R^\mu$. This vector is normal to the event horizon and it is future pointing, i.e. it determines the forward direction in time. Thus the statement that the particle with momentum $p^{(2)}$ crosses the event horizon moving forward in time means that $-p^{(2)\mu}\chi_\mu \geq 0$. The analogue statement in a static spacetime is that particles with positive energy move forward in time, i.e. $E = -p^{(2)\mu} K_\mu \geq 0$. The condition $-p^{(2)\mu}\chi_\mu \geq 0$ is equivalent to

$$
L^{(2)} \leq \frac{E^{(2)}}{\Omega_H} < 0. \tag{2.209}
$$

Since $E^{(2)}$ is assumed to be negative and Ω_H is positive the angular momentum $L^{(2)}$ is negative and hence the particle with momentum $p^{(2)}$ is actually moving against the rotation of the black hole. After the particle with momentum $p^{(1)}$ escapes to infinity and the particle with momentum $p^{(2)}$ falls into the black hole the mass and the angular momentum of the Kerr black hole change (decrease) by the amounts

$$
\Delta M = E^{(2)}\ ,\ \Delta J = L^{(2)}. \tag{2.210}
$$

The bound $L^{(2)} \leq E^{(2)}/\Omega_H$ becomes

$$
\Delta J \leq \frac{\Delta M}{\Omega_H}. \tag{2.211}
$$

Thus extracting energy from the black hole (or equivalently decreasing its mass) is achieved by decreasing its angular momentum, i.e. by making the infalling particle carry angular momentum opposite to the rotation of the black hole.

In the limit when the particle with momentum $p^{(2)}$ becomes null tangent to the event horizon we get the ideal process $\Delta J = \Delta M/\Omega_H$.

2.7 Black Holes Thermodynamics

Let us start this section by calculating the area of the outer event horizon $r = r_+$ of a Kerr black hole. Recall first that

$$r_+ = GM + \sqrt{G^2M^2 - a^2}. \tag{2.212}$$

We need the induced metric γ_{ij} on the outer event horizon. Since the outer event horizon is defined by $r = r_+$ the coordinates on the outer event horizon are θ and ϕ. We set therefore $r = r_+$ $(\Delta = 0)$, $dr = 0$ and $dt = 0$ in the Kerr metric. We obtain the metric

$$\begin{aligned} ds^2|_{r=r_+} &= \gamma_{ij}dx^i dx^j \\ &= g_{\theta\theta}d\theta^2 + g_{\phi\phi}d\phi^2 \\ &= (r_+^2 + a^2\cos^2\theta)d\theta^2 + \frac{(r_+^2 + a^2)^2\sin^2\theta}{r_+^2 + a^2\cos^2\theta}d\phi^2. \end{aligned} \tag{2.213}$$

The area of the horizon can be constructed from the induced metric as follows

$$\begin{aligned} A &= \int \sqrt{|\det\gamma|}d\theta d\phi \\ &= \int (r_+^2 + a^2)\sin\theta d\theta d\phi \\ &= 4\pi(r_+^2 + a^2) \\ &= 8\pi G^2\big(M^2 + \sqrt{M^4 - \frac{M^2a^2}{G^2}}\big) \\ &= 8\pi G^2\big(M^2 + \sqrt{M^4 - \frac{J^2}{G^2}}\big). \end{aligned} \tag{2.214}$$

The area is related to the so-called irreducible mass M_{irr}^2 by

$$\begin{aligned} M_{\text{irr}}^2 &= \frac{A}{16\pi G^2} \\ &= \frac{1}{2}\big(M^2 + \sqrt{M^4 - \frac{J^2}{G^2}}\big). \end{aligned} \tag{2.215}$$

The area (or equivalently the irreducible mass) depends on the two parameters characterizing the Kerr black hole, namely its mass and its angular momentum. From the other hand we know that the mass and the angular momentum of the Kerr black hole decrease in the Penrose process. Thus the area changes in the Penrose process as follows

$$\begin{aligned}\Delta A &= \frac{8\pi G}{\sqrt{G^2M^2-a^2}}\big[2GMr_+\Delta M - a\Delta J\big]\\ &= \frac{8\pi G}{\sqrt{G^2M^2-a^2}}\big[(r_+^2+a^2)\Delta M - a\Delta J\big]\\ &= \frac{8\pi G(r_+^2+a^2)}{\sqrt{G^2M^2-a^2}}\big[\Delta M - \Omega_H\Delta J\big]\\ &= \frac{8\pi G a}{\Omega_H\sqrt{G^2M^2-a^2}}\big[\Delta M - \Omega_H\Delta J\big].\end{aligned} \tag{2.216}$$

This is equivalent to

$$\Delta M_{\rm irr}^2 = \frac{a}{2G\sqrt{G^2M^2-a^2}}\big[\frac{\Delta M}{\Omega_H}-\Delta J\big] \Leftrightarrow \Delta M_{\rm irr} = \frac{a}{4GM_{\rm irr}\sqrt{G^2M^2-a^2}}\big[\frac{\Delta M}{\Omega_H}-\Delta J\big]. \tag{2.217}$$

However we have already found that in the Penrose process we must have $\Delta J \leq \Delta M/\Omega_H$. This leads immediately to

$$\Delta M_{\rm irr} \geq 0. \tag{2.218}$$

The irreducible mass can not decrease. From this result we deduce immediately that

$$\Delta A \geq 0. \tag{2.219}$$

This is the second law of black hole thermodynamics or the area theorem which states that the area of the event horizon is always non decreasing. The area in black hole thermodynamics plays the role of entropy in thermodynamics.

We can use equation (2.215) to express the mass of the Kerr black hole in terms of the irreducible mass $M_{\rm irr}$ and the angular momentum J. We find

$$\begin{aligned}M^2 &= M_{\rm irr}^2 + \frac{J^2}{4G^2M_{\rm irr}^2}\\ &= \frac{A}{16\pi G^2} + \frac{4\pi J^2}{A}.\end{aligned} \tag{2.220}$$

Now we imagine a Penrose process which is reversible, i.e. we reduce the angular momentum of the black hole from J_i to J_f such that $\Delta A = 0$ (clearly $\Delta A > 0$ is not a reversible process simply because the reverse process violates the area theorem). Then

$$M_i^2 - M_f^2 = \frac{4\pi}{A}(J_i^2 - J_f^2). \tag{2.221}$$

If we consider $J_f = 0$ then we obtain

$$M_i^2 - M_f^2 = \frac{4\pi}{A}J_i^2 \Leftrightarrow M_f^2 = \frac{A}{16\pi G^2} = M_{\rm irr}^2. \tag{2.222}$$

In other words if we reduce the angular momentum of the Kerr black hole to zero, i.e. until the black hole stop rotating, then its mass will reduce to a minimum value given precisely by $M_{\rm irr}$. This is why this is called the irreducible mass. In fact $M_{\rm irr}$ is the mass of the resulting Schwarzschild black hole. The maximum energy we can therefore extract from a Kerr black hole via a Penrose process is $M - M_{\rm irr}$. We have

$$E_{\rm max} = M - M_{\rm irr} = M - \frac{1}{\sqrt{2}}\sqrt{M^2 + \sqrt{M^4 - \frac{J^2}{G^2}}}. \tag{2.223}$$

The irreducible mass is minimum at $M^2 = J/G$ or equivalently $GM = a$ (which is the case of extremal Kerr black hole) and as a consequence $E_{\rm max}$ is maximum for $GM = a$. At this point

$$E_{\rm max} = M - M_{\rm irr} = M - \frac{1}{\sqrt{2}}M = 0.29M. \tag{2.224}$$

We can therefore extract at most 29 per cent of the original mass of Kerr black hole via Penrose process.

The first law of black hole thermodynamics is essentially given by equation (2.216). This result can be rewritten as

$$\Delta M = \frac{\kappa}{8\pi G}\Delta A + \Omega_H \Delta J. \tag{2.225}$$

The constant κ is called the surface gravity of the Kerr black hole and it is given by

$$\begin{aligned}\kappa &= \frac{\Omega_H\sqrt{G^2M^2 - a^2}}{a} \\ &= \frac{\sqrt{G^2M^2 - a^2}}{r_+^2 + a^2} \\ &= \frac{\sqrt{G^2M^2 - a^2}}{2GM(GM + \sqrt{G^2M^2 - a^2})}.\end{aligned} \tag{2.226}$$

The above first law of black hole thermodynamics is similar to the first law of thermodynamics $dU = TdS - pdV$ with the most important identifications

$$\begin{aligned} U &\leftrightarrow M \\ S &\leftrightarrow \frac{A}{4G} \\ T &\leftrightarrow \frac{\kappa}{2\pi}.\end{aligned} \tag{2.227}$$

The quantity $\kappa\Delta A/(8\pi G)$ is heat energy while $\Omega_H\Delta J$ is the work done on the black by throwing particles into it.

The zeroth law of black hole thermodynamics states that surface gravity is constant on the horizon. Again this is the analogue of the zeroth law of thermodynamics which states that temperature is constant throughout a system in thermal equilibrium.

Chapter 3

Cosmology I: The Observed Universe

The modern science of cosmology is based on three basic observational results:

- The universe, on very large scales, is homogeneous and isotropic.
- The universe is expanding.
- The universe is composed of: matter, radiation, dark matter and dark energy.

3.1 Homogeneity and Isotropy

The universe is expected to look exactly the same from every point in it. This is the content of the so-called Copernican principle. On the other hand, the universe appears perfectly isotropic to us on Earth. Isotropy is the property that at every point in spacetime all spatial directions look the same, i.e. there are no preferred directions in space. The isotropy of the observed universe is inferred from the cosmic microwave background (CMB) radiation, which is the most distant electromagnetic radiation originating at the time of decoupling, and which is observed at around 3 K, which is found to be isotropic to at least one part in a thousand by various experiments such as COBE, WMAP and PLANK.

The 9 years results of the Wilkinson Microwave Anisotropy Probe (WMAP) for the temperature distribution across the whole sky are shown on figure (3.1). The microwave background is very homogeneous in temperature with a mean of 2.7 K and relative variations from the mean of the order of 5×10^{-5} K. The temperature variations are presented through different colours with the "red" being hotter (2.7281 K) while the "blue" being colder 2.7279 K than the average. These fluctuations about isotropy are extremely important since they will lead, in the theory of inflation, by means gravitational interactions, to structure formation.

The Copernican principle together with the observed isotropy means in particular that the universe on very large scales must look homogeneous and isotropic. Homogeneity is the property that all points of space look the same at every instant of time. This is the content of the so-called cosmological principle. Homogeneity is verified directly by constructing three dimensional maps of the distribution of galaxies such as the 2−Degree-Field Galaxy Redshift survey (2dFGRS) and the Sloan Digital Sky survey (SDSS). A slice through the SDSS 3−dimensional map of the distribution of galaxies with the Earth at the center is shown on figure (3.2).

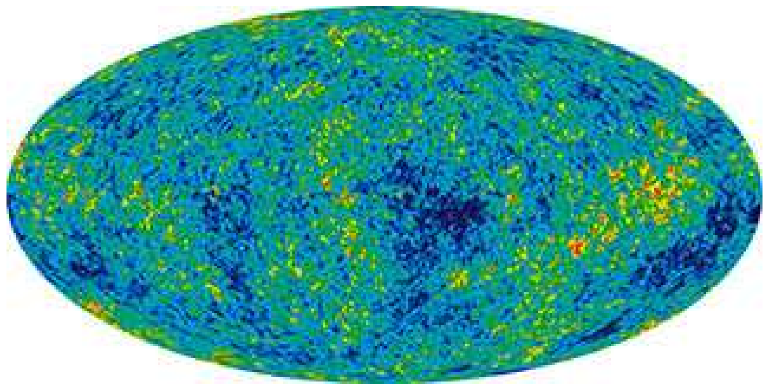

Figure 3.1: The all-sky map of the CMB. Source: http://map.gsfc.nasa.gov/news/.

3.2 Expansion and Distances

3.2.1 Hubble Law

The most fundamental fact about the universe is its expansion. This can be characterized by the so-called scale factor $a(t)$. At the present time t_0 we set $a(t_0) = 1$. At earlier times, when the universe was much smaller, the value of $a(t)$ was much smaller.

Spacetime can be viewed as a grid of points where the so-called comoving distance between the points remains constant with the expansion, since it is associated with the coordinates chosen on the grid, while the physical distance evolves with the expansion of the universe linearly with the scale factor and the comoving distance, viz

$$\text{distance}_{\text{physical}} = a(t) \times \text{distance}_{\text{comoving}}. \tag{3.1}$$

In an expanding universe galaxies are moving away from each other. Thus galaxies must be receding from us. Now, we know from the Doppler effect that the wavelength of sound or light emitted from a receding source is stretched out in the sense that the observed wavelength is larger than the emitted wavelength. Thus the spectra of galaxies, since they are receding from us, must be redshifted. This can be characterized by the so-called redshift z defined by

$$1 + z = \frac{\lambda_{\text{obs}}}{\lambda_{\text{emit}}} \geq 1 \Leftrightarrow z = \frac{\Delta\lambda}{\lambda}. \tag{3.2}$$

For low redshifts $z \longrightarrow 0$, i.e. for sufficiently close galaxies with receding velocities much smaller than the speed of light, the standard Doppler formula must hold, viz

$$z = \frac{\Delta\lambda}{\lambda} \simeq \frac{v}{c}. \tag{3.3}$$

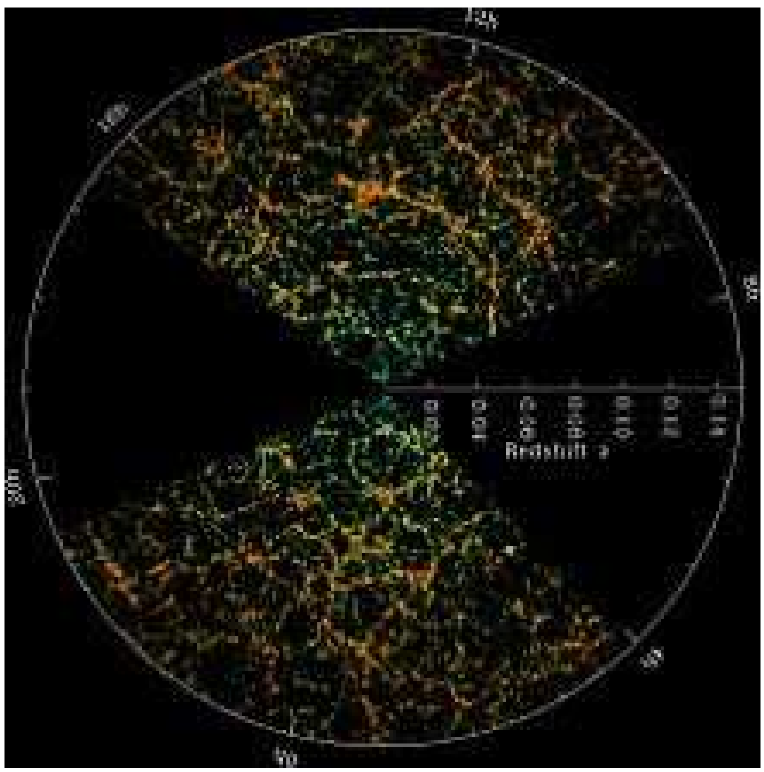

Figure 3.2: The Sloan Digital Sky survey. Source: http://www.sdss3.org/dr10/.

This allows us to determine the expansion velocities of galaxies by measuring the redshifts of absorption and emission lines. This was done originally by Hubble in 1929. He found a linear relation between the velocity v of recession and the distance d given by

$$v = H_0 d. \tag{3.4}$$

This is the celebrated Hubble law exhibited on figure (3.3). The constant H_0 is the Hubble constant given by the value

$$H_0 = 72 \pm 7(\mathrm{km}/s)/\mathrm{Mpc}. \tag{3.5}$$

The Mpc is megapersec which is the standard unit of distances in cosmology. We have

$$1 \text{ parsec(pc)} = 3.08 \times 10^{18}\mathrm{cm} = 3.26 \text{ light} - \text{year}. \tag{3.6}$$

The Hubble law can also be seen as follows. Starting from the formula relating the physical distance to the comoving distance $d = ax$, and assuming no comoving motion $\dot{x} = 0$, we can show immediately that the relative velocity $v = \dot{d}$ is given by

$$v = Hd \ , \ H = \frac{\dot{a}}{a}. \tag{3.7}$$

The Hubble constant sets essentially the age of the universe by keeping a constant velocity. We get the estimate

$$t_H = \frac{1}{H_0} \sim 14 \text{ billion years}. \tag{3.8}$$

This is believed to be the time of the initial singularity known as the big bang where density, temperature and curvature were infinite.

3.2.2 Cosmic Distances from Standard Candles

It is illuminating to start by noting the following distances:

- The distance to the edge of the observable universe is 14Gpc.
- The size of the largest structures in the universe is around 100Mpc.
- The distance to the nearest large cluster, the Virgo cluster which contains several thousands galaxies, is 20Mpc.
- The distance to a typical galaxy in the local group which contains 30 galaxies is $50 - 1000$kpc. For example, Andromeda is 725kpc away.
- The distance to the center of the Milky Way is 10kpc.
- The distance to the nearest star is 1pc.

- The distance to the Sun is 5μpc.

But the fundamental question that one must immediately pose, given the immense expanses of the universe, how do we come up with these numbers?

- **Triangulation:** We start with distances to nearby stars which can be determined using triangulation. The angular position of the star is observed from 2 points on the orbit of Earth giving two angles α and β, and as a consequence, the parallax p is given by

$$2p = \pi - \alpha - \beta. \tag{3.9}$$

For nearby stars the parallax p is a sufficiently small angle and thus the distance d to the star is given by (with a the semi-major axis of Earth's orbit)

$$d = \frac{a}{p}. \tag{3.10}$$

This method was used, by the Hipparchos satellite, to determine the distances to around 120000 stars in the solar neighborhood.

- **Standard Candles:** Most cosmological distances are obtained using the measurements of apparent luminosity of objects of supposedly known intrinsic luminosity. Standard candles are objects, such as stars and supernovae, whose intrinsic luminosity are determined from one of their physical properties, such as color or period, which itself is determined independently. Thus a standard candle is a source with known intrinsic luminosity.

The intrinsic or absolute luminosity L, which is the energy emitted per unit time, of a star is related to its distance d, determined from triangulation, and to the flux l by the equation

$$L = l.4\pi d^2. \tag{3.11}$$

The flux l is the apparent brightness or luminosity which is the energy received per unit time per unit area. By measuring the flux l and the distance d we can calculate the absolute luminosity L.

Now, if all stars with a certain physical property, for example a certain blue color, and for which the distances can be determined by triangulation, turn out to have the same intrinsic luminosity, these stars will constitute standard candles. In other words, all blue color stars will be assumed to have the following luminosity:

$$L_{\text{blue color stars}} = l.4\pi d^2_{\text{triangulation}}. \tag{3.12}$$

This means that for stars farther away with the same blue color, for which triangulation does not work, their distances can be determined by the above formula (3.11) assuming the same intrinsic luminosity (3.12) and only requiring the determination of their flux l at Earth, viz

$$d_{\text{blue color stars}} = \sqrt{\frac{L_{\text{blue color stars}}}{4\pi l}}. \tag{3.13}$$

Some of the standard candles are:

- **Main Sequence Stars:** These are stars who still burn hydrogen at their cores producing helium through nuclear fusion. They obey a characteristic relation between absolute luminosity and color which both depend on the mass. For example, the luminosity is maximum for blue stars and minimum for red stars. The position of a star along the main sequence is essentially determined by its mass. This is summarized in a so-called Hertzsprung-Russell diagram which plots the intrinsic or absolute luminosity against its color index. An example is shown in figure (3.4).

  All main-sequence stars are in hydrostatic equilibrium since the outward thermal pressure from the hot core is exactly balanced by the inward pressure of gravitational collapse. The main-sequence stars with mass less than $0.23M_O$ will evolve into white dwarfs, whereas those with mass less than $10M_O$ will evolve into red giants. Those main-sequence stars with more mass will either gravitationlly collapse into black holes or explode into supernova.

  The HR diagram of main-sequence stars is calibrated using triangulation: The absolute luminosity, for a given color, is measured by measuring the apparent luminosity and the distance from triangulation and then using the inverse square law (3.11).

  By determining the luminosity class of a star, i.e. whether or not it is a main-sequence star, and determining its position on the HR diagram, i.e. its color, we can determine its absolute luminosity. This allows us to calculate its distance from us by measuring its apparent luminosity and using the inverse square law (3.11).

- **Cepheid Variable Stars:** These are massive, bright, yellow stars which arise in a post main-sequence phase of evolution with luminosity of upto $1000 - 10000$ times greater than that of the Sun. These stars are also pulsating, i.e. they grow and shrink in size with periods between 3 and 50 days. They are named after the δ Cephei star in the constellation Cepheus which is the first star of this kind. These stars lie in the so-called instability strip of the HR diagram (3.5).

  The established strong correlation between the luminosity and the period of pulsation allows us to use Cepheid stars as standard candles. By determining the variability of a given Cepheid star, we can determine its absolute luminosity by determining its position on the period-luminosity diagram such as (3.6). From this we can determine its distance from us by determining its apparent luminosity and using the inverse square law (3.11).

  The period-luminosity diagram is calibrated using main-sequence stars and triangulation. For example, Hipparchos satellite had provided true parallaxes for a good

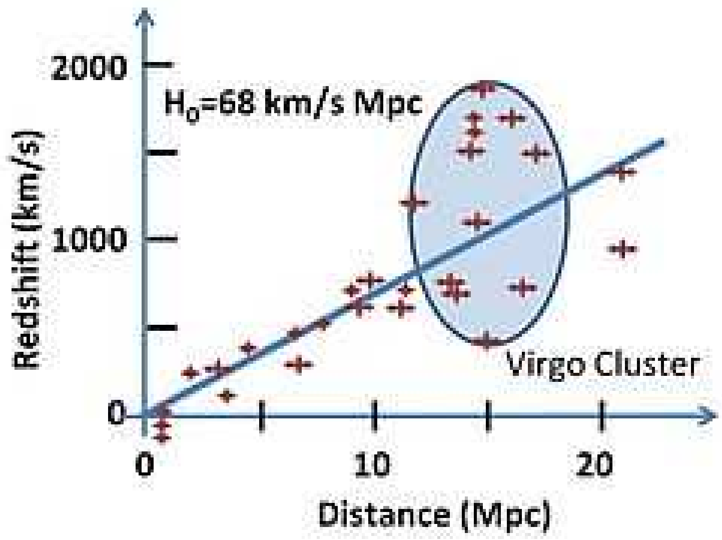

Figure 3.3: The Hubble law. Source: Wikipedia.

sample of Galactic Cepheids.

– **Type Ia Supernovae:** These are the only very far away discrete objects within galaxies that can be resolved due to their brightness which can rival even the brightness of the whole host galaxy. Supernovae are 100000 times more luminous than even the brightest Cepheid, and several billion times more luminous than the Sun.

  Type Ia supernova occurs when a white dwarf star in a binary system accretes sufficient matter from its companion until its mass reaches the Chandrasekhar limit which is the maximum possible mass that can be supported by electron degeneracy pressure. The white dwarf becomes then unstable and explodes. These explosions are infrequent and even in a large galaxy only one supernova per century occurs on average.

  The exploding white dwarf star in a supernova has always a mass close to the Chandrasekhar limit of $1.4M_O$ and as a consequence all supernovae are basically the same, i.e. they have the same absolute luminosity. This absolute luminosity can be calculated by observing supernovae which occur in galaxies whose distances were determined using Cepheid stars. Then we can use this absolute luminosity to measure distances to even farther galaxies, for which Cepheid stars are not available, by observing supernovae in those galaxies and determining their apparent luminosities and using the inverse square law (3.11).

- **Cosmic Distance Ladder:** Triangulation and the standard candles discussed above: main-sequence stars, Cepheid variable stars and type Ia supernovae provide a cosmic distance ladder.

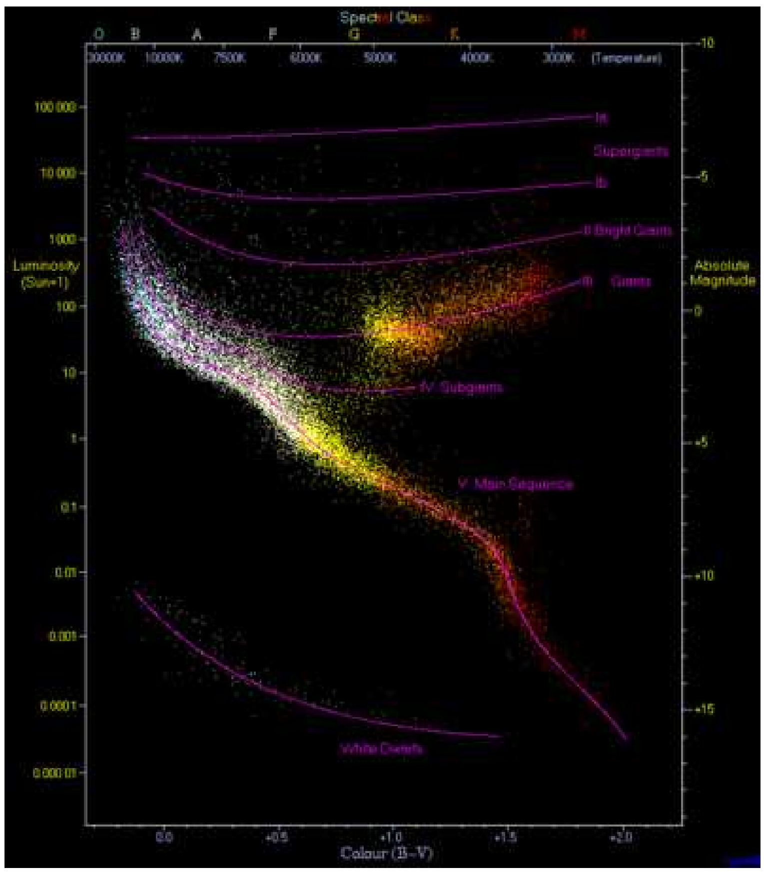

Figure 3.4: The HertzsprungRussell diagram of 22000 stars from the Hipparcos catalogue together with 1000 low-luminosity stars, red and white dwarfs, from the Gliese catalogue of Nearby stars. Source: Wikipedia.

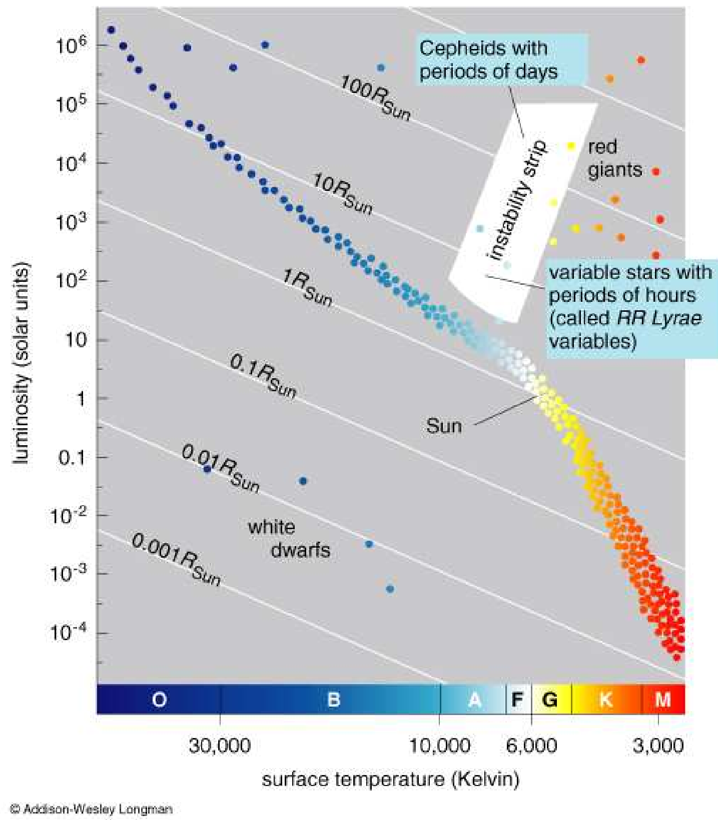

Figure 3.5: The instability strip. Source: http://www.astro.sunysb.edu/metchev/PHY515/cepheidpl.html.

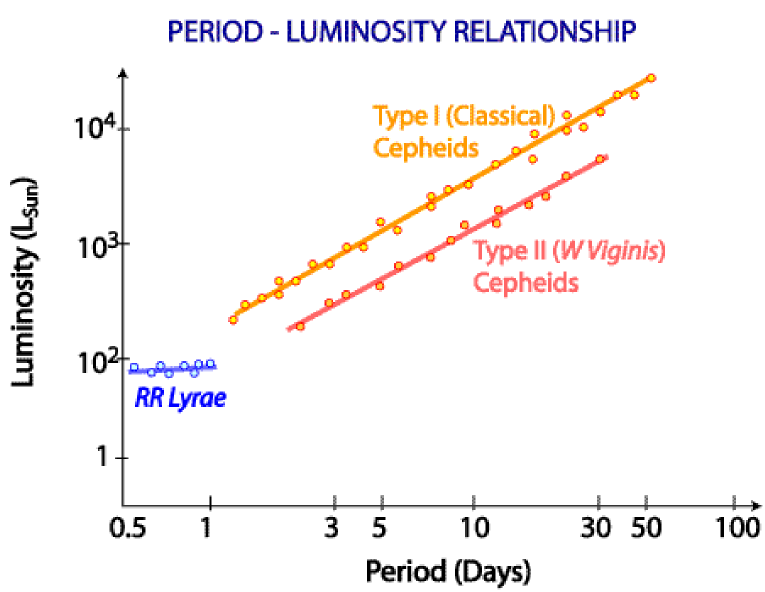

Figure 3.6: The period-luminosity relation. Source:http://www.atnf.csiro.au/outreach/education/senior/astro

3.3 Matter, Radiation, and Vacuum

- **Matter:** This is in the form of stars, gas and dust held together by gravitational forces in bound states called galaxies. The iconic Hubble deep field image, which covers a tiny portion of the sky 1/30th the diameter of the full Moon, is perhaps the most conclusive piece of evidence that galaxies are the most important structures in the universe. See figure (3.7). The observed universe may contain 10^{11} galaxies, each one contains around 10^{11} stars with a total mass of $10^{12}M_O$. The density of this visible matter is roughly given by

$$\rho_{\rm visible} = 10^{-31}g/{\rm cm}^3. \tag{3.14}$$

- **Radiation[1]:** This consists of zero-mass particles such as photons, gravitons (gravitational waves) and in many circumstances (neutrinos) which are not obviously bound by gravitational forces. The most important example of radiation observed in the universe is the cosmic microwave background (CMB) radiation with density given by

$$\rho_{\rm radiation} = 10^{-34}g/{\rm cm}^3. \tag{3.15}$$

  This is much smaller than the observed matter density since we are in a matter dominated phase in the evolution of the universe. This CMB radiation is an electromagnetic radiation left over from the hot big bang, and corresponds to a blackbody spectrum with a temperature of $T = 2.725 \pm 0.001K$. See Figure (3.8).

- **Dark Matter:** This is the most important form of matter in the universe in the sense that most mass in the universe is not luminous (the visible matter) but dark although its effect can still be seen from its gravitational effect.

  It is customary to dynamically measure the mass of a given galaxy by using Kepler's third law:

$$GM(r) = v^2(r)r. \tag{3.16}$$

  In the above equation we have implicitly assumed spherical symmetry, $v(r)$ is the orbital (rotational) velocity of the galaxy at a distance r from the center, and $M(r)$ is the mass inside r. The plot of $v(r)$ as a function of the distance r is known as the rotation curve of the galaxy.

  Applying this law to spiral galaxies, which are disks of stars and dust rotating about a central nucleus, taking r the radius of the galaxy, i.e. the radius within which much of the light emitted by the galaxy is emitted, one finds precisely the mass density $\rho_{\rm visible} = 10^{-31}g/{\rm cm}^3$ quoted above. This is the luminous mass density since it is associated with the emission of light.

[1]Strictly speaking radiation should be included with matter.

Optical observations are obviously limited due to the interstellar dust which does not allow the penetration of light waves. However, this problem does not arise when making radio measurements of atomic hydrogen. More precisely, neutral hydrogen (HI) atoms, which are abundant and ubiquitous in low density regions of the interstellar medium, are detectable in the 21 cm hyperfine line. This transition results from the magnetic interaction between the quantized electron and proton spins when the relative spins change from parallel to antiparallel.

Observations of the 21 cm line from neutral hydrogen regions in spiral galaxies can therefore be used to measure the speed of rotation of objects. More precisely, since objects in galaxies are moving, they are Doppler shifted and the receiver can determine their velocities by comparing the observed wavelengths to the standard wavelength of 21 cm. By extending to distances beyond the point where light emitted from the galaxy effectively ceases, one finds the behavior, shown on figure (3.9) which is given by

$$v \sim \text{constant}. \tag{3.17}$$

We would have expected that outside the radius of the galaxy, with the luminous matter providing the only mass, the velocity should have behaved as

$$v \sim 1/r^{1/2}. \tag{3.18}$$

The result (3.17) indicates that even in the outer region of the galaxies the mass behaves as

$$M(r) \sim r. \tag{3.19}$$

In other words, the mass always grows with r. We conclude that spiral galaxies, and in fact most other galaxies, contain dark, i.e. invisible, matter which permeates the galaxy and extends into the galaxy's halo with a density of at least 3 to 10 times the mass density of the visible matter, viz

$$\rho_{\rm halo} = (3-10) \times \rho_{\rm visible}. \tag{3.20}$$

This form of matter is expected to be 1) mostly nonbaryonic , 2)cold, i.e. nonrelativistic during most of the universe history, so that structure formation is not suppressed and 3) very weakly interacting since they are hard to detect. The most important candidate for dark matter is WIMP (weakly interacting massive particle) such as the neutralinos which is the lightest of the additional stable particles predicted by supersymmetry with mass around 100 GeV.

- **Dark Energy:** This is speculated to be the energy of empty space, i.e. vacuum energy, and is the dominant component in the universe: around 70 per cent. The best candidate for dark energy is usually identified with the cosmological constant.

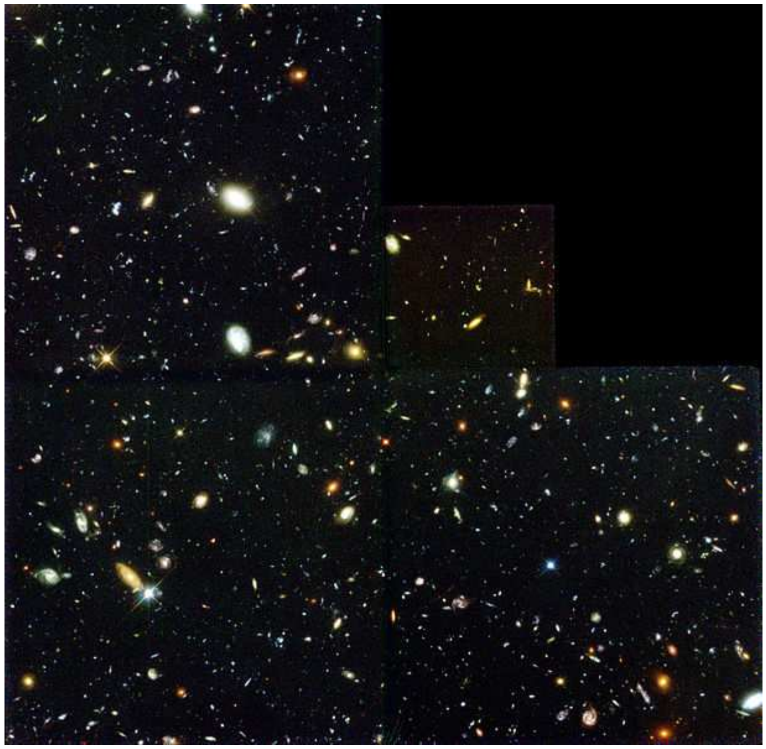

Figure 3.7: The Hubble deep field. Source: Wikipedia.

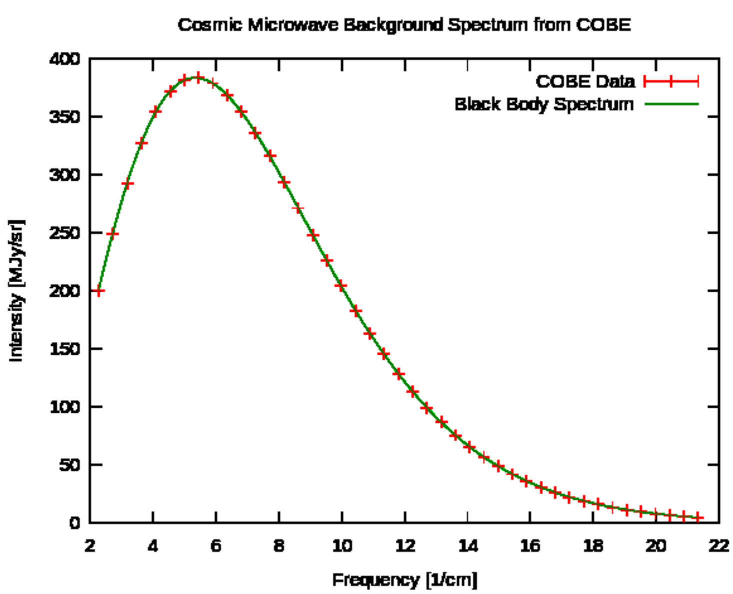

Figure 3.8: The black body spectrum. Source: Wikipedia.

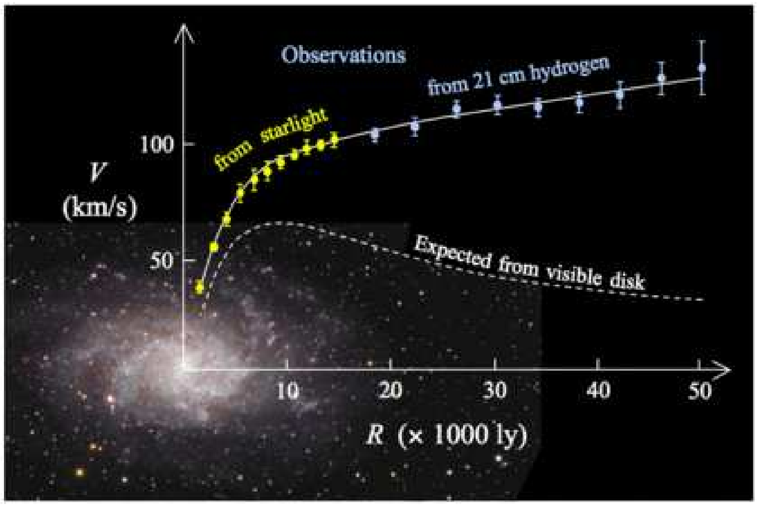

Figure 3.9: The galaxy rotation curve. Source: Wikipedia.

3.4 Flat Universe

The simplest isotropic and homogeneous spacetime is the one in which the line element is given by

$$\begin{aligned} ds^2 &= -dt^2 + a^2(t)(dx^2 + dy^2 + dz^2) \\ &= -dt^2 + a^2(t)(dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)). \end{aligned} \quad (3.21)$$

The function $a(t)$ is the scale factor. This is a flat universe. The homogeneity, isotropy and flatness are properties of the space and not spacetime.

The coordinate or comoving distance between any two points is given by

$$d_{\text{comoving}} = \sqrt{\Delta x^2 + \Delta y^2 + \Delta z^2}. \quad (3.22)$$

This is for example the distance between any pair of galaxies. This distance is constant in time which can be seen as follows. Since we will view the distribution of galaxies as a smoothed out cosmological fluid, and thus a given galaxy is a particle in this fluid with coordinates x^i, the velocity dx^i/dt of the galaxy must vanish, otherwise it will provide a preferred direction contradicting the isotropy property. On the other hand, the physical distance between any two points depends on time and is given by

$$d_{\text{physical}}(t) = a(t) d_{\text{comoving}}. \quad (3.23)$$

Clearly if $a(t)$ increases with time then the physical distance $d_{\text{physical}}(t)$ must increase with time which is what happens in an expanding universe.

The energy of a particle moving in this spacetime will change similarly to the way that the energy of a particle moving in a time-dependent potential will change. For a photon this change in energy is precisely the cosmological redshift. The worldline of the photon satisfies

$$ds^2 = 0. \tag{3.24}$$

By assuming that we are at the origin of the spherical coordinates r, θ and ϕ, and that the photon is emitted in a galaxy a comoving distance $r = R$ away with a frequency ω_e at time t_e, and is received here at time $t = t_0$ with frequency ω_0, the worldline of the photon is therefore the radial null geodesics

$$ds^2 = -dt^2 + a^2(t)dr^2 = 0. \tag{3.25}$$

Integration yields immediately

$$R = \int_{t_e}^{t_0} \frac{dt}{a(t)}. \tag{3.26}$$

For a photon emitted at time $t_e + \delta t_e$ and observed at time $t_0 + \delta t_0$ we will have instead

$$R = \int_{t_e+\delta t_e}^{t_0+\delta t_0} \frac{dt}{a(t)}. \tag{3.27}$$

Thus we get

$$\frac{\delta t_0}{a(t_0)} = \frac{\delta t_e}{a(t_e)}. \tag{3.28}$$

In particular if δt_e is the period of the emitted light, i.e. $\delta t_e = 1/\nu_e$, the period of the observed light will be different given by $\delta t_0 = 1/\nu_0$. The relation between ν_e and ν_0 defines the redshift z through

$$1 + z = \frac{\lambda_0}{\lambda_e} = \frac{\nu_e}{\nu_0} = \frac{a(t_0)}{a(t_e)}. \tag{3.29}$$

This can be rewritten as

$$z = \frac{\Delta\lambda}{\lambda} = \frac{a(t_0) - a(t_e)}{a(t_e)} = \frac{\dot{a}(t_e)}{a(t_e)}(t_e - t_0) + ... \tag{3.30}$$

The physical distance d is related to the comoving distance R by $d = a(t_0)R$. By assuming that R is small we have from $ds^2 = 0$ the result

$$t_e - t_0 = \int_0^R a(t)dr = a(t_0)R + O(R^2). \tag{3.31}$$

Thus

$$z = \frac{\Delta\lambda}{\lambda} = \frac{\dot{a}(t_0)}{a(t_0)}d + ... \tag{3.32}$$

This is Hubble law. The Hubble constant is

$$H_0 = \frac{\dot{a}(t_0)}{a(t_0)}. \tag{3.33}$$

The Hubble time t_H and the Hubble distance d_H are defined by

$$t_H = \frac{1}{H_0}\ ,\ d_H = ct_H. \tag{3.34}$$

The line element (7.74) is called the flat Robertson-Walker metric, and when the scale factor $a(t)$ is specified via Einstein's equations, it is called the flat Friedman-Robertson-Walker metric. The time evolution of the scale factor $a(t)$ is controled by the Friedman equation

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G\rho}{3}. \tag{3.35}$$

For a detailed derivation see next chapter. ρ is the total mass-energy density. At the present time this equation gives the relation between the Hubble constant H_0 and the critical mass density ρ_c given by

$$H_0^2 \equiv \frac{\dot{a}^2(t_0)}{a^2(t_0)} = \frac{8\pi G\rho(t_0)}{3} \Rightarrow \rho(t_0) = \frac{3H_0^2}{8\pi G} \equiv \rho_c. \tag{3.36}$$

We can choose, without any loss of generality, $a(t_0) = 1$. We have the following numerical values

$$H_0 = 100h(\mathrm{km}/s)/\mathrm{Mpc}\ ,\ t_H = 9.78h^{-1}\mathrm{Gyr}\ ,\ d_H = 2998h^{-1}\mathrm{Mpc}\ ,\ \rho_c = 1.88\times 10^{-29}h^2 g/\mathrm{cm}^3. \tag{3.37}$$

The matter, radiation and vacuum contributions to the critical mass density are given by the fractions

$$\Omega_M = \frac{\rho_M(t_0)}{\rho_c}\ ,\ \Omega_R = \frac{\rho_R(t_0)}{\rho_c}\ ,\ \Omega_V = \frac{\rho_V(t_0)}{\rho_c}. \tag{3.38}$$

Obviously

$$1 = \Omega_M + \Omega_R + \Omega_V. \tag{3.39}$$

The generalization of this equation to $t \neq t_0$ is given by

$$\rho(a) = \rho_c(\frac{\Omega_M}{a^3} + \frac{\Omega_R}{a^4} + \Omega_V). \tag{3.40}$$

This can be derived as follows. By employing the principle of local conservation as expressed by the first law of thermodynamics we have

$$dE = -PdV. \tag{3.41}$$

Thus the change in the total energy in a volume V, containing a fixed number of particles and a pressure P, due to any change dV in the volume is equal to the work done on it. The heat flow in any direction is zero because of isotropy. Alternatively because of homogeneity the temperature T depends only on time and thus no place is hotter or colder than any other.

The volume dV is the physical volume and thus it is related to the time-independent co-moving volume $dV_{\rm comoving} = dxdydz$ by $dV = a^3(t)dV_{\rm comoving}$. On the other hand, the energy E is given in terms of the density ρ by $E = \rho dV$. The first law of thermodynamics becomes

$$\frac{d}{dt}\big(\rho a^3(t)\big) = -P\frac{d}{dt}\big(a^3(t)\big). \quad (3.42)$$

We have the following three possibilities

- **Matter-Dominated Universe:** In this case galaxies are approximated by a pressureless dust and thus $P_M = 0$. Also in this case all the energy comes from the rest mass since kinetic motion is neglected. We get then

$$\frac{d}{dt}\big(\rho_M a^3(t)\big) = 0 \Rightarrow \rho_M(t) = \rho_M(t_0)\frac{a^3(t_0)}{a^3(t)}. \quad (3.43)$$

- **Radiation-Dominated Universe:** In this case $P_R = \rho_R/3$ (see below for a proof). Thus

$$\frac{d}{dt}\big(\rho_R a^3(t)\big) = -\frac{1}{3}\rho_R\frac{d}{dt}\big(a^3(t)\big) \Rightarrow \rho_R(t) = \rho_R(t_0)\frac{a^4(t_0)}{a^4(t)}. \quad (3.44)$$

  It is not difficult to check that radiation dominates matter when the scale factor satisfies $a(t) \leq a(t_0)/1000$, i.e. when the universe was $1/1000$ of its present size. Thus over most of the universe history matter dominated radiation.

- **Vacuum-Dominated Universe:** In this case $P_V = -\rho_V$. Thus

$$\frac{d}{dt}\big(\rho_V a^3(t)\big) = \rho_V\frac{d}{dt}\big(a^3(t)\big) \Rightarrow \rho_V(t) = \rho_V(t_0)\frac{a^0(t_0)}{a^0(t)}. \quad (3.45)$$

  In other words, ρ_V is always a constant and thus, unlike matter and radiation, it does not decay away with the expansion of the universe. In particular, the future of any perpetually expanding universe will be dominated by vacuum energy. In the case of a cosmological constant we write

$$\rho_V = \frac{c^4}{8\pi G}\Lambda. \quad (3.46)$$

We compute immediately

$$\Omega_M(t) = \frac{\rho_M(t)}{\rho_c} = \frac{\Omega_M}{a^3}. \quad (3.47)$$

$$\Omega_R(t) = \frac{\rho_R(t)}{\rho_c} = \frac{\Omega_R}{a^4}. \tag{3.48}$$

$$\Omega_V(t) = \frac{\rho_V(t)}{\rho_c} = \frac{\Omega_V}{a^0}. \tag{3.49}$$

The total mass-energy density is given by $\rho(t) = \rho_c \Omega(t) = \rho_c(\Omega_M(t) + \Omega_R(t) + \Omega_V(t)$ or equivalently

$$\rho(a) = \rho_c(\frac{\Omega_M}{a^3} + \frac{\Omega_R}{a^4} + \Omega_V). \tag{3.50}$$

By using this last equation in the Friedmann equation we get the equivalent equation

$$\frac{1}{2H_0^2}\dot{a}^2 + V_{\rm eff}(a) = 0 \ , \ V_{\rm eff}(a) = -\frac{1}{2}\big(\frac{\Omega_M}{a} + \frac{\Omega_R}{a^2} + a^2\Omega_V\big). \tag{3.51}$$

This is effectively the equation of motion of a zero-energy particle moving in one dimension under the influence of the potential $V_{\rm eff}(a)$. The three possible distinct solutions are:

- **Matter-Dominated Universe:** In this case $\Omega_M = 1$, $\Omega_R = \Omega_V = 0$ and thus

$$V_{\rm eff}(a) = -\frac{1}{2a} \Rightarrow \frac{1}{2H_0^2}\dot{a}^2 - \frac{1}{2a} = 0 \Rightarrow a = \big(\frac{t}{t_0}\big)^{2/3} \ , \ t_0 = \frac{2}{3H_0}. \tag{3.52}$$

- **Radiation-Dominated Universe:** In this case $\Omega_R = 1$, $\Omega_M = \Omega_V = 0$ and thus

$$V_{\rm eff}(a) = -\frac{1}{2a^2} \Rightarrow \frac{1}{2H_0^2}\dot{a}^2 - \frac{1}{2a^2} = 0 \Rightarrow a = \big(\frac{t}{t_0}\big)^{1/2} \ , \ t_0 = \frac{1}{2H_0}. \tag{3.53}$$

  In this case, as well as in the matter-dominated case, the universe starts at $t = 0$ with $a = 0$ and thus $\rho = \infty$, and then expands forever. This physical singularity is what we mean by the big bang. Here the expansion is decelerating since the potentials $-1/2a$ and $-1/2a^2$ increase without limit from $-\infty$ to 0, as a increases from 0 to ∞, and thus corresponds to kinetic energies $1/2a$ and $1/2a^2$ which decrease without limit from $+\infty$ to 0 over the same range of a.

- **Vacuum-Dominated Universe:** In this case $\Omega_V = 1$, $\Omega_M = \Omega_R = 0$ and thus

$$V_{\rm eff}(a) = -\frac{a^2}{2} \Rightarrow \frac{1}{2H_0^2}\dot{a}^2 - \frac{a^2}{2} = 0 \Rightarrow a = \exp(H_0(t - t_0)) \ , \ H_0^2 = c^4\frac{\Lambda}{3}. \tag{3.54}$$

  In this case the Hubble constant is truely a constant for all times.

In the actual evolution of the universe the three effects are present. The addition of the vacuum energy results typically in a maximum in the potential $V_{\rm eff}(a)$ when plotted as a function of

a. Thus the universe is initially in a decelerating expansion phase consisting of a radiation-dominated and a matter-dominated regions, then it becomes vacuum dominated with an accelerating expansion. This is because beyond the maximum the potential becomes decreasing function of a and as a consequence the kinetic energy is an increasing function of a.

In a matter-dominated universe the age of the universe is given in terms of the Hubble time by the relation

$$t_0 = \frac{2}{3H_0} = \frac{2}{3}t_H. \tag{3.55}$$

This gives around 9 Gyr which is not correct since there are stars as old as 12 Gyr in our own galaxy.

The size of the universe may be given in terms of the Hubble distance d_H. A more accurate measure will be given now in terms of the conformal time η defined as follows

$$d\eta = \frac{dt}{a(t)}. \tag{3.56}$$

In the $\eta - r$ spacetime diagram, radial geodesics are the 45 degrees lines. In this diagram the big bang is the line $\eta = 0$, while our worldline may be chosen to be the line $r = 0$. At any conformal instant η only signals from points inside the past light cone can be received. To each conformal time η corresponds an instant t given through the equation

$$\eta = \int_0^t \frac{dt'}{a(t')}. \tag{3.57}$$

We have assumed that the big bang occurs at $t = 0$. Since $ds^2 = a^2(t)(-d\eta^2 + dr^2) = 0$, the largest radius $r_{\rm horizon}(t)$ from which a signal could have reached the observer at t since the big bang is given by

$$r_{\rm horizon}(t) = \eta = \int_0^t \frac{dt'}{a(t')}. \tag{3.58}$$

The 3$-$dimensional surface in spacetime with radius $r_{\rm horizon}(t)$ is called the cosmological horizon. This radius $r_{\rm horizon}(t)$ and as a consequence the cosmological horizon grow with time and thus a larger region becomes visible as time goes on. The physical distance to the horizon is obviously given by

$$d_{\rm horizon}(t) = a(t)r_{\rm horizon}(t) = a(t)\int_0^t \frac{dt'}{a(t')}. \tag{3.59}$$

The physical radius at the current epoch in a matter-dominated universe is

$$d_{\rm horizon}(t) = t_0^{2/3}\int_0^{t_0} t^{-2/3}dt = 2t_H = 8{\rm Gpc}. \tag{3.60}$$

This is different from the currently best measured age of 14 Gpc.

3.5 Closed and Open Universes

There are two more possible Friedman-Robertson-Walker universes, beside the flat case, which are isotropic and homogeneous. These are the closed universe given by a 3−sphere and the open universe given by a 3−hyperboloid. The spacetime metric in the three cases is given by

$$ds^2 = -dt^2 + a^2(t)dl^2. \tag{3.61}$$

The spatial metric in the flat case can be rewritten as (with $\chi \equiv r$)

$$dl^2 = d\chi^2 + \chi^2(d\theta^2 + \sin^2\theta d\phi^2). \tag{3.62}$$

Now we discuss the other two cases.

Closed FRW Universe: A 3−sphere can be embedded in R^4 in the usual way by

$$X^2 + Y^2 + Z^2 + W^2 = 1. \tag{3.63}$$

We introduce spherical coordinates $0 \le \theta \le \pi$, $0 \le \phi \le 2\pi$ and $0 \le \chi \le \pi$ by

$$X = \sin\chi \sin\theta \cos\phi \ , \ Y = \sin\chi \sin\theta \sin\phi \ , \ Z = \sin\chi \cos\theta \ , \ W = \cos\chi. \tag{3.64}$$

The line element on the 3−sphere is given by

$$\begin{aligned} dl^2 &= (dX^2 + dY^2 + dZ^2 + dW^2)_{S^3} \\ &= d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta d\phi^2). \end{aligned} \tag{3.65}$$

This is a closed space with finite volume and without boundary. The comoving volume is given by

$$\begin{aligned} dV &= \int \sqrt{\det g} d^4X \\ &= \int_0^{2\pi} d\phi \int_0^{\pi} d\theta \sin\theta \int_0^{\pi} d\chi \sin^2\chi \\ &= 2\pi^2. \end{aligned} \tag{3.66}$$

The physical volume is of course given by $dV(t) = a^3(t)dV$.

Open FRW Universe: A 3−hyperboloid is a 3−surface in a Minkowski spacetime M^4 analogous to a 3−sphere in R^4. It is embedded in M^4 by the relation

$$X^2 + Y^2 + Z^2 - T^2 = -1. \tag{3.67}$$

We introduce hyperbolic coordinates $0 \le \theta \le \pi$, $0 \le \phi \le 2\pi$ and $0 \le \chi \le \infty$ by

$$X = \sinh\chi \sin\theta \cos\phi \ , \ Y = \sinh\chi \sin\theta \sin\phi \ , \ Z = \sinh\chi \cos\theta \ , \ T = \cosh\chi. \tag{3.68}$$

The line element on this 3−surface is given by

$$dl^2 = (dX^2 + dY^2 + dZ^2 - dT^2)_{H^3} = d\chi^2 + \sinh^2\chi(d\theta^2 + \sin^2\theta d\phi^2). \quad (3.69)$$

This is an open space with infinite volume.

The three metrics (3.62), (3.65) and (3.69) can be rewritten collectively as

$$dl^2 = \frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2). \quad (3.70)$$

The variable r and the parameter k, called the spatial curvature, are given by

$$r = \sin\chi\ ,\ k = +1\ :\ \text{closed}. \quad (3.71)$$

$$r = \chi\ ,\ k = 0\ :\ \text{flat}. \quad (3.72)$$

$$r = \sinh\chi\ ,\ k = -1\ :\ \text{open}. \quad (3.73)$$

The metric of spacetime is thus given by

$$ds^2 = -dt^2 + a^2(t)\bigg[\frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\bigg]. \quad (3.74)$$

Thus the open and closed cases are characterized by a non-zero spatial curvature. As before, the scale factor must be given by Friedman equation derived in the next chapter. This is given by

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G\rho}{3} - \frac{kc^2}{a^2}. \quad (3.75)$$

At $t = t_0$ we get

$$H_0^2 = \frac{8\pi G\rho(t_0)}{3} - \frac{kc^2}{a^2(t_0)} \Rightarrow \rho(t_0) - \rho_c = \frac{3kc^2}{8\pi G a^2(t_0)}. \quad (3.76)$$

The critical density is of course defined by

$$\rho_c = \frac{3H_0^2}{8\pi G}. \quad (3.77)$$

Thus for a closed universe the spacetime is positively curved and as a consequence the current energy density is larger than the critical density, i.e. $\Omega = \rho(t_0)/\rho_c > 1$, whereas for an open universe the spacetime is negatively curved and as a consequence the current energy density is smaller than the critical density, i.e. $\Omega = \rho(t_0)/\rho_c < 1$. Only for a flat universe the current energy density is equal to the critical density, i.e. $\Omega=\rho(t_0)/\rho_c=1$. The above equation can also be rewritten as

$$\Omega=1+\frac{kc^2}{H_0^2a^2(t_0)}. \tag{3.78}$$

Equivalently

$$\Omega_M+\Omega_R+\Omega_V+\Omega_C=1\Rightarrow\Omega_C=1-\Omega_M-\Omega_R-\Omega_V. \tag{3.79}$$

The density parameter Ω_C associated with the spatial curvature is defined by

$$\Omega_C=-\frac{kc^2}{H_0^2a^2(t_0)}. \tag{3.80}$$

We use now the formula

$$\begin{aligned}\rho(t) &= \rho_c\Omega(t)\\ &= \rho_c\bigg(\frac{\Omega_M}{\tilde{a}^3(t)}+\frac{\Omega_R}{\tilde{a}^4(t)}+\Omega_V\bigg)\ ,\ \tilde{a}(t)=\frac{a(t)}{a(t_0)}.\end{aligned} \tag{3.81}$$

The Friedman equation can then be put in the form (with $\tilde{t}=t/t_H=H_0t$)

$$\frac{1}{2}\big(\frac{d\tilde{a}}{d\tilde{t}}\big)^2+V_{\rm eff}(\tilde{a})=\frac{\Omega_C}{2}. \tag{3.82}$$

$$V_{\rm eff}(\tilde{a})=-\frac{1}{2}\big(\frac{\Omega_M}{\tilde{a}}+\frac{\Omega_R}{\tilde{a}^2}+\tilde{a}^2\Omega_V\big). \tag{3.83}$$

We need to solve (3.79), (3.82) and (3.83). This is a generalization of the potential problem (6.48) corresponding to the flat FRW model to the generic curved FRW models. This is effectively the equation of motion of a particle moving in one dimension under the influence of the potential $V_{\rm eff}(\tilde{a})$ with energy $\Omega_C/2$. There are therefore four independent cosmological parameters Ω_M, Ω_R, Ω_V and H_0. The solution of the above equation determines the scale factor $a(t)$ as well as the present age t_0.

There are two general features worth of mention here:

- **Open and Flat:** In this case $\Omega\leq 1$ and thus $\Omega_C=1-\Omega\geq 0$. From the other hand, $V_{\rm eff}<0$. Thus $V_{\rm eff}$ is strictly below the line $\Omega_C/2$. In other words, there are no turning points where "the total energy" $\Omega_C/2$ becomes equal to the "potential energy" $V_{\rm eff}$, i.e. "the kinetic energy" $\dot{a}^2/2$ never vanishes and thus we never have $\dot{a}=0$. The universe starts from a big bang singularity at $a=0$ and keeps expanding forever.

- **Closed:** In this case $\Omega>1$ and thus $\Omega_C=1-\Omega<0$. There are here two scenarios:

  - The potential is strictly below the line $\Omega_C/2$ and thus there are no turning points. The universe starts from a big bang singularity at $a=0$ and keeps expanding forever.

- The potential intersects the line $\Omega_C/2$. There are two turning points given by the intersection points. We have two possibilities depending on where $\tilde{a}=1$ is located below the smaller turning point or above the larger turning point.
  * $\tilde{a}=1$ is below the smaller turning point. The universe starts from a big bang singularity at $a=0$, expands to a maximum radius corresponding to the smaller turning point, then recollapse to a big crunch singularity at $a=0$.
  * $\tilde{a}=1$ is above the larger turning point. The universe collapses from a larger value of a, it bounces when it hits the largest turning point and then reexpands forever. There is no singularity in this case. This case is ruled out by current observations.

For an FRW universe dominated by matter and vacuum like ours the above possibilities are sketched in the plane of the least certain cosmological parameters Ω_M and Ω_V on figure (3.10). Flat FRW models are on the line $\Omega_V=1-\Omega_M$. Open models lie below this line while closed models lie above it.

3.6 Aspects of The Early Universe

The most central property of the universe is expansion. The evidence for the expansion of the universe comes from three main sets of observations. Firstly, light from distant galaxies is shifted towards the red which can be accounted for by the expansion of the universe. Secondly, the observed abundance of light elements can be calculated from big bang nucleosynthesis. Thirdly, the cosmic microwave background radiation can be interpreted as the afterglow of a hot early universe. The temperature of the universe at any instant of time t is inversely proportional to the scale factor $a(t)$, viz

$$T\propto\frac{1}{a(t)}. \tag{3.84}$$

The early universe is obviously radiation-dominated because of the relativistic energies involved. During this era the temperature is related to time by

$$\frac{t}{1\ \mathrm{s}}=(\frac{10^{10}\ \mathrm{K}}{T})^2. \tag{3.85}$$

In particle physics accelerators we can generate temperatures up to $T=10^{15}K$ which means that we can probe the conditions of the early universe down to $10^{-10}s$. From $10^{-10}s$ to today the history of the universe is based on well understood and well tested physics. For example at $1s$ the big bang nucleosynthesis (BBN) begins where light nuclei start to form, and at 10^4 years matter-radiation equality is reached where the density of photons drops below that of matter. After matter-radiation equality, which corresponds to a scale factor of about $a=10^{-4}$, the relation between temperature and time changes to

$$t\propto\frac{1}{T^{3/2}}. \tag{3.86}$$

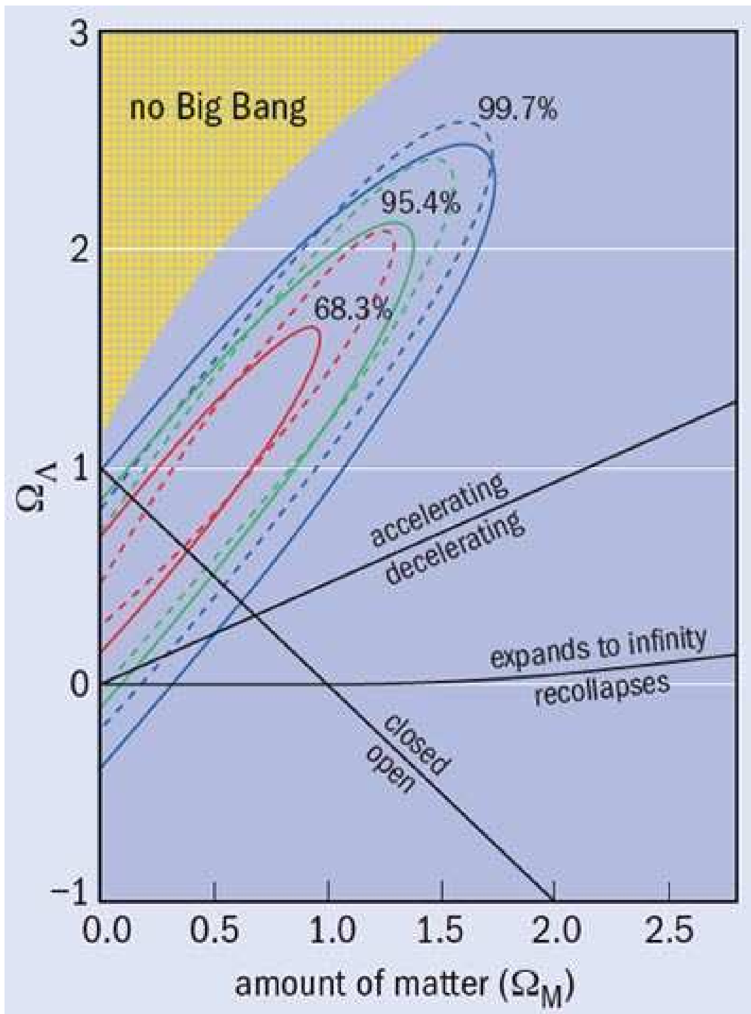

Figure 3.10: The FRW models in the $\Omega_M - \Omega_V$ plane.

The universe after the big bang was a hot and dense plasma of photons, electrons and protons which was very opaque to electromagnetic radiation. As the universe expanded it cooled down until it reached the stage where the formation of neutral hydrogen was energetically favored and the ratio of free electrons and protons to neutral hydrogen decreased to 1/10000. This event is called recombination and it occurred at around $T \simeq 0.3$ eV or equivalently 378000 years ago which corresponds to a scale factor $a = 1/1200$.

After recombination the universe becomes fully matter-dominated, and shortly after recombination, photons decouple from matter and as a consequence the mean free path of photons approaches infinity. In other words after photon decoupling the universe becomes effectively transparent. These photons are seen today as the cosmic microwave background (CMB) radiation. The decoupling period is also called the surface of last scattering.

3.7 Concordance Model

From a combination of cosmic microwave background (CMB) and large scale structure (LSS) observations we deduce that the universe is spatially flat and is composed of 4 per cent ordinary mater, 23 per cent dark matter and 73 per cent dark energy (vaccum energy or cosmological constant Λ), i.e.

$$\Omega_k \sim 0. \tag{3.87}$$

$$\Omega_M \sim 0.04 \ , \ \Omega_{DM} \sim 0.23 \ , \ \Omega_\Lambda \sim 0.73. \tag{3.88}$$

Chapter 4

Cosmology II: The Expanding Universe

4.1 Friedmann-Lemaître-Robertson-Walker Metric

The universe on very large scales is homogeneous and isotropic. This is the cosmological principle.

A spatially isotropic spacetime is a manifold in which there exists a congruence of timelike curves representing observers with tangents u^a such that given any two unit spatial tangent vectors s_1^a and s_2^a at a point p, orthogonal to u^a, there exists an isometry of the metric g_{ab} which rotates s_1^a into s_2^a while leaving p and u^a fixed. The fact that we can rotate s_1^a into s_2^a means that there is no preferred direction in space.

On the other hand, a spacetime is spatially homogeneous if there exists a foliation of spacetime, i.e. a one-parameter family of spacelike hypersurfaces Σ_t foliating spacetime such that any two points $p, q \in \Sigma_t$ can be connected by an isometry of the metric g_{ab}. The surfaces of homogeneity Σ_t are orthogonal to the isotropic observers with tangents u^a and they must be unique. In flat spacetime the isotropic observers and the surfaces of homogeneity are not unique.

A manifold can be homogeneous but not isotropic such as $R \times S^2$ or it can be isotropic around a point but not homogeneous such as the cone around its vertex. However, a spacetime which is isotropic everywhere must be also homogeneous, and a spacetime which is isotropic at a point and homogeneous must be isotropic everywhere.

The requirement of spatial isotropy and homogeneity of spacetime means that there exists a foliation of spacetime consisting of $3-$dimensional maximally symmetric spatial slices Σ_t. The universe is therefore given by the manifold $R \times \Sigma$ with metric

$$ds^2 = -c^2 dt^2 + R^2(t) d\vec{u}^2. \tag{4.1}$$

The metric on Σ is given by

$$d\sigma^2 = d\vec{u}^2 = \gamma_{ij} du^i du^j. \tag{4.2}$$

The scale factor $R(t)$ gives the volume of the spatial slice Σ at the instant of time t. The coordinates t, u^1, u^2 and u^3 are called comoving coordinates. An observer whose spatial coordinates

u^i remain fixed is a comoving observer. Obviously, the universe can look isotropic only with respect to a comoving observer. It is obvious that the relative distance between particles at fixed spatial coordinates grows with time t as $R(t)$. These particles draw worldlines in spacetime which are said to be comoving. Similarly, a comoving volume is a region of space which expands along with its boundaries defined by fixed spatial coordinates with the expansion of the universe.

A maximally symmetric metric is certainly a spherically symmetric metric. Recall that the metric $d\vec{x}^2 = dx^2 + dy^2 + dz^2$ of the flat $3-$dimensional space in spherical coordinates is $d\vec{x}^2 = dr^2 + r^2 d\Omega^2$ where $d\Omega^2 = d\theta^2 + \sin^2\theta d\phi^2$. A general $3-$dimensional metric with spherical symmetry is therefore necessarily of the form

$$d\vec{u}^2 = e^{2\beta(r)}dr^2 + r^2 d\Omega^2. \tag{4.3}$$

The Christoffel symbols are computed to be given by

$$\Gamma^r{}_{rr} = \partial_r\beta \ , \ \Gamma^r{}_{\theta\theta} = -re^{-2\beta(r)} \ , \ \Gamma^r{}_{\phi\phi} = -r\sin^2\theta e^{-2\beta(r)}, \Gamma^r{}_{r\theta} = \Gamma^r{}_{r\phi} = \Gamma^r{}_{\theta\phi} = 0. \tag{4.4}$$

$$\Gamma^\theta{}_{r\theta} = \frac{1}{r} \ , \ \Gamma^\theta{}_{\phi\phi} = -\sin\theta\cos\theta \ , \ \Gamma^\theta{}_{rr} = \Gamma^\theta{}_{r\phi} = \Gamma^\theta{}_{\theta\theta} = \Gamma^\theta{}_{\theta\phi} = 0. \tag{4.5}$$

$$\Gamma^\phi{}_{r\phi} = \frac{1}{r} \ , \ \Gamma^\phi{}_{\theta\phi} = \frac{\cos\theta}{\sin\theta} \ , \ \Gamma^\phi{}_{rr} = \Gamma^\phi{}_{r\theta} = \Gamma^\phi{}_{\theta\theta} = \Gamma^\phi{}_{\phi\phi} = 0. \tag{4.6}$$

The Ricci tensor is then given by

$$R_{rr} = \frac{2}{r}\partial_r\beta. \tag{4.7}$$

$$R_{r\theta} = 0 \ , \ R_{r\phi} = 0. \tag{4.8}$$

$$R_{\theta\theta} = 1 + e^{-2\beta}(r\partial_r\beta - 1). \tag{4.9}$$

$$R_{\theta\phi} = 0. \tag{4.10}$$

$$R_{\phi\phi} = \sin^2\theta[1 + e^{-2\beta}(r\partial_r\beta - 1)]. \tag{4.11}$$

The above spatial metric is a maximally symmetric metric. Hence, we know that the $3-$dimensional Riemann curvature tensor must be of the form

$$R^{(3)}_{ijkl} = \frac{R^{(3)}}{3(3-1)}(\gamma_{ik}\gamma_{jl} - \gamma_{il}\gamma_{jk}). \tag{4.12}$$

In other words, the Ricci tensor is actually given by

$$R^{(3)}_{ik} = (R^{(3)})_{ijk}{}^{j} = R^{(3)}_{ijkl}\gamma^{lj} = \frac{R^{(3)}}{3}\gamma_{ik}. \tag{4.13}$$

By comparison we get the two independent equations (with $k = R^{(3)}/6$)

$$2ke^{2\beta} = \frac{2}{r}\partial_r\beta. \tag{4.14}$$

$$2kr^2 = 1 + e^{-2\beta}(r\partial_r\beta - 1). \tag{4.15}$$

From the first equation we determine that the solution must be such that $\exp(-2\beta) = -kr^2 +$ constant, whereas from the second equation we determine that constant $= 1$. We get then the solution

$$\beta = -\frac{1}{2}\ln(1 - kr^2). \tag{4.16}$$

The spatial metric becomes

$$d\vec{u}^2 = \frac{dr^2}{1 - kr^2} + r^2 d\Omega^2. \tag{4.17}$$

The constant k is proportional to the scalar curvature which can be positive, negative or 0. It also obviously sets the size of the spatial slices. Without any loss of generality we can normalize it such that $k = +1, 0, -1$ since any other scale can be absorbed into the scale factor $R(t)$ which multiplies the length $|d\vec{u}|$ in the formula for ds^2.

We introduce a new radial coordinate χ by the formula

$$d\chi = \frac{dr}{\sqrt{1 - kr^2}}. \tag{4.18}$$

By integrating both sides we obtain

$$\begin{aligned} r &= \sin\chi\ ,\ k = +1 \\ r &= \chi\ ,\ k = 0 \\ r &= \sinh\chi\ ,\ k = -1. \end{aligned} \tag{4.19}$$

Thus the metric becomes

$$\begin{aligned} d\vec{u}^2 &= d\chi^2 + \sin^2\chi d\Omega^2\ ,\ k = +1 \\ d\vec{u}^2 &= d\chi^2 + \chi^2 d\Omega^2\ ,\ k = 0 \\ d\vec{u}^2 &= d\chi^2 + \sinh^2\chi d\Omega^2\ ,\ k = -1. \end{aligned} \tag{4.20}$$

The physical interpretation of this result is as follows:

- The case $k = +1$ corresponds to a constant positive curvature on the manifold Σ and is called closed. We recognize the metric $d\vec{u}^2 = d\chi^2 + \sin^2\chi d\Omega^2$ to be that of a three sphere, i.e. $\Sigma = S^3$. This is obviously a closed sphere in the same sense that the two sphere S^2 is closed.

- The case $k=0$ corresponds to 0 curvature on the manifold Σ and as such is naturally called flat. In this case the metric $d\vec{u}^2 = d\chi^2 + \chi^2 d\Omega^2$ corresponds to the flat three dimensional Euclidean space, i.e. $\Gamma = R^3$.

- The case $k=-1$ corresponds to a constant negative curvature on the manifold Σ and is called open. We recognize the metric $d\vec{u}^2 = d\chi^2 + \sinh^2\chi d\Omega^2$ to be that of a 3−dimensional hyperboloid, i.e. $\Sigma = H^3$. This is an open space.

The so-called Robertson-Walker metric on a spatially homogeneous and spatially isotropic spacetime is therefore given by

$$ds^2 = -c^2dt^2 + R^2(t)\big[\frac{dr^2}{1-kr^2} + r^2 d\Omega^2\big]. \tag{4.21}$$

4.2 Friedmann Equations

4.2.1 The First Friedmann Equation

The scale factor $R(t)$ has units of distance and thus r is actually dimensionless. We reinstate a dimensionful radius ρ by $\rho = R_0 r$. The scale factor becomes dimensionless given by $a(t) = R(t)/R_0$ whereas the curvature becomes dimensionful $\kappa = k/R_0^2$. The Robertson-Walker metric becomes

$$ds^2 = -c^2dt^2 + a^2(t)\big[\frac{d\rho^2}{1-\kappa\rho^2} + \rho^2 d\Omega^2\big]. \tag{4.22}$$

The non-zero components of the metric are $g_{00} = -1$, $g_{\rho\rho} = a^2/(1-\kappa\rho^2)$, $g_{\theta\theta} = a^2\rho^2$, $g_{\phi\phi} = a^2\rho^2\sin^2\theta$. We compute now the non-zero Christoffel symbols

$$\Gamma^0{}_{\rho\rho} = \frac{a\dot{a}}{c(1-\kappa\rho^2)} \ , \ \Gamma^0{}_{\theta\theta} = \frac{a\dot{a}\rho^2}{c} \ , \ \Gamma^0{}_{\phi\phi} = \frac{a\dot{a}\rho^2\sin^2\theta}{c}. \tag{4.23}$$

$$\Gamma^\rho{}_{0\rho} = \frac{\dot{a}}{ca} \ , \ \Gamma^\rho{}_{\rho\rho} = \frac{\kappa\rho}{1-\kappa\rho^2} \ , \ \Gamma^\rho{}_{\theta\theta} = -\rho(1-\kappa\rho^2) \ , \ \Gamma^\rho{}_{\phi\phi} = -\rho(1-\kappa\rho^2)\sin^2\theta. \tag{4.24}$$

$$\Gamma^\theta{}_{0\theta} = \frac{\dot{a}}{ca} \ , \ \Gamma^\theta{}_{\rho\theta} = \frac{1}{\rho} \ , \ \Gamma^\theta{}_{\phi\phi} = -\sin\theta\cos\theta. \tag{4.25}$$

$$\Gamma^\phi{}_{0\phi} = \frac{\dot{a}}{ca} \ , \ \Gamma^\phi{}_{\rho\phi} = \frac{1}{\rho} \ , \ \Gamma^\phi{}_{\theta\phi} = \frac{\cos\theta}{\sin\theta}. \tag{4.26}$$

The non-zero components of the Ricci tensor are

$$R_{00} = -\frac{3}{c^2}\frac{\ddot{a}}{a}. \tag{4.27}$$

$$R_{\rho\rho} = \frac{1}{c^2(1-\kappa\rho^2)}(a\ddot{a}+2\dot{a}^2+2\kappa c^2). \tag{4.28}$$

$$R_{\theta\theta} = \frac{\rho^2}{c^2}(a\ddot{a}+2\dot{a}^2+2\kappa c^2)\ ,\ R_{\phi\phi} = \frac{\rho^2\sin^2\theta}{c^2}(a\ddot{a}+2\dot{a}^2+2\kappa c^2). \tag{4.29}$$

The scalar curvature is therefore given by

$$R = g^{\mu\nu}R_{\mu\nu} = \frac{6}{c^2}\bigg(\frac{\ddot{a}}{a}+\bigg(\frac{\dot{a}}{a}\bigg)^2+\frac{\kappa c^2}{a^2}\bigg). \tag{4.30}$$

The Einstein's equations are

$$R_{\mu\nu} = \frac{8\pi G}{c^4}(T_{\mu\nu}-\frac{1}{2}g_{\mu\nu}T). \tag{4.31}$$

The stress-energy-momentum tensor

$$T^{\mu\nu} = (\rho+\frac{P}{c^2})U^\mu U^\nu + Pg^{\mu\nu}. \tag{4.32}$$

The fluid is obviously at rest in comoving coordinates. In other words, $U^\mu = (c,0,0,0)$ and hence

$$T^{\mu\nu} = \mathrm{diag}(\rho c^2, Pg^{11}, Pg^{22}, Pg^{33}) \Rightarrow T_\mu{}^\lambda = \mathrm{diag}(-\rho c^2, P, P, P). \tag{4.33}$$

Thus $T = T_\mu{}^\mu = -\rho c^2 + 3P$. The $\mu=0$, $\nu=0$ component of Einstein's equations is

$$R_{00} = \frac{8\pi G}{c^4}(T_{00}+\frac{1}{2}T) \Rightarrow -3\frac{\ddot{a}}{a} = 4\pi G(\rho+3\frac{P}{c^2}). \tag{4.34}$$

The $\mu=\rho$, $\nu=\rho$ component of Einstein's equations is

$$R_{\rho\rho} = \frac{8\pi G}{c^4}(T_{\rho\rho}-\frac{1}{2}g_{\rho\rho}T) \Rightarrow a\ddot{a}+2\dot{a}^2+2\kappa c^2 = 4\pi G(\rho-\frac{P}{c^2})a^2. \tag{4.35}$$

There are no other independent equations. The Einstein's equation (4.34) is known as the second Friedmann equation. This is given by

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho+3\frac{P}{c^2}). \tag{4.36}$$

Using this result in the Einstein's equation (4.35) yields immediately the first Friedmann equation. This is given by

$$\frac{\dot{a}^2}{a^2} = \frac{8\pi G\rho}{3} - \frac{\kappa c^2}{a^2}. \tag{4.37}$$

In most cases, in which we know how ρ depends on a, the first Friedmann equation is sufficient to solve for the problem.

4.2.2 Cosmological Parameters

We introduce the following cosmological parameters:

- **The Hubble parameter H:** This is given by

$$H = \frac{\dot{a}}{a}. \tag{4.38}$$

This provides the rate of expansion. At present time we have

$$H_0 = 100h \text{ km sec}^{-1} \text{ Mpc}^{-1}. \tag{4.39}$$

The dimensionless Hubble parameter h is around 0.7 ± 0.1. The megaparsec Mpc is 3.09×10^{24}cm.

- **The density parameter Ω and the critical density ρ_c:** These are defined by

$$\Omega = \frac{8\pi G}{3H^2}\rho = \frac{\rho}{\rho_c}. \tag{4.40}$$

$$\rho_c = \frac{3H^2}{8\pi G}. \tag{4.41}$$

- **The deceleration parameter q:** This provides the rate of change of the rate of the expansion of the universe. This is defined by

$$q = -\frac{a\ddot{a}}{\dot{a}^2}. \tag{4.42}$$

Using the first two parameters in the first Friedmann equation we obtain

$$\frac{\rho - \rho_c}{\rho_c} = \Omega - 1 = \frac{\kappa c^2}{H^2 a^2}. \tag{4.43}$$

We get immediately the behavior

$$\textbf{closed universe} \ : \ \kappa > 0 \leftrightarrow \Omega > 1 \leftrightarrow \rho > \rho_c. \tag{4.44}$$

$$\textbf{flat universe} \ : \ \kappa = 0 \leftrightarrow \Omega = 1 \leftrightarrow \rho = \rho_c. \tag{4.45}$$

$$\textbf{open universe} \ : \ \kappa < 0 \leftrightarrow \Omega < 1 \leftrightarrow \rho < \rho_c. \tag{4.46}$$

4.2.3 Energy Conservation

Let us now consider the conservation law $\nabla_\mu T^\mu{}_\nu = \partial_\mu T^\mu{}_\nu + \Gamma^\mu{}_{\mu\alpha} T^\alpha{}_\nu - \Gamma^\alpha{}_{\mu\nu} T^\mu{}_\alpha = 0$. In the comoving coordinates we have $T_\mu{}^\nu = \mathrm{diag}(-\rho c^2, P, P, P)$. The $\nu = 0$ component of the conservation law is

$$-c\dot{\rho} - \frac{3\dot{a}}{ca}(\rho c^2 + P) = 0. \tag{4.47}$$

In cosmology the pressure P and the rest mass density ρ are generally related by the equation of state

$$P = w\rho c^2. \tag{4.48}$$

The conservation of energy becomes

$$\frac{\dot{\rho}}{\rho} = -3(1+w)\frac{\dot{a}}{a}. \tag{4.49}$$

For constant w the solution is of the form

$$\rho \propto a^{-3(1+w)}. \tag{4.50}$$

There are three cases of interest:

- **The matter-dominated universe:** Matter (also called dust) is a set of collision-less non-relativistic particles which have zero pressure. For example, stars and galaxies may be considered as dust since pressure can be neglected to a very good accuracy. Since $P_M = 0$ we have $w = 0$ and as a consequence

  $$\rho_M \propto a^{-3}. \tag{4.51}$$

  This can be seen also as follows. The energy density for dust comes entirely from the rest mass of the particles. The mass density is $\rho = nm$ where n is the number density which is inversely proportional to the volume. Hence, the mass density must go as the inverse of a^3 which is the physical volume of a comoving region.

- **The radiation-dominated universe:** Radiation consists of photons (obviously) but also includes any particles with speeds close to the speed of light. For an electromagnetic field we can show that the stress-energy-tensor satisfies $T^\mu{}_\mu = 0$. However, the stress-energy-momentum tensor of a perfect fluid satisfies $T_\mu{}^\mu = -\rho c^2 + 3P$. Thus for radiation we must have the equation of state $P_R = \rho_R c^2/3$ and as a consequence $w = 1/3$ and hence

  $$\rho_R \propto a^{-4}. \tag{4.52}$$

  In this case the energy of each photon will redshifts away as $1/a$ (see below) as the universe expands which is the extra factor that multiplies the original factor $1/a^3$ coming from number density.

- **The vacuum-dominated universe:** The vacuum energy is a perfect fluid with equation of state $P_V = -\rho_V$, i.e. $w = -1$ and hence

$$\rho_V \propto a^0. \tag{4.53}$$

The vacuum energy is an unchanging form of energy in any physical volume which does not redshifts.

The null dominant energy condition allows for densities which satisfy the requirements $\rho \geq 0$, $\rho \geq |P|/c^2$ or $\rho \leq 0$, $P = -c^2\rho < 0$, thus in particular allowing the vacuum energy to be either positive or negative, and as a consequence we must have in all the above discussed cases $|w| \leq 1$.

In general matter, radiation and vacuum can contribute simultaneously to the evolution of the universe. Let us simply assume that all densities evolve as power laws, viz

$$\rho_i = \rho_{i0}a^{-n_i} \Leftrightarrow w_i = \frac{n_i}{3} - 1. \tag{4.54}$$

The first Friedmann equation can then be put in the form

$$\begin{aligned} H^2 &= \frac{8\pi G}{3}\sum_i \rho_i - \frac{\kappa c^2}{a^2} \\ &= \frac{8\pi G}{3}\sum_{i,C} \rho_i. \end{aligned} \tag{4.55}$$

In the above equation the spatial curvature is thought of as giving another contribution to the rest mass density given by

$$\rho_C = -\frac{3}{8\pi G}\frac{\kappa c^2}{a^2}. \tag{4.56}$$

This rest mass density corresponds to the values $w_C = -1/3$ and $n_C = 2$. The total density parameter Ω is defined by $\Omega = \sum_i 8\pi G\rho_i/3H^2$. By analogy the density parameter of the spatial curvature is given by

$$\Omega_C = \frac{8\pi G\rho_C}{3H^2} = -\frac{\kappa c^2}{H^2a^2}. \tag{4.57}$$

The first Friedmann equation becomes the identity

$$\sum_{i,C}\Omega_i = 1 \Leftrightarrow \Omega_C = 1 - \Omega = 1 - \Omega_M - \Omega_R - \Omega_V. \tag{4.58}$$

The rest mass densities of matter and radiation are always positive whereas the rest mass densities corresponding to vacuum and curvature can be either positive or negative.

The Hubble parameter is the rate of expansion of the universe. The derivative of the Hubble parameter is

$$\begin{aligned}
\dot{H} &= \frac{\ddot{a}}{a} - \left(\frac{\dot{a}}{a}\right)^2 \\
&= -\frac{4\pi G}{3}\sum_i \rho_i(1+3w_i) - \frac{8\pi G}{3}\sum_i \rho_i + \frac{\kappa c^2}{a^2} \\
&= -4\pi G\sum_i \rho_i(1+w_i) + \frac{\kappa c^2}{a^2} \\
&= -4\pi G\sum_{i,C} \rho_i(1+w_i). \qquad (4.59)
\end{aligned}$$

This is effectively the second Friedmann equation. In terms of the deceleration parameter this reads

$$\frac{\dot{H}}{H^2} = -1-q. \qquad (4.60)$$

An open or flat universe $\rho_C \geq 0$ ($\kappa \leq 0$) with $\rho_i > 0$ will never contract as long as $\sum_{i,C}\rho_i \neq 0$ since $H^2 \propto \sum_{i,C}\rho_i$ from the first Friedmann equation (4.55). On the other hand, we have $|w_i| \leq 1$, and thus we deduce from the second Friedmann equation (4.59) the condition $\dot{H} \leq 0$ which indicates that the expansion of the universe decelerates.

For a flat universe dominated by a single component w_i we can show that the deceleration parameter is given by

$$q_i = \frac{1}{2}(1+3w_i). \qquad (4.61)$$

This is positive and thus the expansion is accelerating for a matter dominated universe ($w_i = 0$) whereas it is negative and thus the expansion is decelerating for a vacuum dominated universe ($w_i = -1$). The current cosmological data strongly favors the second possibility.

4.3 Examples of Scale Factors

Matter and Radiation Dominated Universes: From observation we know that the universe was radiation-dominated at early times whereas it is matter-dominated at the current epoch. Let us then consider a single kind of rest mass density $\rho \propto a^{-n}$. The Friedmann equation gives therefore $\dot{a} \propto a^{1-n/2}$. The solution behaves as

$$a \sim t^{\frac{2}{n}}. \qquad (4.62)$$

For a flat (since $\rho_C = 0$) universe dominated by matter we have $\Omega = \Omega_M = 1$ and $n = 3$. In this case

$$a \sim t^{\frac{2}{3}}\ ,\ \text{Matter} - \text{Dominated Universe}. \qquad (4.63)$$

This is also known as the Einstein-de Sitter universe. For a flat universe dominated by radiation we have $\Omega=\Omega_R=1$ and $n=4$ and hence $a\sim t^{1/2}$.

$$a\sim t^{\frac{1}{2}}\ ,\ \text{Radiation}-\text{Dominated Universe}. \tag{4.64}$$

These solutions exhibit a singularity at $a=0$ known as the big bang. Indeed the rest mass density diverges as $a\longrightarrow 0$. At this regime general relativity breaks down and quantum gravity takes over. The so-called cosmological singularity theorems show that any universe with $\rho>0$ and $p\geq 0$ must start at a singularity.

Vacuum Dominated Universes: For a flat universe dominated by vacuum energy we have $H=\text{constant}$ since $\rho_\Lambda=\text{constant}$ and hence $a=\exp(Ht)$. The universe expands exponentially. The metric reads explicitly $ds^2=-c^2dt^2+\exp(Ht)(dx^2+dy^2+dz^2)$. This is the maximally symmetric spacetime known as de Sitter spacetime. Indeed, the corresponding Riemann curvature tensor has the characteristic form of a maximally symmetric spacetime in $4-$dimension. Since de Sitter spacetime has a positive scalar curvature whereas this space has zero curvature the coordinates (t,x,y,z) must only cover part of the de Sitter spacetime. Indeed, we can show that these coordinates are incomplete in the past.

From observation $\Omega_R<<\Omega_{M,V,C}$. We will therefore neglect the effect of radiation and set $\Omega=\Omega_M+\Omega_V$. The curvature is $\Omega_C=1-\Omega_M-\Omega_V$. Recall that $\Omega_C\propto 1/a^2$, $\Omega_M\propto 1/a^3$ and $\Omega_V\propto 1/a^0$. Thus in the limit $a\longrightarrow 0$ (the past), matter dominates and spacetime approaches Einstein-de Sitter spacetime whereas in the limt $a\longrightarrow\infty$ (the future), vacuum dominates and spacetime approaches de Sitter spacetime.

Milne Universe: For an empty space with spatial curvature we have

$$H^2=-\frac{\kappa c^2}{a^2}. \tag{4.65}$$

The curvature must be negative. This corresponds to the so-called Milne universe with a rest mass density $\rho_C\propto a^{-2}$, i.e. $n=2$. Hence the Milne universe expands linearly, viz

$$a\sim t\ ,\ \text{Milne universe}. \tag{4.66}$$

The Milne universe can only be Minkowski spacetime in a certain incomplete coordinate system which can be checked by showing that its Riemann curvature tensor is actually 0. In fact Milne universe is the interior of the future light cone of some fixed event in spacetime foliated by hyperboloids which have negative scalar curvature.

The Static Universe: A static universe satisfies $\dot{a}=\ddot{a}=0$. The Friedmann equations become

$$\frac{\kappa c^2}{a^2}=\frac{8\pi G}{3}\sum_i\rho_i\ ,\ \sum_i(\rho_i+3\frac{P}{c^2})=0. \tag{4.67}$$

Again by neglecting radiation the second equation leads to

$$\rho_M+\rho_V=-\frac{3}{c^2}(P_M+P_V)=3\rho_V\Rightarrow\rho_M=2\rho_V. \tag{4.68}$$

The first equation gives the scalar curvature

$$\kappa=\frac{4\pi G\rho_M a^2}{c^2}. \tag{4.69}$$

Expansion versus Recollapse: Recall that $H=\dot{a}/a$. Thus if $H>0$ the universe is expanding while if $H<0$ the universe is collapsing. The point a_* at which the universe goes from expansion to collapse corresponds to $H=0$. By using the Friedmann equation this gives the condition

$$\rho_{M0}a_*^{-3}+\rho_{V0}+\rho_{C0}a_*^{-2}=0. \tag{4.70}$$

Recall also that $\Omega_{C0}=1-\Omega_{M0}-\Omega_{V0}$ and $\Omega_i\propto\rho_i/H^2$. By dividing the above equation on H_0^2 we get

$$\Omega_{M0}a_*^{-3}+\Omega_{V0}+(1-\Omega_{M0}-\Omega_{V0})a_*^{-2}=0\Rightarrow\Omega_{V0}a_*^3+(1-\Omega_{M0}-\Omega_{V0})a_*+\Omega_{M0}=0. \tag{4.71}$$

First we consider $\Omega_{V0}=0$. For open and flat universes we have $\Omega_0=\Omega_{M0}\leq 1$ and thus the above equation has no solution. In other words, open and flat universes keep expanding. For a closed universe $\Omega_0=\Omega_{M0}>1$ and the above equation admits a solution a_* and as a consequence the closed universe will recollapse.

For $\Omega_{V0}>0$ the situation is more complicated. For $0\leq\Omega_{M0}\leq 1$ the universe will always expand whereas for $\Omega_{M0}>1$ the universe will always expand only if Ω_Λ is further bounded from below as

$$\Omega_{V0}\geq\hat{\Omega}_{V0}=4\Omega_{M0}\cos^3\left[\frac{1}{3}\cos^{-1}\left(\frac{1-\Omega_{M0}}{\Omega_{M0}}\right)+\frac{4\pi}{3}\right]. \tag{4.72}$$

This means in particular that the universe with sufficiently large Ω_{M0} can recollapse for $0<\Omega_{V0}<\hat{\Omega}_{V0}$. Thus a sufficiently large Ω_M can halt the expansion before Ω_V becomes dominant.

Note also from the above solution that the universe will always recollapse for $\Omega_{V0}<0$. Indeed, the effect of $\Omega_{V0}<0$ is to cause deceleration and recollapse.

4.4 Redshift, Distances and Age

4.4.1 Redshift in a Flat Universe

Let us consider the metric

$$ds^2=-c^2dt^2+a^2(t)[dx^2+dy^2+dz^2]. \tag{4.73}$$

Thus space at each fixed instant of time t is the Euclidean $3-$dimensional space R^3. The universe described by this metric is expanding in the sense that the volume of the $3-$dimensional spatial slice, which is given by the so-called scale factor $a(t)$, is a function of time. The above metric is also rewritten as

$$g_{00} = -1 \ , \ g_{ij} = a^2(t)\delta_{ij} \ , \ g_{0i} = 0. \tag{4.74}$$

It is obvious that the relative distance between particles at fixed spatial coordinates grows with time t as $a(t)$. These particles draw worldlines in spacetime which are said to be comoving. Similarly a comoving volume is a region of space which expands along with its boundaries defined by fixed spatial coordinates with the expansion of the universe.

We recall the formula of the Christoffel symbols

$$\Gamma^{\lambda}{}_{\mu\nu} = \frac{1}{2}g^{\lambda\rho}(\partial_\mu g_{\nu\rho} + \partial_\nu g_{\mu\rho} - \partial_\rho g_{\mu\nu}). \tag{4.75}$$

We compute

$$\Gamma^{0}{}_{\mu\nu} = -\frac{1}{2}(\partial_\mu g_{\nu 0} + \partial_\nu g_{\mu 0} - \partial_0 g_{\mu\nu}) \Rightarrow \Gamma^0{}_{00} = \Gamma^0{}_{0i} = \Gamma^0{}_{i0} = 0 \ , \ \Gamma^0{}_{ij} = \frac{a\dot{a}}{c}\delta_{ij}. \tag{4.76}$$

$$\Gamma^{i}{}_{\mu\nu} = \frac{1}{2a^2}(\partial_\mu g_{\nu i} + \partial_\nu g_{\mu i} - \partial_i g_{\mu\nu}) \Rightarrow \Gamma^i{}_{00} = \Gamma^i{}_{jk} = 0 \ , \ \Gamma^i{}_{0j} = \frac{\dot{a}}{ac}\delta_{ij}. \tag{4.77}$$

The geodesic equation reads

$$\frac{d^2x^\lambda}{d\tau^2} + \Gamma^{\lambda}{}_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau} = 0. \tag{4.78}$$

In particular

$$\frac{d^2x^0}{d\lambda^2} + \Gamma^{0}{}_{ij}\frac{dx^i}{d\lambda}\frac{dx^j}{d\lambda} = 0 \Rightarrow \frac{d^2t}{d\lambda^2} + \frac{a\dot{a}}{c^2}\left(\frac{d\vec{x}}{d\lambda}\right)^2 = 0. \tag{4.79}$$

For null geodesics (which are paths followed by massless particle such as photons) we have $ds^2 = -c^2dt^2 + a^2(t)d\vec{x}^2 = 0$. In other words we must have along a null geodesic with parameter λ the condition $a^2(t)d\vec{x}^2/d\lambda^2 = c^2dt^2/d\lambda^2$. We get then the equation

$$\frac{d^2t}{d\lambda^2} + \frac{\dot{a}}{a}\left(\frac{dt}{d\lambda}\right)^2 = 0. \tag{4.80}$$

The solution is immediately given by (with ω_0 a constant)

$$\frac{dt}{d\lambda} = \frac{\omega_0}{c^2 a}. \tag{4.81}$$

The energy of the photon as measured by an observer whose velocity is U^μ is given by $E = -p^\mu U_\mu$ where p^μ is the $4-$vector energy-momentum of the photon. A comoving observer is an observer with fixed spatial coordinates and thus $U^\mu=(U^0,0,0,0)$. Since $g_{\mu\nu}U^\mu U^\nu=-c^2$ we must have $U^0=\sqrt{-c^2/g_{00}}=c$. Furthermore we choose the parameter λ along the null geodesic such that the 4−vector energy-momentum of the photon is $p^\mu=dx^\mu/d\lambda$. We get then

$$\begin{aligned}E &= -g_{\mu\nu}p^\mu U^\nu\\ &= p^0U^0\\ &= \frac{dx^0}{d\lambda}c\\ &= \frac{\omega_0}{a}.\end{aligned} \tag{4.82}$$

Thus if a photon is emitted with energy E_1 at a scale factor a_1 and then observed with energy E_2 at a scale factor a_2 we must have

$$\frac{E_1}{E_2}=\frac{a_2}{a_1}. \tag{4.83}$$

In terms of wavelengths this reads

$$\frac{\lambda_2}{\lambda_1}=\frac{a_2}{a_1}. \tag{4.84}$$

This is the phenomena of cosmological redshift: In an expanding universe we have $a_2>a_1$ and as a consequence we must have $\lambda_2>\lambda_1$, i.e. the wavelength of the photon grows with time. The amount of redshift is

$$z=\frac{E_1-E_2}{E_2}=\frac{a_2}{a_1}-1. \tag{4.85}$$

This effect allows us to measure the change in the scale factor between distant galaxies (where the photons are emitted) and here (where the photons are observed). Also it can be used to infer the distance between us and distant galaxies. Indeed a greater redshift means a greater distance. For example z close to 0 means that there was not sufficient time for the universe to expand because the emitter and observer are very close to each other.

The scale factor $a(t)$ as a function of time might be of the form

$$a(t)=t^q\ ,\ 0<q<1. \tag{4.86}$$

In the limit $t\longrightarrow 0$ we have $a(t)\longrightarrow 0$. In fact the time $t=0$ is a true singularity of this geometry, which represents a big bang event, and hence it must be excluded. The physical range of t is

$$0<t<\infty. \tag{4.87}$$

The light cones of this curved spacetime are defined by the null paths $ds^2=-c^2dt^2+a^2(t)d\vec{x}^2=0$. In $1+1$ dimensions this reads

$$ds^2=-c^2dt^2+a^2(t)dx^2=0\Rightarrow\frac{dx}{dt}=\pm ct^{-q}. \tag{4.88}$$

The solution is

$$t=\Big(\pm\frac{1-q}{c}(x-x_0)\Big)^{\frac{1}{1-q}}. \tag{4.89}$$

These are the light cones of our expanding universe. Compare with the light cones of a flat Minkowski universe obtained by setting $q=0$ in this formula. These light cones are tangent to the singularity at $t=0$. As a consequence the light cones in this curved geometry of any two points do not necessarily need to intersect in the past as opposed to the flat Minkowski universe where the light cones of any two points intersect both in the past and in the future.

4.4.2 Cosmological Redshift

Recall that a Killing vector is any vector which satisfies the Killing equation $\nabla_\mu K_\nu+\nabla_\nu K_\mu=0$. This Killing vector generates an isometry of the metric which is associated with the conservation of the momentum $p_\mu K^\mu$ along the geodesic whose tangent vector is p^μ.

In an FLRW universe there could be no Killing vector along timelike geodesic and thus no concept of conserved energy. However we can define Killing tensor along timelike geodesic. We introduce the tensor

$$K_{\mu\nu}=a^2(t)(g_{\mu\nu}+\frac{U_\mu U_\nu}{c^2}). \tag{4.90}$$

We have

$$\begin{aligned}\nabla_{(\sigma}K_{\mu\nu)} &= \nabla_\sigma K_{\mu\nu}+\nabla_\mu K_{\sigma\nu}+\nabla_\nu K_{\mu\sigma}\\ &= \partial_\sigma K_{\mu\nu}+\partial_\mu K_{\sigma\nu}+\partial_\nu K_{\mu\sigma}-2\Gamma^\rho{}_{\sigma\mu}K_{\rho\nu}-2\Gamma^\rho{}_{\sigma\nu}K_{\mu\rho}-2\Gamma^\rho{}_{\mu\nu}K_{\sigma\rho}.\end{aligned} \tag{4.91}$$

Since U^μ is the 4−vector velocity of comoving observers in the FLRW universe we have $U^\mu=(c,0,0,0)$ and $U_\mu=(-c,0,0,0)$ and as a consequence $K_{\mu\nu}=a^4\mathrm{diag}(0,1/(1-\kappa\rho^2),\rho^2,\rho^2\sin^2\theta)$. In other words $K_{ij}=a^2g_{ij}=a^4\gamma_{ij}$, $K_{0i}=K_{00}=0$. The first set of non trivial components of $\nabla_{(\sigma}K_{\mu\nu)}$ are

$$\begin{aligned}\nabla_{(0}K_{ij)}=\nabla_{(i}K_{0j)}=\nabla_{(j}K_{i0)} &= \nabla_0K_{ij}\\ &= \partial_0K_{ij}-2\Gamma^k{}_{0i}K_{kj}-2\Gamma^k{}_{0j}K_{ik}.\end{aligned} \tag{4.92}$$

By using the result $\Gamma^k{}_{0i}=\dot{a}\delta_i^k/ca$ we get

$$\begin{aligned}\nabla_{(0}K_{ij)} &= \partial_0K_{ij}-4\frac{\dot{a}a}{c}g_{ij}\\ &= \partial_0K_{ij}-\frac{d}{dx^0}a^4.\gamma_{ij}\\ &= 0.\end{aligned} \tag{4.93}$$

The other set of non trivial components of $\nabla_{(\sigma}K_{\mu\nu)}$ are

$$\begin{aligned}\nabla_{(i}K_{jk)} &= \nabla_i K_{jk} + \nabla_j K_{ik} + \nabla_k K_{ij}\\ &= a^4(\nabla_i\gamma_{jk} + \nabla_j\gamma_{ik} + \nabla_k\gamma_{ij})\\ &= 0.\end{aligned} \tag{4.94}$$

In the last step we have used the fact that the 3−dimensional metric γ_{ij} is covariantly constant which can be verified directly.

We conclude therefore that the tensor $K_{\mu\nu}$ is a Killing tensor and hence $K^2 = K_{\mu\nu}V^\mu V^\nu$ where $V^\mu = dx^\mu/d\tau$ is the 4−vector velocity of a particle is conserved along its geodesic. We have two cases to consider:

- Massive Particles: In this case $V^\mu V_\mu = -c^2$ and thus $(V^0)^2 = c^2 + g_{ij}V^iV^j = c^2 + \vec{V}^2$. But since $U_\mu V^\mu = -cV^0$ we have

$$\begin{aligned}K^2 &= K_{\mu\nu}V^\mu V^\nu\\ &= a^2(V^\mu V_\mu + \frac{(U_\mu V^\mu)^2}{c^2})\\ &= a^2\vec{V}^2.\end{aligned} \tag{4.95}$$

  We get then the result

$$|\vec{V}| = \frac{K}{a}. \tag{4.96}$$

  In other worlds particles slow down with respect to comoving coordinates as the universe expands. This is equivalent to the statement that the universe cools down as it expands.

- Massless Particles: In this case $V^\mu V_\mu = 0$ and hence

$$\begin{aligned}K^2 &= K_{\mu\nu}V^\mu V^\nu\\ &= a^2(V^\mu V_\mu + \frac{(U_\mu V^\mu)^2}{c^2})\\ &= \frac{a^2}{c^2}(U_\mu V^\mu)^2.\end{aligned} \tag{4.97}$$

  We get now the result

$$|U_\mu V^\mu| = \frac{cK}{a}. \tag{4.98}$$

  However recall that the energy E of the photon as measured with respect to the comoving observer whose 4−vector velocity is U^μ is given by $E = -p^\mu U_\mu$. But the 4−vector energy-momentum of the photon is given by $p^\mu = V^\mu$. Hence we obtain

$$E = \frac{cK}{a}. \tag{4.99}$$

An emitted photon with energy $E_{\rm em}$ will be observed with a lower energy $E_{\rm ob}$ as the universe expands, viz

$$\frac{E_{\rm em}}{E_{\rm ob}}=\frac{a_{\rm ob}}{a_{\rm em}}>1. \tag{4.100}$$

We define the redshift

$$z_{\rm em}=\frac{E_{\rm em}-E_{\rm ob}}{E_{\rm ob}}. \tag{4.101}$$

This means that

$$a_{\rm em}=\frac{a_{\rm ob}}{1+z_{\rm em}}. \tag{4.102}$$

Recall that $a(t)=R(t)/R_0$. Thus if we are observing the photon today we must have $a_{\rm ob}(t)=1$ or equivalently $R_{\rm ob}(t)=R_0$. We get then

$$a_{\rm em}=\frac{1}{1+z_{\rm em}}. \tag{4.103}$$

The redshift is a direct measure of the scale factor at the time of emission.

4.4.3 Comoving and Instantaneous Physical Distances

Note that the above described redshift is due to the expansion of the universe and not to the relative velocity between the observer and emitter and thus it is not the same as the Doppler effect. However over distances which are much smaller than the Hubble radius $1/H_0$ and the radius of spatial curvature $1/\sqrt{\kappa}$ we can view the expansion of the universe as galaxies moving apart and as a consequence the redshift can be thought of as a Doppler effect. The redshift can therefore be thought of as a relative velocity. We stress that this picture is only an approximation which is valid at sufficiently small distances.

The distance d from us to a given galaxy can be taken to be the instantaneous physical distance d_p. Recall the metric of the FLRW universe given by

$$ds^2=-c^2dt^2+R_0^2a^2(t)(d\chi^2+S_k^2(\chi)d\Omega^2). \tag{4.104}$$

$$\begin{aligned}
S_k(\chi)&=\sin\chi\ ,\ k=+1\\
S_k(\chi)&=\chi\ ,\ k=0\\
S_k(\chi)&=\sinh\chi\ ,\ k=-1.
\end{aligned} \tag{4.105}$$

Clearly the instantaneous physical distance d_p from us at $\chi=0$ to a galaxy which lies on a sphere centered on us of radius χ is

$$d_p=R_0a(t)\chi. \tag{4.106}$$

The radial coordinate χ is constant since we are assuming that us and the galaxy are perfectly comoving. The relative velocity (which we can define only within the approximation that the redshift is a Doppler effect) is therefore

$$v = \dot{d}_p = R_0 \dot{a} \chi = \frac{\dot{a}}{a} d_p = H d_p. \tag{4.107}$$

At present time this law reads

$$v = H_0 d_p. \tag{4.108}$$

This is the famous Lemaître-Hubble law: Galaxies which are not very far from us move away from us with a recess velocity which is linearly proportional to their distance.

The instantaneous physical distance d_p is obviously not a measurable quantity since measurement relates to events on our past light cone whereas d_p relates events on our current spatial hyper surface.

4.4.4 Luminosity Distance

The luminosity distance is the distance inferred from comparing the proper luminosity to the observed brightness if we were in flat and non expanding universe. Recall that the luminosity L of a source is the amount of energy emitted per unit time. This is the proper or intrinsic luminosity of the source. We will assume that the source radiates equally in all directions. In flat space the flux of the source as measured by an observer a distance d away is the amount of energy per unit time per unit area given by $F = L/4\pi d^2$. This is the apparent brightness at the location of the observer. We write this result as

$$\frac{F}{L} = \frac{1}{4\pi d^2}. \tag{4.109}$$

Now in the FLRW universe the flux will be diluted by two effects. First the energy of each photon will be redshift by the factor $1/a = 1 + z$ due to the expansion of the universe. In other words the luminosity L must be changed as $L \longrightarrow (1 + z)L$. In a comoving system light will travel a distance $|d\vec{u}| = cdt/(R_0 a)$ during a time dt. Hence two photons emitted a time δt apart will be observed a time $(1 + z)\delta t$ apart. The flux F must therefore be changed as $F \longrightarrow F/(1 + z)$. Hence in the FLRW universe we must have

$$\frac{F}{L} = \frac{1}{(1 + z)^2 A}. \tag{4.110}$$

The area A of a sphere of radius χ in the comoving system of coordinates is from the FLRW metric

$$A = 4\pi R_0^2 a^2(t) S_k^2(\chi) = 4\pi R_0^2 S_k^2(\chi). \tag{4.111}$$

Again we used the fact that at the current epoch $a(t)=1$. The luminosity distance d_L is the analogue of d and thus it must be defined by

$$d_L^2=\frac{L}{4\pi F}\Rightarrow d_L=(1+z)R_0S_k(\chi). \tag{4.112}$$

Next on a null radial geodesic we have $-c^2dt^2+a^2(t)R_0^2d\chi^2=0$ and thus we obtain (by using $dt=da/(aH)$ and remembering that at the emitter position $\chi^{'}=0$ and $a=a(t)$ whereas at our position $\chi^{'}=\chi$ and $a=1$)

$$\int_0^{\chi}d\chi^{'}=\frac{c}{R_0}\int_{t_a}^{t_1}\frac{dt^{'}}{a(t^{'})}=\frac{c}{R_0}\int_a^1\frac{da^{'}}{a^{'2}H(a^{'})}. \tag{4.113}$$

We convert to redshift by the formula $a=1/(1+z^{'})$. We get

$$\chi=\frac{c}{R_0}\int_0^z\frac{dz^{'}}{H(z^{'})}. \tag{4.114}$$

The Friedmann equation is

$$\begin{aligned}H^2&=\frac{8\pi G}{3}\sum_{i,c}\rho_i\\&=\frac{8\pi G}{3}\sum_{i,c}\rho_{i0}a^{-n_i}\\&=\frac{8\pi G}{3}\sum_{i,c}\rho_{i0}(1+z^{'})^{n_i}\\&=H_0^2\sum_{i,c}\Omega_{i0}(1+z^{'})^{n_i}\\&=H_0^2E^2(z^{'}).\end{aligned} \tag{4.115}$$

Thus

$$H(z^{'})=H_0E(z^{'})\ ,\ E(z^{'})=\bigg[\sum_{i,c}\Omega_{i0}(1+z^{'})^{n_i}\bigg]^{\frac{1}{2}}. \tag{4.116}$$

Hence

$$\chi=\frac{c}{R_0H_0}\int_0^z\frac{dz^{'}}{E(z^{'})}. \tag{4.117}$$

The luminosity distance becomes

$$d_L=(1+z)R_0S_k\bigg(\frac{c}{R_0H_0}\int_0^z\frac{dz^{'}}{E(z^{'})}\bigg). \tag{4.118}$$

Recall that the curvature density is $\Omega_c = -\kappa c^2/(H^2 a^2) = -k^2c^2/(H^2R^2(t))$. Thus

$$\Omega_{c0} = -\frac{kc^2}{H_0^2 R_0^2} \Rightarrow R_0 = \frac{c}{H_0}\sqrt{-\frac{k}{\Omega_{c0}}} = \frac{c}{H_0}\sqrt{\frac{1}{|\Omega_{c0}|}}. \quad (4.119)$$

The above formula works for $k = \pm 1$. This formula will also lead to the correct result for $k = 0$ as we will now show. Thus for $k = \pm 1$ we have

$$\frac{c}{R_0 H_0} = \sqrt{|\Omega_{c0}|}. \quad (4.120)$$

In other words

$$d_L = (1+z)\frac{c}{H_0}\frac{1}{\sqrt{|\Omega_{c0}|}} S_k\bigg(\sqrt{|\Omega_{c0}|}\int_0^z \frac{dz'}{E(z')}\bigg). \quad (4.121)$$

For $k = 0$, the curvature density Ω_{c0} vanishes but it cancels exactly in this last formula for d_L and we get therefore the correct answer which can be checked by comparing with the original formula (4.118).

The above formula allows us to compute the distance to any source at redshift z given H_0 and Ω_{i0} which are the Hubble parameter and the density parameters at our epoch. Conversely given the distance d_L at various values of the redshift we can extract H_0 and Ω_{i0}.

4.4.5 Other Distances

Proper Motion Distance: This is the distance inferred from the proper and observable motion of the source. This is given by

$$d_M = \frac{u}{\dot{\theta}}. \quad (4.122)$$

The u is the proper transverse velocity whereas $\dot{\theta}$ is the observed angular velocity. We can check that

$$d_M = \frac{d_L}{1+z}. \quad (4.123)$$

The Angular Diameter Distance: This is the distance inferred from the proper and observed size of the source. This is given by

$$d_A = \frac{S}{\theta}. \quad (4.124)$$

The S is the proper size of the source and θ is the observed angular diameter. We can check that

$$d_A = \frac{d_L}{(1+z)^2}. \quad (4.125)$$

4.4.6 Age of the Universe

Let t_0 be the age of the universe today and let t_* be the age of the universe when the photon was emitted. The difference $t_0 - t_*$ is called lookback time. This is given by

$$\begin{aligned} t_0 - t_* &= \int_{t_*}^{t_0} dt \\ &= \int_{a_*}^{1} \frac{da}{aH(a)} \\ &= \frac{1}{H_0}\int_0^{z_*} \frac{dz^{'}}{(1+z^{'})E(z^{'})}. \end{aligned} \tag{4.126}$$

For a flat ($k=0$) matter-dominated ($\rho \simeq \rho_M = \rho_{M0}a^{-3}$) universe we have $\Omega_{M0} \simeq 1$ and hence $E(z^{'}) = \sqrt{\Omega_{M0}(1+z^{'})^3 + ...} = (1+z^{'})^{3/2}$. Thus

$$\begin{aligned} t_0 - t_* &= \frac{1}{H_0}\int_0^{z_*} \frac{dz^{'}}{(1+z^{'})^{\frac{5}{2}}} \\ &= \frac{2}{3H_0}\bigg[1 - \frac{1}{(1+z_*)^{\frac{3}{2}}}\bigg]. \end{aligned} \tag{4.127}$$

By allowing $t_* \longrightarrow 0$ we get the actual age of the universe. This is equivalent to $z_* \longrightarrow \infty$ since a photon emitted at the time of the big bang will be infinitely redshifted , i.e. unobservable. We get then

$$t_0 = \frac{2}{3H_0}. \tag{4.128}$$

Chapter 5

Cosmology III: The Inflationary Universe

5.1 Cosmological Puzzles

The isotropy and homogeneity of the universe and its spatial flatness are two properties which are highly non generic and as such they can only arise from very special set of initial conditions which is a very unsatisfactory state of affair. Inflation is a dynamical mechanism which allows us to go around this problem by permitting the universe to evolve to the state of isotropy/homogeneity and spatial flatness from a wide range of initial conditions.

Another problem solved by inflation is the relics problem. Relics refer to magnetic monopoles, domain walls and supersymmetric particles which are assumed to be produced during the early universe yet they are not seen in observations.

As it turns out inflation does also provide the mechanism for the formation of large scale structure in the universe starting from minute quantum fluctuations in the early universe.

5.1.1 Homogeneity/Horizon Problem

The metric of the FLRW universe can be put in the form

$$ds^2 = -dt^2 + R^2(t)(d\chi^2 + S_k^2(\chi)d\Omega^2). \tag{5.1}$$

$$\begin{aligned} S_k(\chi) &= \sin\chi \ , \ k = +1 \\ S_k(\chi) &= \chi \ , \ k = 0 \\ S_k(\chi) &= \sinh\chi \ , \ k = -1. \end{aligned} \tag{5.2}$$

We introduce the conformal time τ by

$$\tau = \int \frac{dt}{R(t)}. \tag{5.3}$$

The FLRW metric becomes

$$\begin{aligned} ds^2 &= R^2(\tau)\Big[-d\tau^2+d\chi^2+S_k^2(\chi)d\Omega^2\Big] \\ &\equiv R^2(\tau)\Big[-d\tau^2+d\vec{\chi}^2\Big]. \end{aligned} \tag{5.4}$$

The motion of photons in the Friedmann-Lemaître-Robertson-Walker universe is given by null geodesics $ds^2=0$. In an isotropic universe it is sufficient to consider only radial motion. The condition $ds^2=0$ is then equivalent to $d\tau=d\chi$. The maximum comoving distance a photon can travel since the initial singularity at $t=t_i$ ($R(t_i)=0$) is

$$\chi_{\rm hor}(t)\equiv\tau-\tau_i=\int_{t_i}^t\frac{dt_1}{R(t_1)}. \tag{5.5}$$

The is called the particle horizon. Indeed, particles separated by distances larger than $\chi_{\rm hor}$ could have never been in causal contact. On the other hand, the comoving Hubble radius $1/aH$ is such that particles separated by distances larger than $1/aH$ can not communicate to each other now. The physical size of the particle horizon is

$$d_{\rm hor}=R\chi_{\rm hor}. \tag{5.6}$$

The existence of particle horizons is at the heart of the so-called horizon problem, i.e. of the problem of why the universe is isotropic and homogeneous.

The universe has a finite age and thus photons can only travel a finite distance since the big bang singularity. This distance is precisely $d_{\rm hor}(t)$ which can be rewritten as

$$\begin{aligned} d_{\rm hor}(t) &= a(t)\int_{t_i}^t\frac{dt_1}{a(t_1)} \\ &= a(t)\int_{a(t_i)=0}^{a(t)}\frac{1}{a(t_1)H(t_1)}d\ln a(t_1). \end{aligned} \tag{5.7}$$

The number $1/aH$ is precisely the comoving Hubble radius. The distance $d_{\rm hor}(t_0)$ is effectively the distance to the surface of last scattering which corresponds to the decoupling event.

The first Friedmann equation can be rewritten as $H^2a^2=8\pi G\rho_0a^{-(1+3w)}/3-\kappa$. For a flat universe we have

$$\frac{1}{aH}=\frac{a^{\frac{1}{2}(1+3w)}}{H_0}. \tag{5.8}$$

It is then clear that the particle horizon is given by

$$\chi_{\rm hor} = \frac{2}{H_0(1+3w)}a^{\frac{1}{2}(1+3w)}. \tag{5.9}$$

For a matter-dominated flat universe we have $w=0$ and hence $H=H_0a^{-3/2}$ or equivalently $a=(t/t_0)^{2/3}$. In this case

$$\chi_{\rm hor} = \frac{2}{H_0}a^{\frac{1}{2}}\Rightarrow d_{\rm hor}=\frac{2}{H}. \tag{5.10}$$

For a radiation-dominated flat universe we have $w = 1/3$ and hence $H = H_0 a^{-2}$ or equivalently $a = (t/t_0)^{1/2}$. In this case

$$\chi_{\rm hor} = \frac{1}{H_0} a \Rightarrow d_{\rm hor} = \frac{1}{H}. \tag{5.11}$$

For a flat universe containing both matter and radiation we should get then

$$d_{\rm hor} \sim \frac{1}{H}. \tag{5.12}$$

In other words

$$d_{\rm hor} \sim d_H. \tag{5.13}$$

The so-called Hubble distance d_H is defined simply as the inverse of the Hubble parameter H. This is the source of the horizon problem. Inflation solves this problem by making $d_{\rm hor} >> d_H$.

Let us put this important point in different words. The cosmic microwave background (CMB) radiation consists of photons from the epochs of recombination and photon decoupling. The CMB radiation comes uniformly from every direction of the sky. The physical distance at the time of emission t_e of the source of the CMB radiation as measured from an observer on Earth making an observation at time t_0 is given by

$$\begin{aligned} \Delta d(t_e) &= a(t_e) \int_{t_e}^{t_0} \frac{dt_1}{a(t_1)} \\ &= 3t_e^{2/3}(t_0^{1/3} - t_e^{1/3}) \ : \ \text{MD}. \end{aligned} \tag{5.14}$$

The physical distance between sources of CMB radiation coming from opposite directions of the sky at the time of emission is therefore given by

$$2\Delta d(t_e) = 6t_e^{2/3}(t_0^{1/3} - t_e^{1/3}) \ : \ \text{MD}. \tag{5.15}$$

At the time of emission t_e the maximum distance a photon had traveled since the big bang is

$$\begin{aligned} d_{\rm hor}(t_e) &= a(t_e) \int_0^{t_e} \frac{dt_1}{a(t_1)} \\ &= 3t_e \ : \ \text{MD}. \end{aligned} \tag{5.16}$$

This is the particle horizon at the time of emission, i.e. the maximum distance that light can travel at the time of emission.

We compute

$$\frac{2\Delta d(t_e)}{d_{\rm hor}(t_e)} = 2(a(t_e)^{-1/2} - 1). \tag{5.17}$$

By looking at CMB we are looking at the universe at a scale factor $a_{\rm CMB} \equiv a(t_e) = 1/1200$. Thus

$$\frac{2\Delta d(t_e)}{d_{\rm hor}(t_e)} \simeq 67.28. \tag{5.18}$$

In other words $2\Delta d(t_e) > d_{\rm hor}(t_e)$. The two widely separated parts of the CMB considered above have therefore non overlapping horizons and as such they have no causal contact at recombination (the time of emission t_e), yet these two widely separated parts of the CMB have the same temperature to an incredible degree of precision (this is the observed isotropy/homogeneity property). See figure $1COS, 1$. How did they know how to do that?. This is precisely the horizon problem.

5.1.2 Flatness Problem

The first Friedmann equation can be rewritten as

$$\Omega - 1 = \frac{\kappa}{a^2 H^2}. \tag{5.19}$$

The density parameter is

$$\Omega = \frac{\rho}{\rho_c}. \tag{5.20}$$

The critical density is

$$\rho_c = \frac{3}{8\pi G} H^2. \tag{5.21}$$

We know that $1/(a^2 H^2) = a^{1+3w}/H_0^2$ and thus as the universe expands the quantity $\Omega - 1$ increases, i.e. Ω moves away from 1. The value $\Omega = 1$ is therefore a repulsive (unstable) fixed point since any deviation from this value will tend to increase with time. Indeed we compute (with $g = \Omega - 1$)

$$a\frac{dg}{da} = (1 + 3w)g. \tag{5.22}$$

By assuming the strong energy condition we have $\rho + P \geq 0$ and $\rho + 3P \geq 0$, i.e. $1 + 3w > 0$. The value $\Omega = 1$ is then clearly a repulsive fixed point since $d\Omega/d\ln a > 0$.

As a consequence the value $\Omega \sim 1$ observed today can only be obtained if the value of Ω in the early universe is fine-tuned to be extremely close to 1. This is the flatness problem.

5.2 Elements of Inflation

5.2.1 Solving the Flatness and Horizon Problems

The second Friedmann equation can be put into the form

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P). \tag{5.23}$$

For matter satisfying the strong energy condition, i.e. $\rho + 2P \geq 0$ we have $\ddot{a} < 0$. Inflation is any epoch with $\ddot{a} > 0$. We explain this further below.

We have shown that the first Friedmann equation can be put into the form $|\Omega - 1| = \kappa/(a^2H^2)$. The problem with the hot big bang model (big bang without inflation) is simply that aH is always decreasing so that Ω is always flowing away from 1. Indeed, in a universe filled with a fluid with an equation of state $w = P/\rho$ with the strong energy condition $1+3w > 0$ the comoving Hubble radius is given by

$$\frac{1}{aH} \propto a^{\frac{1+3w}{2}}. \tag{5.24}$$

Thus $\ddot{a} = d(aH)/dt$ is always negative. Inflation is the hypothesis that during the early universe there was a period of accelerated expansion $\ddot{a} > 0$. We write this condition as

$$\ddot{a} = \frac{d(aH)}{dt} > 0 \Leftrightarrow P < -\frac{\rho}{3}. \tag{5.25}$$

Thus the comoving Hubble length $1/(aH)$ is decreasing during inflation whereas in any other epoch it will be increasing. This behavior holds in a vacuum-dominated flat universe ($P = -\rho$, $\rho \propto a^0$, $a(t) \propto \exp(Ht)$). However inflation can only be a phenomena of the early universe and thus must terminate quickly in order for the hot big bang theory to proceed normally.

Inflation solves the flatness problem by construction since in the first Friedmann equation $|\Omega - 1| = \kappa/(a^2H^2)$ the right hand side decreases rapidly during inflation and thus driving Ω towards 1 (towards flatness) quite fast. Another way of putting it using the first Friedmann equation in the form $H^2 = 8\pi G\rho/3 - \kappa/a^2$ is as follows. In a vacuum-dominated (for example) universe the mass density $\rho \propto a^0$ grows very fast with respect to the spatial curvature term $-\kappa/a^2$ and hence the universe becomes flatter very quickly.

The horizon problem is also solved by inflation. Recall that this problem arises from the fact that the physical horizon length $d_{\rm hor}(t_e)$ grows more rapidly with the scale factor (in a matter-dominated or radiation-dominated universe) than the physical distance $2\Delta d(t_e)$ between any two comoving objects. We need therefore to reverse this situation so that

$$\Delta d(t_e) << d_{\rm hor}(t_e). \tag{5.26}$$

Or equivalently

$$\int_0^{t_e} \frac{dt_1}{a(t_1)} >> \int_{t_e}^{t_0} \frac{dt_1}{a(t_1)}. \tag{5.27}$$

This means in particular that we want a situation where photons can travel much further before recombination/decoupling than it can afterwards. Equivalently, if the Hubble radius is decreasing then the strong energy condition is viloated and as a consequence the Big Bang singularity is pushed to infinite negative conformal time since

$$\tau = \int \frac{dt}{a(t)} \propto \frac{2}{1+3w} a^{\frac{1+3w}{2}}. \tag{5.28}$$

In other words, there is much more conformal time between the initial Big Bang singularity and decoupling with inflation.

In a universe with a period of inflation the comoving Hubble length $1/(aH)$ is decreasing during inflation. Thus if we start with a large Hubble length then a sufficiently large and smooth patch within the Hubble length can form by ordinary causal interactions. Inflation will cause this Hubble length to shrink enormously to within the smooth patch and after inflation comes to an end the Hubble length starts increasing again but remains within the smooth patch. See figure 1COS, 2.

This can also be stated as follows. All comoving scales k^{-1} which are relevant today were larger than the Hubble radius until $a = 10^{-5}$ (start of inflation). At earlier times these scales were within the Hubble radius and thus were casually connected whereas at recent times these scales re-entered again within the Hubble radius. See figure 1COS, 3.

The observable universe is therefore one causal patch of a much larger unobservable universe. In other words there are parts of the universe which cannot communicate with us yet but they will eventually come into view as the cosmological horizon moves out and which will appear to us no different from any other region of space we have already seen since they are within the smooth patch. This explains homogeneity or the horizon problem. However there are possibly other parts of the universe outside the smooth patch which are different from the observable universe.

5.2.2 Inflaton

Inflation can be driven by a field called inflaton. This is a scalar field coupled to gravity with dynamics given by the usual action

$$S_\phi = \int d^4x\sqrt{-\det g}\,\bigg[-\frac{1}{2}g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi - V(\phi)\bigg]. \tag{5.29}$$

The equations of motion read

$$\begin{aligned}\frac{\delta S_\phi}{\delta\phi} &\equiv \nabla_\nu(g^{\mu\nu}\nabla_\mu\phi) - \frac{\delta V}{\delta\phi}\\ &= \frac{1}{\sqrt{-\det g}}\partial_\mu(\sqrt{-\det g}\;\partial^\mu\phi) - \frac{\delta V}{\delta\phi}\\ &= \partial_\mu\partial^\mu\phi + \frac{1}{2}g^{\alpha\beta}\partial_\mu g_{\alpha\beta}\partial^\mu\phi - \frac{\delta V}{\delta\phi}\\ &= 0.\end{aligned} \tag{5.30}$$

For a homogeneous field $\phi \equiv \phi(t,\vec{x}) = \phi(t)$ we obtain

$$\partial_0\partial^0\phi + \frac{1}{2}g^{\alpha\beta}\partial_0 g_{\alpha\beta}\partial^0\phi - \frac{\delta V}{\delta\phi} = 0. \tag{5.31}$$

In the RW metric we obtain

$$\ddot{\phi}+3H\dot{\phi}+\frac{\delta V}{\delta\phi}=0. \tag{5.32}$$

The corresponding stress-energy-momentum tensor is calculated to be given by

$$T^{(\phi)}_{\mu\nu}=\nabla_{\mu}\phi\nabla_{\nu}\phi-\frac{1}{2}g_{\mu\nu}g^{\rho\sigma}\nabla_{\rho}\phi\nabla_{\sigma}\phi-g_{\mu\nu}V(\phi). \tag{5.33}$$

Explicit calculation shows that this stress-energy-momentum tensor is of the form of the stress-energy-momentum tensor of a perfect fluid $T_{\mu}{}^{\nu}=(-\rho_{\phi},P_{\phi},P_{\phi},P_{\phi})$ with

$$\rho_{\phi}=\frac{1}{2}\dot{\phi}^2+V\ ,\ P_{\phi}=\frac{1}{2}\dot{\phi}^2-V. \tag{5.34}$$

The equation of state is therefore

$$w_{\phi}=\frac{P_{\phi}}{\rho_{\phi}}=\frac{\frac{1}{2}\dot{\phi}^2-V}{\frac{1}{2}\dot{\phi}^2+V}. \tag{5.35}$$

We can have accelerated expansion $w_{\phi}<-1/3$ if the potential dominates over the kinetic energy. In other words we will have inflation whenever the potential dominates. The first Friedmann equation in this case reads (assuming also flatness)

$$H^2=\frac{8\pi G}{3}(\frac{1}{2}\dot{\phi}^2+V). \tag{5.36}$$

The second Friedmann equation reads

$$\begin{aligned}\frac{\ddot{a}}{a} &= -\frac{8\pi G}{3}(\dot{\phi}^2-V)\\ &= H^2(1-\epsilon).\end{aligned} \tag{5.37}$$

The so-called slow-roll parameter is given by

$$\begin{aligned}\epsilon &= 4\pi G\frac{\dot{\phi}^2}{H^2}\\ &= \frac{3}{2}(1+w_{\phi}).\end{aligned} \tag{5.38}$$

This can also be expressed as

$$\epsilon = -\frac{\dot{H}}{H^2}. \tag{5.39}$$

Let us introduce $N=\ln a$, i.e. $dN=Hdt$. Then we can show that

$$\epsilon = -\frac{d\ln H}{dN}. \tag{5.40}$$

Inflation corresponds to $\epsilon < 1$. In the so-called de Sitter limit $P_\phi \longrightarrow -\rho_\phi$ ($w_\phi \longrightarrow -1$, $\epsilon \longrightarrow 0$) we observe that the kinetic energy can be neglected compared to the potential energy, i.e. $\dot{\phi}^2 << V$. We have then

$$\epsilon << 1 \Leftrightarrow \dot{\phi}^2 << V. \tag{5.41}$$

This condition means that the scalar field is moving very slowly because for example the potential is flat.

In order to maintain accelerated expansion for a sufficient long time we require also that the second derivative of ϕ is small enough, viz $|\ddot{\phi}| << |3H\dot{\phi}|$ and $|\ddot{\phi}| << |\delta V/\delta\phi|$. This second condition means that the field keeps moving slowly over a wide range of its values and hence the term slowly rolling. We compute

$$\frac{1}{2\epsilon}\frac{d\epsilon}{dt} = \frac{\ddot{\phi}}{\dot{\phi}} - \frac{\dot{H}}{H} \Rightarrow \frac{1}{2\epsilon}\frac{d\epsilon}{dN} = \frac{\ddot{\phi}}{H\dot{\phi}} + \epsilon. \tag{5.42}$$

We introduce a second slow-roll parameter by

$$\begin{aligned} \eta &= -\frac{\ddot{\phi}}{\dot{\phi}H} \\ &= \epsilon - \frac{1}{2\epsilon}\frac{d\epsilon}{dN}. \end{aligned} \tag{5.43}$$

It is clear that sustained acceleration is equivalent to the condition $\eta << 1$. In other words

$$\eta << 1 \Leftrightarrow |\ddot{\phi}| << |3H\dot{\phi}| \ , \ |\ddot{\phi}| << |\frac{\delta V}{\delta\phi}|. \tag{5.44}$$

The equations of motion in the slow-roll regime are

$$H^2 \simeq \frac{8\pi G}{3}V \ , \ \dot{\phi} = -\frac{\delta V/\delta\phi}{3H}. \tag{5.45}$$

Since ϕ is almost constant during slow-roll we can assume that $H^2 \simeq$ constant in this regime and hence $a(t) \simeq \exp(Ht)$. This is de Sitter spacetime.

Instead of the Hubble slow-roll parameters ϵ and η we can work with the potential slow-roll parameters ϵ_V and η_V defined as follows. The first slow-roll parameter ϵ is equivalent to

$$\begin{aligned} \epsilon &\simeq \frac{3}{2}\frac{\dot{\phi}^2}{V} \\ &\simeq \frac{1}{6H^2}\frac{(\delta V/\delta\phi)^2}{V} \\ &\simeq \frac{1}{16\pi G}\frac{(\delta V/\delta\phi)^2}{V^2} \\ &\simeq \epsilon_V. \end{aligned} \tag{5.46}$$

We compute

$$\frac{1}{2\epsilon}\frac{d\epsilon}{dt}\simeq\frac{\dot{\phi}}{\delta V/\delta\phi}\bigg(\frac{\delta^2 V}{\delta\phi^2}-\frac{1}{V}\big(\frac{\delta V}{\delta\phi}\big)^2\bigg)\Leftrightarrow\frac{1}{2\epsilon}\frac{d\epsilon}{dN}\simeq\frac{\dot{\phi}}{H\delta V/\delta\phi}\bigg(\frac{\delta^2 V}{\delta\phi^2}-\frac{1}{V}\big(\frac{\delta V}{\delta\phi}\big)^2\bigg). \tag{5.47}$$

The second slow-roll parameter η is therefore equivalent to

$$\begin{aligned}\eta &\simeq \frac{1}{8\pi G}\frac{\delta^2 V/\delta\phi^2}{V}-\epsilon_V\\ &\simeq \eta_V-\epsilon_V.\end{aligned} \tag{5.48}$$

The slow-roll conditions $\epsilon, |\eta| << 1$ are equivalent to

$$\epsilon_V, |\eta_V| << 1. \tag{5.49}$$

These are obviously conditions on the shape of the inflationary potential. The first (inflation) states that the slope of the potential is small whereas the second (prolonged inflation) states that the curvature of the potential is small. These are necessary conditions for the slow-roll state but they are not sufficient. For example a potential could be very flat but the velocity of the field is very large.

The amount of inflation is defined by the logarithm of the expansion or equivalently the number of $e-$foldings N given by

$$\begin{aligned}N=\ln\frac{a(t_{\rm end})}{a(t_{\rm initial})} &= \int_{t_i}^{t_e}H dt\\ &= \int_{\phi(t_i)}^{\phi(t_e)}\frac{H}{\dot{\phi}}d\phi\\ &= 8\pi G\int_{\phi(t_i)}^{\phi(t_e)}\frac{V}{\delta V/\delta\phi}d\phi\\ &= \sqrt{8\pi G}\int_{\phi(t_i)}^{\phi(t_e)}\frac{1}{\sqrt{2\epsilon_V}}d\phi.\end{aligned} \tag{5.50}$$

In order to solve the horizon and the flatness problems we need a minimum amount of inflation of at least 60 $e-$foldings which is equivalent to an expansion by a factor of 10^{30}.

Inflation ends at the value of the field $\phi_{\rm end}$ where the kinetic energy becomes comparable to the potential energy. This is the value where the slow-roll conditions breaks down, viz $\epsilon(\phi_{\rm end})=1$, $\epsilon_V(\phi_{\rm end})\simeq 1$. After inflation ends the scalar field starts oscillating around the minimum of the potential and then decays into conventional matter. This is the process of reheating which is followed by the usual hot big bang theory. See figure 1COS, 3.

The most simple and interesting models of inflation involve 1) a single rolling scalar field and 2) a potential V which satisfies the slow-roll conditions in some regions and possesses a minimum with zero potential where inflation terminates. Some of these models are

$$V=\lambda_p\phi^p\ ,\ \text{chaotic inflation}. \tag{5.51}$$

$$V = V_0(1 + \cos\frac{\phi}{f}) \ , \ \text{natural inflation.} \tag{5.52}$$

$$V = V_0 \exp(\alpha\phi) \ , \ \text{power} - \text{law inflation.} \tag{5.53}$$

5.2.3 Amount of Inflation

As pointed out above in order to solve the horizon and the flatness problems we need a minimum amount of inflation of at least 60 $e-$foldings which is equivalent to an expansion by a factor of 10^{30}. A clean argument is found in [41].

We imagine a universe in which inflation starts at t_i with a scale factor $a(t_i)$ and ends at t_f with a scale factor $a(t_f)$. The current time is t_0 with a scale factor $a(t_0)$. During inflation we can assume that H is constant (de Sitter spacetime) and as a consequence the (vacuum) mass density ρ_V is constant. We assume for simplicity that between the end of inflation and the current moment the universe is radiation-dominated with a density ρ_R behaving as $1/a^4$. We assume that at t_f the vacuum density is fully converted into radiation, viz $\rho_R(t_f) = \rho_V$. We recall that the density parameter Ω_C associated with curvature is given by

$$\Omega_C = \frac{\rho_C}{\rho_c} = -\frac{\kappa}{a^2H^2} = -\frac{\kappa}{\dot{a}^2}. \tag{5.54}$$

During inflation the expansion is accelerating, since gravity is acting as repulsive due to the dominance of the vacuum energy, and thus $\dot{a}$ increases and hence Ω_C decreases. Thus $\Omega_C << 1$ today at t_0 can be easily explained with more inflation. On the other hand, if inflation is preceded with a long phase of deceleration in which gravity acts in the usual way as attractive, then $\dot{a}$ at t_i must be very small and hence $\Omega_C >> 1$ at the beginning of inflation. This case also would only require more inflation to explain. Thus it is sufficient to assume that $\Omega_C \sim 1$ at t_0 and t_i. This means that at t_0 and t_i ρ_C is equal to the critical density ρ_c. In other words, at t_0 and t_i we have nothing but curvature, viz

$$\rho_V(t_i) = \rho_C(t_i) \ , \ \rho_R(t_0) = \rho_C(t_0). \tag{5.55}$$

We compute now

$$\begin{aligned}
\frac{\rho_C(a_0)}{\rho_C(a_i)} &= (\frac{a_i}{a_0})^2 \\
&= \frac{\rho_R(a_0)}{\rho_V(a_i)} \\
&= \frac{\rho_R(a_0)}{\rho_V(a_f)} \\
&= \frac{\rho_R(a_0)}{\rho_R(a_f)} \\
&= (\frac{a_f}{a_0})^4.
\end{aligned} \tag{5.56}$$

The solution is thus

$$\frac{a_f}{a_i}=\sqrt{\frac{a_0}{a_i}}=\frac{a_0}{a_f}. \tag{5.57}$$

In general we obtain

$$\frac{a_f}{a_i}\geq\frac{a_0}{a_f}. \tag{5.58}$$

In terms of the e-fold number $N=\ln a$ this reads

$$N_f-N_i\geq N_0-N_f. \tag{5.59}$$

The amount of inflation is precisely $\Delta N=N_f-N_i$. Thus we have more expansion during inflation than since the end of inflation. Although there is no upper bound on the amount of inflation, there is a lower bound given by

$$\begin{aligned}\Delta N_{\rm min} &= N_0-N_f\\ &= \ln\frac{a(t_0)}{a(t_f)}\\ &= \frac{1}{4}\ln\frac{\rho_R(a_f)}{\rho_R(a_0)}.\end{aligned} \tag{5.60}$$

As we will show later the energy scale of inflation is $\rho_R(a_f)\sim 10^{-12}\rho_{\rm pl}$. Also we have already seen that the energy density contained in radiation is $\rho_{\rm rad}=10^{-34}g/{\rm cm}^3=10^{-127}\rho_{\rm pl}$ where $\rho_{\rm pl}=10^{93}g/{\rm cm}^3$. Hence the minimum amount of inflation is

$$\Delta N_{\rm min} = \frac{1}{4}\ln\frac{10^{-12}}{10^{-127}}\sim 66. \tag{5.61}$$

From this approach we can get another important estimation. We have

$$\begin{aligned}\Delta N=\ln\frac{a(t_f)}{a(t_i)} &= \int H dt\\ &= H_i(t_f-t_i)+\frac{H_i^2(t_f-t_i)^2}{2}\frac{\dot{H}_i}{H_i^2}+...\end{aligned} \tag{5.62}$$

We must then have

$$H_i(t_f-t_i)>66. \tag{5.63}$$

$$\frac{|\dot{H}_i|}{H_i^2}<\frac{1}{66}. \tag{5.64}$$

However, from the Friedmann equations $H^2=8\pi G\bar{\rho}/3$, $\dot{H}=-4\pi G(\bar{\rho}+\bar{\mathcal{P}})$ we have

$$\frac{\dot{H}}{H^2}=-\frac{3}{2}(1+\frac{\bar{\mathcal{P}}}{\bar{\rho}}). \tag{5.65}$$

We get immediately the estimate

$$1+\frac{\bar{\mathcal{P}}_i}{\bar{\rho}_i}<\frac{1}{99}\sim 10^{-2}. \tag{5.66}$$

5.2.4 End of Inflation: Reheating and Scalar-Matter-Dominated Epoch

We start by summarizing the main results of the previous section. The main equations are the equation of motion of the inflaton scalar field and the Friedmann equation given respectively by

$$\ddot{\phi}+3H\dot{\phi}+\frac{\partial V}{\partial\phi}=0. \tag{5.67}$$

$$H^2=\frac{8\pi G}{3}(\frac{1}{2}\dot{\phi}^2+V). \tag{5.68}$$

The slow-roll approximation is given by the conditions

$$\epsilon<<1\ ,\ \eta<<1\Leftrightarrow\dot{\phi}^2<<V\ ,\ |\ddot{\phi}|<<|3H\dot{\phi}|\ ,\ |\ddot{\phi}|<<|\frac{\delta V}{\delta\phi}|. \tag{5.69}$$

Equivalently

$$\epsilon_V<<1\ ,\ \eta_V<<1\Leftrightarrow\frac{(\delta V/\delta\phi)^2}{V^2}<<1\ ,\ \frac{\delta^2V/\delta\phi^2}{V}<<1. \tag{5.70}$$

The equations of motion during slow-roll are:

$$H^2\simeq\frac{8\pi G}{3}V\ ,\ \dot{\phi}=-\frac{\delta V/\delta\phi}{3H}. \tag{5.71}$$

These two equations can be combined to give the equation of motion

$$\frac{d\ln a}{d\phi}=-8\pi G\frac{V}{\frac{\partial V}{\partial\phi}}. \tag{5.72}$$

The solution is

$$a(\phi)=a_i\exp\bigg(-8\pi G\int_{\phi_i}^{\phi}\frac{V}{\frac{\partial V}{\partial\phi}}d\phi\bigg). \tag{5.73}$$

For a power-law potential $V=\lambda\phi^n/n$ the slow-roll conditions are equivalent to $\phi>>1$ and the above solution is given by

$$a(\phi)=a_i\exp\bigg(-\frac{4\pi G}{n}(\phi^2-\phi_i^2)\bigg). \tag{5.74}$$

Let us consider a quadratic potential $V=m^2\phi^2/2$ at the end of inflation. By combining the Friedmann equation and the equation of motion of the scalar field we obtain a differential equation for $\dot{\phi}$ as a function of ϕ given by

$$\frac{d\dot{\phi}}{d\phi}=-\frac{m^2\phi+\sqrt{12\pi G(\dot{\phi}^2+m^2\phi^2)}}{\dot{\phi}}. \tag{5.75}$$

In slow-roll we have $d\dot{\phi}/d\phi \sim 0$, $\dot{\phi}=$ constant and $\phi >> 1$. A solution is given by

$$\dot{\phi}=-\frac{m}{\sqrt{12\pi G}}\ ,\ \phi=-\frac{m}{\sqrt{12\pi G}}(t_f-t). \tag{5.76}$$

Since $|\phi| >> 1$ during inflation we must have $mt >> 1$. The pressure is given by

$$P=-\rho+\dot{\phi}^2=-\rho+\frac{m^2}{12\pi G}. \tag{5.77}$$

When the scalar field drops to its Planck value $\phi \sim 1/\sqrt{12\pi G}$ we observe that the energy density drops to $m^2/12\pi G$ and hence the pressure vanishes. Inflation is then over. Thus inflation ends when the scalar field becomes of order 1 in Planck units. The duration of inflation is

$$\Delta t=t_f-t_i=-\frac{\sqrt{12\pi G}}{m}\Delta\phi=\sqrt{12\pi G}\frac{\phi_i}{m}. \tag{5.78}$$

By substituting the above solution into Friedmann equation we get

$$H=\frac{1}{3}m^2(t_f-t)=-\sqrt{\frac{4\pi G}{3}}m\phi. \tag{5.79}$$

$$a(t)=a(t_f)\exp\big(-\frac{m^2}{6}(t_f-t)^2\big) \tag{5.80}$$

We get immediately

$$\frac{a_f}{a_i}=\exp\big(2\pi G\phi_i^2\big). \tag{5.81}$$

Thus in order to get a 75 e-folds we must start with a value of the scalar field which is four times the Planck value, viz $\phi_i \sim 4/\sqrt{G}$.

Alternatively, the Friedmann equation can be immediately solved by the ansatz

$$m\phi=\sqrt{\frac{3}{4\pi G}}H\cos\theta. \tag{5.82}$$

$$\dot{\phi}=\sqrt{\frac{3}{4\pi G}}H\sin\theta. \tag{5.83}$$

By taking the time derivative of the first equation and comparing with the second one we get

$$\frac{\dot{H}}{H}\cos\theta-\dot{\theta}\sin\theta=m\sin\theta. \tag{5.84}$$

By taking the time derivative of the second equation and comparing with the value of $\ddot{\phi}$ obtained from the equation of motion of the inflaton field we get

$$\frac{\dot{H}}{H}\sin\theta+\dot{\theta}\cos\theta=-3H\sin\theta-m\cos\theta. \tag{5.85}$$

Solving the above two equations for $\dot{H}$ and $\dot{\theta}$ in terms of the original variables H and θ we get

$$\dot{H} = -3H^2 \sin^2\theta \ , \ \dot{\theta} = -m - \frac{3}{2}H\sin 2\theta. \tag{5.86}$$

In terms of α defined by $\theta = -mt + \alpha$ these read

$$\dot{H} = -3H^2 \sin^2(mt-\alpha) \ , \ \dot{\alpha} = \frac{3}{2}H\sin 2(mt-\alpha). \tag{5.87}$$

For $mt >> 1$, i.e. towards the end of inflation, we can neglect α:

$$\dot{H} = -3H^2 \sin^2(mt). \tag{5.88}$$

The solution is

$$H = \frac{2}{3t}\bigg(1 - \frac{\sin 2mt}{2mt}\bigg)^{-1} = \frac{2}{3t}\bigg(1 + \frac{\sin 2mt}{2mt} + ...\bigg). \tag{5.89}$$

Now we can check directly that α corresponds to oscillations with decaying amplitude. We can also show that the scalar field oscillates with a frequency $\omega = m$ with slowly decaying amplitude. On the other hand, the scale factor behaves as

$$a = t^{2/3}\big(1 + O(\frac{1}{(m^2t^2)})\big). \tag{5.90}$$

We get therefore a graceful exit into a matter-dominated phase. In summary, if the mass is sufficiently small compared to the Planck mass the inflationary phase will last sufficiently long and is followed by a matter-dominated phase. Furthermore, a quadratic potential gives rise naturally to a post-inflationary matter-dominated universe consisting of heavy scalar particles which will eventually be converted into photons, baryons and leptons (reheating). The usual radiation-dominated, matter-dominated and vacuum-dominated phases follow after reheating.

Other power-law potentials will also give oscillatory stages with scale factors behaving as $a \sim t^p$. For example, a quartic potential will give an oscillating scalar field with the scale factor of a radiation-dominated universe, viz $a \sim t^{1/2}$.

5.3 Perfect Fluid Revisited

Let ρ be the mass density of a perfect fluid, P its pressure, S its entropy per unit mass and $\vec{u}$ its flow velocity vector, i.e. the velocity of an element of fluid at a point $\vec{x}$ at a time t. The equation of state of the perfect fluid allows us to determine the pressure in terms of the mass density ρ and the entropy S, viz

$$P = P(\rho, S). \tag{5.91}$$

The state of the prefect fluid is therefore completely determined by the mass density ρ, the entropy per unit mass S and the flow velocity vector $\vec{u}$. In the absence of dissipation the entropy is conserved, i.e.

$$\begin{aligned} \frac{dS}{dt} &= \frac{\partial S}{\partial t} + (\vec{u}.\vec{\nabla})S \\ &= 0. \end{aligned} \tag{5.92}$$

The mass M contained in a volume V is given by

$$M = \int_V dV \rho. \tag{5.93}$$

The rate of change of the mass contained in V is obviously given by

$$\frac{dM}{dt} = \int_V dV \frac{\partial \rho}{\partial t}. \tag{5.94}$$

This rate of change is also obviously given by the mass flowing through the surface Σ which encloses the volume V. Since the amount of mass flowing per unit area per unit time is $\vec{J} = \rho\vec{u}$ the rate of change dM/dt can be rewritten as

$$\frac{dM}{dt} = -\oint_\Sigma d\vec{\sigma}\vec{J} = -\int_V dV \vec{\nabla}(\rho\vec{u}). \tag{5.95}$$

We get therefore the continuity equation

$$\frac{\partial \rho}{\partial t} + \vec{\nabla}(\rho\vec{u}) = 0. \tag{5.96}$$

The Newtonian gravitational potential Φ generated by the mass density ρ is given by the Poisson equation

$$\nabla\Phi = 4\pi G\rho. \tag{5.97}$$

The force exerted by this potential Φ on a mass ΔM is given by Newton's law of gravitation, viz

$$\vec{F}_{\rm gr} = -\Delta M \vec{\nabla}\Phi. \tag{5.98}$$

The other force acting on ΔM is due to the pressure P of the perfect fluid and is given by

$$\begin{aligned} \vec{F}_{\rm pr} &= -\oint_{\Delta\Sigma} P d\vec{\sigma} \\ &= -\oint_{\Delta V} \vec{\nabla} P dV \\ &\simeq -\vec{\nabla} P \Delta V. \end{aligned} \tag{5.99}$$

Newton's second law then reads

$$-\Delta M\vec{\nabla}\Phi - \vec{\nabla}P\Delta V = \Delta M\vec{g}. \tag{5.100}$$

However,

$$\vec{g} = \frac{d\vec{u}}{dt} = \frac{\partial\vec{u}}{\partial t} + (\vec{u}\vec{\nabla})\vec{u}. \tag{5.101}$$

By identification we obtain Euler's equation

$$\frac{\partial\vec{u}}{\partial t} + (\vec{u}\vec{\nabla})\vec{u} + \frac{\vec{\nabla}P}{\rho} + \vec{\nabla}\Phi = 0. \tag{5.102}$$

We have seven equations: equation of state (5.91), conservation of the entropy (5.92), continuity equation (5.96), Poisson's equation (5.97) and three Euler's equations (5.102) for seven unknowns: ρ, P, S, $\vec{u}$ and Φ.

Linearization of these equations around an expanding homogeneous and isotropic universe with mass density $\rho_0 = \rho_0(t)$ and flow velocity vector $\vec{u}_0$ obeying the Hubble law, i.e. $\vec{u}_0 = H(t)\vec{x}$, leads to a Newtonian theory of gravitational instabilities. This topic is discussed at length in [2]. In the following we will concentrate instead on the corresponding general relativistic theory following mostly [2].

5.4 Cosmological Perturbations

5.4.1 Metric Perturbations

The universe is isotropic and homogeneous and spatially flat with a gravitational field described by the RobertsonWalker metric. This is the punch line so far. However, this is just an approximation which neglects the most obvious fact we observe directly around us which is the presence of structure: galaxies, stars and us. All departures from homogeneity and isotropy will be assumed to be small given by weak first order fluctuations. The perturbed metric is given by

$$g_{\mu\nu} = \bar{g}_{\mu\nu} + \delta g_{\mu\nu}. \tag{5.103}$$

$$d\bar{s}^2 = \bar{g}_{\mu\nu}dx^\mu dx^\nu = -dt^2 + a^2(t)dx^i dx^i. \tag{5.104}$$

The inverse is given by

$$g^{\mu\nu} = \bar{g}^{\mu\nu} - \delta g^{\mu\nu} \ , \ \delta g^{\mu\nu} = \bar{g}^{\mu\alpha}\bar{g}^{\nu\beta}\delta g_{\alpha\beta}. \tag{5.105}$$

The metric is a symmetric $(0,2)$ tensor containing 10 degrees of freedom. The 00 component is a scalar under spatial rotations, the $0i$ (or equivalently $i0$) components form a vector under spatial rotations and the ij components form a rank 2 tensor under spatial rotations. We introduce two scalars Φ and Ψ, a vector B_i and a traceless rank 2 tensor E_{ij} under the action of spatial rotations by the relations

$$\delta g_{00} = -2\Phi$$
$$\delta g_{0i} = aB_i$$
$$\delta g_{ij} = a^2[2E_{ij} - 2\Psi\delta_{ij}]. \tag{5.106}$$

The metric takes the form

$$ds^2 = -(1+2\Phi)dt^2 + 2aB_i dt dx^i + a^2[2E_{ij} + (1-2\Psi)\delta_{ij}]dx^i dx^j. \tag{5.107}$$

$\Psi\delta_{ij}$ and E_{ij} form together a rank 2 tensor under spatial rotations where Ψ is precisely its trace. We call Ψ the spatial curvature perturbation, E_{ij} spatial shear tensor, B_i the shift and Φ the lapse.

We can decompose any 3−vector such as B_i into a divergenceless 3−vector S_i satisfying $\partial^i S_i = 0$, and a total derivative $\partial_i B$ as follows

$$B_i = -S_i + \partial_i B. \tag{5.108}$$

This is Helmholtz's decomposition. Similarly, we can Hodge decompose any symmetric traceless 3−tensor such as E_{ij} into a divergenceless symmetric traceless 3−tensor h_{ij} satisfying $\partial^i h_{ij} = 0$ and $h^i_i = 0$, and a divergenceless 3−vector F_i satisfying $\partial^i F_i = 0$, and a scalar E as follows

$$E_{ij} = \partial_i\partial_j E + \partial_i F_j + \partial_j F_i + \frac{1}{2}h_{ij}. \tag{5.109}$$

We get then the metric

$$ds^2 = -(1+2\Phi)dt^2 + 2a(\partial_i B - S_i)dt dx^i + a^2\left[h_{ij} + 2\partial_i\partial_j E + 2\partial_i F_j + 2\partial_j F_i + (1-2\Psi)\delta_{ij}\right]dx^i dx^j. \tag{5.110}$$

There are 4 scalar degrees of freedom Φ, B, E and Ψ, 4 vector degrees of freedom contained in S_i and F_i which satisfy two constraints, and 2 tensor degrees of freedom contained in h_{ij} which satisfies 4 constraints. Thus the total number of degrees of freedom is $4+4+2=10$ which is precisely the correct number of degrees of freedom contained in the perturbed metric.

As we will see scalars lead to, or are induced by, density fluctuations, and since they can suffer from gravitational instabilities they can also lead to structure formation. On the other hand, tensors lead to gravitational waves and as such they are absent in the Newtonian theory. Further, it can be shown that the vector perturbations S_i and F_i, which are related to the rotational motions of the fluid, decay as $1/a^2$ with the expansion of the universe, which holds true already in Newtonian theory, and thus they do not play an important role in cosmology. The scalar, vector and tensor perturbations evolve independently of each other at the linear order and as such they can be treated separately.

In the following we will mostly neglect vector perturbations for simplicity. The metric will then read

$$ds^2 = -(1+2\Phi)dt^2 + 2a\partial_i B dt dx^i + a^2[h_{ij} + 2\partial_i\partial_j E + (1-2\Psi)\delta_{ij}]dx^i dx^j. \tag{5.111}$$

5.4.2 Gauge Transformations

We recall that coordinate transformations are given by

$$x^\mu \longrightarrow x^{'\mu} = x^\mu + \epsilon^\mu(x). \tag{5.112}$$

$$g_{\mu\nu}(x) \longrightarrow g^{'}_{\mu\nu}(x^{'}) = \frac{\partial x^\alpha}{\partial x^{'\mu}}\frac{\partial x^\beta}{\partial x^{'\nu}} g_{\alpha\beta}(x). \tag{5.113}$$

Explicitly we have

$$\begin{aligned} g^{'}_{\mu\nu}(x^{'}) &= [\delta^\alpha_\mu - \frac{\partial \epsilon^\alpha}{\partial x^{'\mu}}][\delta^\beta_\nu - \frac{\partial \epsilon^\beta}{\partial x^{'\nu}}] g_{\alpha\beta}(x) \\ &= g_{\mu\nu}(x) - \frac{\partial \epsilon^\alpha}{\partial x^{'\nu}} g_{\mu\beta}(x) - \frac{\partial \epsilon^\alpha}{\partial x^{'\mu}} g_{\alpha\nu}(x) \\ &= g_{\mu\nu}(x) - \frac{\partial \epsilon^\alpha}{\partial x^{\nu}} g_{\mu\alpha}(x) - \frac{\partial \epsilon^\alpha}{\partial x^{\mu}} g_{\alpha\nu}(x). \end{aligned} \tag{5.114}$$

We will reinterpret these coordinate transformations as gauge transformations where all change is encoded only in the field perturbations, viz

$$\delta g_{\mu\nu}(x) \longrightarrow \delta g^{'}_{\mu\nu}(x) = \delta g_{\mu\nu}(x) + \Delta\delta g_{\mu\nu}(x). \tag{5.115}$$

Equivalently

$$\begin{aligned} \Delta\delta g_{\mu\nu}(x) &= g^{'}_{\mu\nu}(x) - g_{\mu\nu}(x) \\ &= \delta g^{'}_{\mu\nu}(x) - \delta g_{\mu\nu}(x). \end{aligned} \tag{5.116}$$

We compute

$$\begin{aligned} \Delta\delta g_{\mu\nu}(x) &= g^{'}_{\mu\nu}(x^{'} - \epsilon) - g_{\mu\nu}(x) \\ &= g^{'}_{\mu\nu}(x^{'}) - \epsilon^\lambda \frac{\partial}{\partial x^\lambda} g_{\mu\nu}(x) - g_{\mu\nu}(x) \\ &= -\frac{\partial \epsilon^\beta}{\partial x^\nu}\bar{g}_{\mu\beta}(x) - \frac{\partial \epsilon^\alpha}{\partial x^\mu}\bar{g}_{\alpha\nu}(x) - \epsilon^\lambda \frac{\partial}{\partial x^\lambda}\bar{g}_{\mu\nu}(x). \end{aligned} \tag{5.117}$$

Explicitly we have

$$\Delta\delta g_{ij}(x) = -\frac{\partial \epsilon_i}{\partial x^j} - \frac{\partial \epsilon_j}{\partial x^i} + 2a\dot{a}\epsilon_0\delta_{ij}. \tag{5.118}$$

$$\Delta\delta g_{i0}(x) = -\frac{\partial\epsilon_i}{\partial t} - \frac{\partial\epsilon_0}{\partial x^i} + 2\frac{\dot{a}}{a}\epsilon_i. \tag{5.119}$$

$$\Delta\delta g_{00}(x) = -2\frac{\partial\epsilon_0}{\partial t}. \tag{5.120}$$

Let us consider the metric (5.110) with the vector and tensor perturbations set to zero. We have then

$$ds^2 = -(1+2\Phi)dt^2 + 2a\partial_i B dt dx^i + a^2[2\partial_i\partial_j E + (1-2\Psi)\delta_{ij}]dx^i dx^j. \tag{5.121}$$

The coordinate transformations (5.112) are given explicitly by $x^0 \longrightarrow x^{'0} = x^0 + \epsilon^0$, $x^i \longrightarrow x^{'i} = x^i + \epsilon^i$. We will write $\epsilon^0 = \alpha$. The vector ϵ^i can be decomposed as $\epsilon_i = a^2\partial_i\beta + \epsilon_i^V$ where $\partial^i\epsilon_i^V = 0$. As before we will neglect the vector contribution coming from ϵ_i^V since it will only contribute to the gauge transformations of the vector perturbations which we have dropped. By setting $\epsilon_i^V = 0$ the coordinate transformations (5.112) take now the form

$$t \longrightarrow t^{'} = t + \alpha\ ,\ x^i \longrightarrow x^{'i} = x^i + \partial_i\beta. \tag{5.122}$$

The corresponding gauge transformations are (with $\bar{g}_{ij} = a^2\delta_{ij}$, $\bar{g}_{i0} = 0$, $\bar{g}_{00} = -1$ and $H = \dot{a}/a$)

$$\begin{aligned}
\Delta\Phi &= -\frac{\partial\alpha}{\partial t}\\
\Delta B &= a^{-1}\alpha - a\frac{\partial\beta}{\partial t}\\
\Delta E &= -\beta\\
\Delta\Psi &= H\alpha.
\end{aligned} \tag{5.123}$$

This depends only on two functions α and β. Thus by choosing α and β appropriately we can make any two of the four scalar perturbations E, B, Φ and Ψ vanish. In other words, the space of the physical scalar perturbations is two dimensional. This space is spanned by the two gauge-invariant linear combinations Φ_B and Ψ_B known as Bardeen potentials which are defined by

$$\begin{aligned}
\Phi_B &= \Phi - \frac{d}{dt}\bigg[a(a\partial_t E - B)\bigg]\\
\Psi_B &= \Psi + \dot{a}\bigg[a\partial_t E - B\bigg].
\end{aligned} \tag{5.124}$$

Indeed, we compute

$$\begin{aligned}
\Delta\Phi_B &= \Delta\Phi - \frac{d}{dt}\bigg[a(a\partial_t\Delta E - \Delta B)\bigg] = 0\\
\Delta\Psi_B &= \Delta\Psi + \dot{a}\bigg[a\partial_t\Delta E - \Delta B\bigg] = 0.
\end{aligned} \tag{5.125}$$

Let us now consider the metric (5.110) with the scalar and vector perturbations set to zero. We obtain

$$ds^2 = -dt^2 + a^2[\delta_{ij} + h_{ij}]dx^i dx^j. \tag{5.126}$$

It is obvious from the above discussion that the tensor h_{ij} is invariant under gauge transformations, viz

$$\Delta h_{ij} = 0 \tag{5.127}$$

Some of the used and most useful gauge choices are as follows:

- **Longitudinal, Conformal-Newtonian Gauge:** We can choose β so that $E = 0$ and then choose α so that $B = 0$. These are unique choices which fix the gauge uniquely. This gauge is therefore given by

$$E = B = 0. \tag{5.128}$$

  The metric becomes

$$ds^2 = -(1 + 2\Phi)dt^2 + a^2(1 - 2\Psi)\delta_{ij}dx^i dx^j. \tag{5.129}$$

- **Synchronous Gauge:** We can choose β so that $B = 0$ and then choose α so that $\Phi = 0$. This gauge is therefore given by

$$B = \Phi = 0. \tag{5.130}$$

  The metric becomes

$$ds^2 = -dt^2 + a^2[2\partial_i\partial_j E + (1 - 2\Psi)\delta_{ij}]dx^i dx^j. \tag{5.131}$$

  The synchronous gauge does not fix the gauge completely. Indeed, we can check immediately that the choice $B = \Phi = 0$ remains intact under the gauge transformations

$$\alpha(t, x^i) = f_1(x^i)\ ,\ \beta(t, x^i) = f_1(x^i)\int_{-\infty}^{t}\frac{dt'}{a^2(t')} + f_2(x^i), \tag{5.132}$$

  for any functions $f_i(x^i)$.

5.4.3 Linearized Einstein Equations

We want to linearize the Einstein Equations

$$G^\mu_\nu = R^\mu_\nu - \frac{1}{2}Rg^\mu_\nu = 8\pi G T^\mu_\nu, \tag{5.133}$$

around the perturbed metric (5.103). We have already computed the components of the unperturbed Ricci tensor. We recall (with the prime denoting differentiation with respect to the conformal time, $d\eta = adt$ and $\mathcal{H} = a^{'}/a$)

$$\bar{R}_0^0 = 3\frac{\ddot{a}}{a} = 3\frac{\mathcal{H}^{'}}{a^2}. \tag{5.134}$$

$$\bar{R}_j^i = \frac{\delta_j^i}{a^2}(a\ddot{a} + 2\dot{a}^2) = \frac{\delta_j^i}{a^2}(\mathcal{H}^{'2} + 2\mathcal{H}^2). \tag{5.135}$$

$$\bar{R}_0^i = 0. \tag{5.136}$$

The background stress-energy-momentum must also be diagonal by the background Einstein equations, viz

$$\bar{T}_0^0 \neq 0 \ , \ \bar{T}_0^i = 0 \ , \ \bar{T}_j^i \propto \delta_j^i. \tag{5.137}$$

The linearized Einstein's equations are of the form

$$\delta G_\nu^\mu = 8\pi G \delta T_\nu^\mu. \tag{5.138}$$

Both δG_ν^μ and δT_ν^μ are not gauge invariant. Indeed, under the gauge transformations (5.112) the tensors $\delta X_\nu^\mu = \delta G_\nu^\mu, \delta T_\nu^\mu$ will transform as second rank tensors similarly to δg_ν^μ, i.e. as (5.115) with

$$\begin{aligned}
\Delta\delta X_{\mu\nu}(x) &= X^{'}_{\mu\nu}(x^{'} - \epsilon) - X_{\mu\nu}(x) \\
&= -\frac{\partial\epsilon^\beta}{\partial x^\nu}\bar{X}_{\mu\beta}(x) - \frac{\partial\epsilon^\alpha}{\partial x^\mu}\bar{X}_{\alpha\nu}(x) - \epsilon^\lambda\frac{\partial}{\partial x^\lambda}\bar{X}_{\mu\nu}(x).
\end{aligned} \tag{5.139}$$

More explicitly we have

$$\Delta\delta X_{ij}(x) = -2\partial_i\partial_j\beta.\frac{\bar{X}_{kk}}{3} - \alpha\partial_t\bar{X}_{ij}. \tag{5.140}$$

$$\Delta\delta X_{i0}(x) = -\partial_t\partial_i\beta.\frac{\bar{X}_{kk}}{3} - \partial_i\alpha\bar{X}_{00}. \tag{5.141}$$

$$\Delta\delta X_{00}(x) = -2\partial_t\alpha\bar{X}_{00} - \alpha\partial_t\bar{X}_{00}. \tag{5.142}$$

We can construct gauge invariant quantities as follows. We observe that

$$\Delta I = -\alpha \ , \ I = a(a\partial_t E - B) \text{ and } \Delta E = -\beta. \tag{5.143}$$

Thus the following combinations are gauge invariant:

$$\Delta\delta\hat{X}_{ij} = 0 \ , \ \delta\hat{X}_{ij} = \delta X_{ij} - 2\partial_i\partial_j E.\frac{\bar{X}_{kk}}{3} - I\partial_t\bar{X}_{ij}. \tag{5.144}$$

$$\Delta\delta\hat{X}_{i0}=0\ ,\ \delta\hat{X}_{i0} = \delta X_{i0}-\partial_i\big(\partial_t E.\frac{\bar{X}_{kk}}{3}+I\bar{X}_{00}\big). \tag{5.145}$$

$$\Delta\delta\hat{X}_{00}=0\ ,\ \delta\hat{X}_{00} = \delta X_{00}-2\bar{X}_{00}\partial_t I-I\partial_t\bar{X}_{00}. \tag{5.146}$$

We use now the result

$$\begin{aligned}\Delta\delta X^{\alpha}_{\nu} &= \bar{g}^{\alpha\mu}\Delta\delta X_{\mu\nu}-\Delta\delta g^{\alpha\mu}.\bar{X}_{\mu\nu}\\ &= \bar{g}^{\alpha\mu}\Delta\delta X_{\mu\nu}+\bigg(\partial^{\mu}\epsilon^{\alpha}+\partial^{\alpha}\epsilon^{\mu}+\bar{g}^{\alpha\rho}\bar{g}^{\mu\sigma}\epsilon^{\lambda}\partial_{\lambda}\bar{g}_{\rho\sigma}\bigg)\bar{X}_{\mu\nu}.\end{aligned} \tag{5.147}$$

We get now the gauge invariant observables

$$\Delta\delta\hat{X}^0_0=0\ ,\ \delta\hat{X}^0_0 = \delta X^0_0-I\partial_t\bar{X}^0_0. \tag{5.148}$$

$$\Delta\delta\hat{X}^0_i=0\ ,\ \delta\hat{X}^0_i = \delta X^0_i-\partial_i I(\bar{X}^0_0-\frac{1}{3}\bar{X}^k_k). \tag{5.149}$$

$$\Delta\delta\hat{X}^i_j=0\ ,\ \delta\hat{X}^i_j = \delta X^i_j-I\partial_t\bar{X}^i_j. \tag{5.150}$$

We can then write the linearized Einstein equations in a gauge invariant way as follows

$$\delta\hat{G}^{\mu}_{\nu}=8\pi G\delta\hat{T}^{\mu}_{\nu}. \tag{5.151}$$

We start from

$$ds^2 = -dt^2+a^2\delta_{ij}dx^idx^j=a^2\big[-d\eta^2+\delta_{ij}dx^idx^j\big]. \tag{5.152}$$

From here on the subscript 0 indicates conformal time. We compute

$$\bar{\Gamma}^0_{00}=\frac{\dot{a}}{a}\ ,\ \bar{\Gamma}^0_{ij}=\frac{\dot{a}}{a}\delta_{ij}\ ,\ \bar{\Gamma}^i_{0j}=\frac{\dot{a}}{a}\delta_{ij}. \tag{5.153}$$

$$\bar{R}_{00}=-3\partial_0(\frac{\dot{a}}{a})\ ,\ \bar{R}_{ij}=\bigg(\partial_0(\frac{\dot{a}}{a})+2\frac{\dot{a}^2}{a^2}\bigg)\delta_{ij}\ ,\ \bar{R}=\frac{6}{a^2}(\partial_0(\frac{\dot{a}}{a})+\frac{\dot{a}^2}{a^2}). \tag{5.154}$$

Now we have the perturbations

$$\delta R_{\mu\nu} = \partial_{\alpha}\delta\Gamma^{\alpha}_{\mu\nu}-\partial_{\mu}\delta\Gamma^{\alpha}_{\alpha\nu}+\delta\Gamma^{\beta}_{\mu\nu}\bar{\Gamma}^{\alpha}_{\alpha\beta}+\bar{\Gamma}^{\beta}_{\mu\nu}\delta\Gamma^{\alpha}_{\alpha\beta}-\delta\Gamma^{\beta}_{\alpha\nu}\bar{\Gamma}^{\alpha}_{\mu\beta}-\bar{\Gamma}^{\beta}_{\alpha\nu}\delta\Gamma^{\alpha}_{\mu\beta}. \tag{5.155}$$

And[1]

$$\delta\Gamma^{\rho}_{\mu\nu} = \frac{1}{2}\bar{g}^{\rho\sigma}(\partial_{\mu}\delta g_{\nu\sigma}+\partial_{\nu}\delta g_{\mu\sigma}-\partial_{\sigma}\delta g_{\mu\nu})-\frac{1}{2}\bar{g}^{\rho\alpha}\bar{g}^{\sigma\beta}\delta g_{\alpha\beta}(\partial_{\mu}\bar{g}_{\nu\sigma}+\partial_{\nu}\bar{g}_{\mu\sigma}-\partial_{\sigma}\bar{g}_{\mu\nu}) \tag{5.156}$$

[1]The minus sign in the second term is due to our "bad" definition: $-\delta g^{\mu\nu}=g^{\mu\nu}-\bar{g}^{\mu\nu}$.

$$\delta G_{\mu\nu} = \delta R_{\mu\nu} - \frac{1}{2}\bar{g}_{\mu\nu}\delta R - \frac{3}{a^2}(\partial_0(\frac{\dot{a}}{a}) + \frac{\dot{a}^2}{a^2})\delta g_{\mu\nu}. \quad (5.157)$$

We will work in the Conformal-Newtonian gauge $E = B = 0$ in which the metric takes the form

$$ds^2 = -(1+2\Phi)dt^2 + a^2(1-2\Psi)\delta_{ij}dx^i dx^j = a^2\bigg[-(1+2\Phi)d\eta^2 + (1-2\Psi)\delta_{ij}dx^i dx^j\bigg]. \quad (5.158)$$

We have

$$\delta g_{00} = -2a^2\Phi \ , \ \delta g_{ij} = -2a^2\Psi\delta_{ij}. \quad (5.159)$$

$$\delta g^{00} = -\frac{2}{a^2}\Phi \ , \ \delta g^{ij} = -\frac{2}{a^2}\Psi\delta_{ij}. \quad (5.160)$$

We compute immediately

$$\delta\Gamma^0_{\mu\nu} = -\frac{1}{2a^2}(\partial_\mu\delta g_{\nu 0} + \partial_\nu\delta g_{\mu 0} - \partial_0\delta g_{\mu\nu}) - \frac{1}{2a^4}\delta g_{00}(\partial_\mu\bar{g}_{\nu 0} + \partial_\nu\bar{g}_{\mu 0} - \partial_0\bar{g}_{\mu\nu}). \quad (5.161)$$

$$\delta\Gamma^i_{\mu\nu} = \frac{1}{2a^2}(\partial_\mu\delta g_{\nu i} + \partial_\nu\delta g_{\mu i} - \partial_i\delta g_{\mu\nu}) - \frac{1}{2a^4}\delta g_{ik}(\partial_\mu\bar{g}_{\nu k} + \partial_\nu\bar{g}_{\mu k} - \partial_k\bar{g}_{\mu\nu}). \quad (5.162)$$

Step 1: We set $a = 1$. In this case we compute

$$\delta\Gamma^0_{00} = \partial_0\Phi \ , \ \delta\Gamma^0_{0i} = \partial_i\Phi \ , \ \delta\Gamma^0_{ij} = -\partial_0\Psi\delta_{ij} \quad (5.163)$$

$$\delta\Gamma^i_{00} = \partial_i\Phi \ , \ \delta\Gamma^i_{0j} = -\partial_0\Psi\delta_{ij} \ , \ \delta\Gamma^i_{jl} = \partial_i\Psi\delta_{jl} - \partial_j\Psi\delta_{li} - \partial_l\Psi\delta_{ij}. \quad (5.164)$$

Thus

$$\delta R_{00} = \partial_i^2\Phi + 3\partial_0^2\Psi. \quad (5.165)$$

$$\delta R_{0i} = 2\partial_0\partial_i\Psi. \quad (5.166)$$

$$\delta R_{ij} = \delta_{ij}\bigg(-\partial_0^2\Psi + \partial_k^2\Psi\bigg) - \partial_i\partial_j(\Phi - \Psi). \quad (5.167)$$

And

$$\delta R = -\delta R_{00} + \delta R_{ii}. \quad (5.168)$$

Thus

$$\delta G_{00} = \frac{1}{2}(\delta R_{00} + \delta R_{ii}) = 2\partial_i^2\Psi. \quad (5.169)$$

$$\delta G_{0i} = \delta R_{0i} = 2\partial_i\partial_0\Psi. \quad (5.170)$$

$$\delta G_{ij} = \delta R_{ij} + \frac{1}{2}\delta_{ij}(\delta R_{00} - \delta R_{kk}) = 2\delta_{ij}\big(\partial_0^2\Psi + \frac{1}{2}\partial_i^2(\Phi - \Psi)\big) - \partial_i\partial_j(\Phi - \Psi). \quad (5.171)$$

Step 2**:** The next step we perform the conformal transformation

$$g_{\mu\nu} \longrightarrow \tilde{g}_{\mu\nu} = F g_{\mu\nu} \ , \ F = a^2. \tag{5.172}$$

Under this transformation we have

$$\Gamma^{\rho}_{\mu\nu} \longrightarrow \tilde{\Gamma}^{\rho}_{\mu\nu} = \Gamma^{\rho}_{\mu\nu} + \frac{1}{2}(\partial_{\mu} \ln F.g^{\rho}_{\nu} + \partial_{\nu} \ln F.g^{\rho}_{\mu} - \partial^{\rho} \ln F.g_{\mu\nu}). \tag{5.173}$$

Also (by using also the fact that the metric is covariantly constant)

$$R_{\mu\nu} \longrightarrow \tilde{R}_{\mu\nu} = R_{\mu\nu} - \frac{1}{F}\nabla_{\mu}\nabla_{\nu}F - \frac{1}{2F}g_{\mu\nu}\nabla_{\alpha}\nabla^{\alpha}F + \frac{3}{2F^2}\nabla_{\mu}F\nabla_{\nu}F. \tag{5.174}$$

Thus

$$R \longrightarrow \tilde{R} = \frac{1}{F}R - \frac{3}{F^2}\nabla_{\mu}\nabla^{\mu}F + \frac{3}{2F^3}\nabla_{\mu}F\nabla^{\mu}F. \tag{5.175}$$

$$G_{\mu\nu} \longrightarrow \tilde{G}_{\mu\nu} = G_{\mu\nu} - \frac{1}{F}\nabla_{\mu}\nabla_{\nu}F + \frac{1}{F}g_{\mu\nu}\nabla_{\alpha}\nabla^{\alpha}F + \frac{3}{2F^2}\nabla_{\mu}F\nabla_{\nu}F - \frac{3}{4F^2}g_{\mu\nu}\nabla_{\alpha}F\nabla^{\alpha}F. \tag{5.176}$$

For our case we need

$$\delta\bigg(-\frac{1}{F}\nabla_{\mu}\nabla_{\nu}F\bigg) = 2\frac{\dot{a}}{a}\delta\Gamma^{0}_{\mu\nu}. \tag{5.177}$$

$$\begin{aligned}
\delta\bigg(\frac{1}{F}g_{\mu\nu}\nabla_{\alpha}\nabla^{\alpha}F\bigg) &= \delta\bigg(\frac{1}{F}g_{\mu\nu}g^{\alpha\beta}\nabla_{\alpha}\partial_{\beta}F\bigg) \\
&= \delta\bigg(\frac{1}{F}g_{\mu\nu}g^{\alpha\beta}\big(\partial_{\alpha}\partial_{\beta}F - \Gamma^{\rho}_{\alpha\beta}\partial_{\rho}F\big)\bigg) \\
&= \bigg(-2\frac{\ddot{a}}{a} - 2\frac{\dot{a}^2}{a^2}\bigg)\delta g_{\mu\nu} + \bigg(4\frac{\ddot{a}}{a}\Phi + 4\frac{\dot{a}^2}{a^2}\Phi + 2\frac{\dot{a}}{a}\partial_0\Phi + 6\frac{\dot{a}}{a}\partial_0\Psi\bigg)\bar{g}_{\mu\nu}.
\end{aligned} \tag{5.178}$$

$$\delta\bigg(-\frac{3}{4F^2}g_{\mu\nu}\nabla_{\alpha}F\nabla^{\alpha}F\bigg) = 3\frac{\dot{a}^2}{a^2}\delta g_{\mu\nu} - 6\frac{\dot{a}^2}{a^2}\Phi\bar{g}_{\mu\nu}. \tag{5.179}$$

$$\delta\tilde{G}_{\mu\nu} = \delta G_{\mu\nu} + 2\frac{\dot{a}}{a}\delta\Gamma^{0}_{\mu\nu} + \bigg(-2\frac{\ddot{a}}{a} + \frac{\dot{a}^2}{a^2}\bigg)\delta g_{\mu\nu} + \bigg(4\frac{\ddot{a}}{a}\Phi - 2\frac{\dot{a}^2}{a^2}\Phi + 2\frac{\dot{a}}{a}\partial_0\Phi + 6\frac{\dot{a}}{a}\partial_0\Psi\bigg)\bar{g}_{\mu\nu}. \tag{5.180}$$

Explicitly we have

$$\delta\tilde{G}_{00} = \delta G_{00} - 6\frac{\dot{a}}{a}\partial_0\Psi. \tag{5.181}$$

$$\delta\tilde{G}_{0i}=\delta G_{0i}+2\frac{\dot{a}}{a}\partial_i\Phi. \tag{5.182}$$

$$\delta\tilde{G}_{ij}=\delta G_{ij}+\bigg(4\frac{\dot{a}}{a}\partial_0\Psi+2\frac{\dot{a}}{a}\partial_0\Phi+\big(4\frac{\ddot{a}}{a}-2\frac{\dot{a}^2}{a^2}\big)\Phi+\big(4\frac{\ddot{a}}{a}-2\frac{\dot{a}^2}{a^2}\big)\Psi\bigg)\delta_{ij}. \tag{5.183}$$

We rewrite these as

$$-a^2\delta\tilde{G}^0_0=\delta G_{00}-6\frac{\dot{a}^2}{a^2}\Phi=2\partial_i^2\Psi-6\frac{\dot{a}}{a}\partial_0\Psi-6\frac{\dot{a}^2}{a^2}\Phi. \tag{5.184}$$

$$-a^2\delta\tilde{G}^0_i=\delta\tilde{G}_{0i}=2\partial_i\big(\partial_0\Psi+\frac{\dot{a}}{a}\Phi\big). \tag{5.185}$$

$$\begin{aligned}
a^2\delta\tilde{G}^0_0 &= \delta G_{ij}+\big(-4\frac{\ddot{a}}{a}+2\frac{\dot{a}^2}{a^2}\big)\Psi\delta_{ij}\\
&= 2\delta_{ij}\big(\partial_0^2\Psi+\frac{1}{2}\partial_i^2(\Phi-\Psi)\big)-\partial_i\partial_j(\Phi-\Psi)+\bigg(4\frac{\dot{a}}{a}\partial_0\Psi+2\frac{\dot{a}}{a}\partial_0\Phi+\big(4\frac{\ddot{a}}{a}-2\frac{\dot{a}^2}{a^2}\big)\Phi\bigg)\delta_{ij}\\
&= 2\delta_{ij}\bigg(\partial_0^2\Psi+\frac{1}{2}\partial_i^2(\Phi-\Psi)+2\frac{\dot{a}}{a}\partial_0\Psi+\frac{\dot{a}}{a}\partial_0\Phi+\big(2\partial_0(\frac{\dot{a}}{a})+\frac{\dot{a}^2}{a^2}\big)\Phi\bigg)-\partial_i\partial_j(\Phi-\Psi).
\end{aligned} \tag{5.186}$$

The linearized Einstein's equations are therefore given by

$$\partial_i^2\Psi-3\frac{\dot{a}}{a}\partial_0\Psi-3\frac{\dot{a}^2}{a^2}\Phi=-4\pi Ga^2\delta T^0_0. \tag{5.187}$$

$$\partial_i\big(\partial_0\Psi+\frac{\dot{a}}{a}\Phi\big)=-4\pi Ga^2\delta T^0_i. \tag{5.188}$$

$$\delta_{ij}\bigg(\partial_0^2\Psi+\frac{1}{2}\partial_i^2(\Phi-\Psi)+2\frac{\dot{a}}{a}\partial_0\Psi+\frac{\dot{a}}{a}\partial_0\Phi+\big(2\partial_0(\frac{\dot{a}}{a})+\frac{\dot{a}^2}{a^2}\big)\Phi\bigg)-\frac{1}{2}\partial_i\partial_j(\Phi-\Psi)=4\pi Ga^2\delta T^i_j. \tag{5.189}$$

The gauge invariant objects are obtained by replacing Φ and Ψ by the Bardeen potentials Φ_B and Ψ_B respectively.

5.4.4 Matter Perturbations

Now we discuss matter perturbations. The stress-energy-momentum tensor $T^{\mu\nu}$ of a perfect fluid is given by

$$T^{\mu\nu}=(\rho+P)U^\mu U^\nu+Pg^{\mu\nu}\ ,\ g_{\mu\nu}U^\mu U^\nu=-1. \tag{5.190}$$

Again we will work with the conformal time denoted by 0 for simplicity. The unperturbed velocity satisfies $\bar{g}_{\mu\nu}\bar{U}^\mu\bar{U}^\nu=-1$ and thus $\bar{U}^\mu=(1/a,0,0,0)$. We compute then from $2\bar{g}_{\mu\nu}\bar{U}^\mu\delta U^\nu+\delta g_{\mu\nu}\bar{U}^\mu\bar{U}^\nu=0$ the result $\delta U^0=\delta U_0/a^2=\delta g_{00}/2a^3$ while δU^i is an independent dynamical variable. We have then $U^0=(1-\Phi)/a$, $U_0=-a-a\Phi$ and $U^i=g^{i\mu}U_\mu\Rightarrow\delta U^i=\delta U_i/a^2-B_i/a$. We will use the notation $\delta U_i=av_i$ and thus $\delta U^i=(v_i-B_i)/a$. The first order perturbation of the stress-energy-momentum tensor is

$$\delta T_{\mu\nu}=(\delta\rho+\delta P)\bar{U}_\mu\bar{U}_\nu+(\bar{\rho}+\bar{P})\delta U_\mu\bar{U}_\nu+(\bar{\rho}+\bar{P})\bar{U}_\mu\delta U_\nu+\delta P\bar{g}_{\mu\nu}+\bar{P}\delta g_{\mu\nu}. \tag{5.191}$$

Explicitly we have (using $\delta g_\mu{}^\nu=0$)

$$\delta T_{00}=a^2(\delta\rho+2\Phi\bar{\rho})\Leftrightarrow\delta T_0{}^0=-\delta\rho. \tag{5.192}$$

$$\delta T_{i0}=-a^2(\bar{\rho}+\bar{P})v_i+a^2\bar{P}B_i\Leftrightarrow\delta T_i^0=(\bar{\rho}+\bar{P})v_i\ ,\ \delta T_0^i=-(\bar{\rho}+\bar{P})(v_i-B_i). \tag{5.193}$$

$$\delta T_{ij}=a^2\delta_{ij}\delta P+\bar{P}\delta g_{ij}\Leftrightarrow\delta T_i^j=\delta_{ij}\delta P. \tag{5.194}$$

There is an extra contribution to the stress-energy-momentum tensor $T_{\mu\nu}$ which is the anisotropic stress tensor $\Sigma_{\mu\nu}$ which vanishes in the unperturbed theory. This tensor is therefore a first order perturbation which is constrained to satisfy $\Sigma_{\mu\nu}U^\nu=0$ and $\Sigma_\mu{}^\mu=0$ and as a consequence $\Sigma_{00}=\Sigma_{i0}=0$ and $\Sigma_i{}^i=0$. The anisotropic stress tensor is therefore a traceless symmetric $3-$tensor Σ_{ij}. In other words we need to change equation (5.194) as follows

$$\delta T_{ij}=a^2\delta_{ij}\delta P+\bar{P}\delta g_{ij}+\Sigma_{ij}\Leftrightarrow\delta T_i^j=\delta_i^j\delta P+\Sigma_i{}^j. \tag{5.195}$$

It is obvious that the tensor $\delta T_{\mu\nu}$ must transform under gauge transformations in the same way as the tensor $\delta g_{\mu\nu}$. These have been already computed in (5.148), (5.149) and (5.149). In conformal time we need to make the replacements $\bar{X}_0^0\longrightarrow\bar{X}_0^0$,$\bar{X}_i^0\longrightarrow\bar{X}_i^0/a$,$\bar{X}_j^i\longrightarrow\bar{X}_j^i$ where 0 stands now for conformal time. The gauge invariant quantities are given by

$$\Delta\delta\hat{X}_0^0=0\ ,\ \delta\hat{X}_0^0\ =\ \delta X_0^0+(B-E^{'})(\bar{X}_0^0)^{'}. \tag{5.196}$$

$$\Delta\delta\hat{X}_i^0=0\ ,\ \delta\hat{X}_i^0\ =\ \delta X_i^0+\partial_i(B-E^{'})(\bar{X}_0^0-\frac{1}{3}\bar{X}_k^k). \tag{5.197}$$

$$\Delta\delta\hat{X}_j^i=0\ ,\ \delta\hat{X}_j^i\ =\ \delta X_j^i+(B-E^{'})(\bar{X}_j^i)^{'}. \tag{5.198}$$

The gauge invariant linearized Einstein's equations becomes given by

$$\partial_i^2\Psi_B-3\mathcal{H}\Psi_B^{'}-3\mathcal{H}^2\Phi_B=4\pi Ga^2\delta\hat{\rho}. \tag{5.199}$$

$$\partial_i\big(\Psi_B^{'}+\mathcal{H}\Phi_B\big)=-4\pi Ga^2(\bar{\rho}+\bar{P})\frac{\delta\hat{U}_i}{a}. \quad (5.200)$$

$$\delta_{ij}\Bigg(\Psi_B^{''}+\frac{1}{2}\partial_i^2(\Phi_B-\Psi_B)+2\mathcal{H}\Psi_B^{'}+\mathcal{H}\Phi_B^{'}+\big(2\mathcal{H}^{'}+\mathcal{H}^2\big)\Phi_B\Bigg)-\frac{1}{2}\partial_i\partial_j(\Phi_B-\Psi_B)=4\pi Ga^2\delta\hat{P}\delta_{ij}. \quad (5.201)$$

$$\delta\hat{T}_0^0=-\delta\hat{\rho}=-\delta\rho-\bar{\rho}^{'}(B-E^{'}). \quad (5.202)$$

$$\delta\hat{T}_i^0 = (\bar{\rho}+\bar{P})\Bigg(\frac{\delta U_i}{a}-\partial_i(B-E^{'})\Bigg)=(\bar{\rho}+\bar{P})\frac{\delta\hat{U}_i}{a}. \quad (5.203)$$

$$\delta\hat{T}_j^i = \delta_{ij}\big(\delta P+\bar{P}(B-E^{'})\big)=\delta_{ij}\delta\hat{P}. \quad (5.204)$$

In the above second equation $\delta\hat{U}_i$ is the gauge invariant velocity perturbation. As before only the parallel part of this velocity, which is of the form $a^2\partial_i\gamma$ for some scalar function γ, will contribute to scalar perturbation. Remember that we are neglecting vector perturbations throughout.

5.5 Matter-Radiation Equality

We recall the two Friedmann equations and the energy conservation law

$$\mathcal{H}^2=\frac{8\pi Ga^2}{3}\bar{\rho}\ ,\ \mathcal{H}^2-\mathcal{H}^{'}=4\pi Ga^2(\bar{\rho}+\bar{P})\ ,\ \bar{\rho}^{'}=-3\mathcal{H}(\bar{\rho}+\bar{P}). \quad (5.205)$$

By combining the two Friedmann equations as $\mathcal{H}^2-(\mathcal{H}^2-\mathcal{H}^{'})$ we obtain

$$a^{''}=\frac{4\pi Ga^3}{3}(\bar{\rho}-3\bar{P}). \quad (5.206)$$

We have already shown that the density of radiation falls off as $1/a^4$ whereas the density of matter falls off as $1/a^3$ and thus in a universe filled with matter and radiation we have the energy density

$$\begin{aligned}\bar{\rho} &= \bar{\rho}_m+\bar{\rho}_r\\ &= \frac{\bar{\rho}_{\rm eq}}{2}\big(\frac{a_{\rm eq}^3}{a^3}+\frac{a_{\rm eq}^4}{a^4}\big).\end{aligned} \quad (5.207)$$

The $\rho_{\rm eq}$ is the energy density at the time of equality $\eta_{\rm eq}$ at which matter and radiation densities become equal and $a_{\rm eq}=a(\eta_{\rm eq})$ is the corresponding scale factor. We also recall that the pressure of matter (dust) is 0 whereas the pressure of radiation is $\bar{P}_r=\bar{\rho}_r/3$ and thus

$$\bar{P} = \bar{P}_m+\bar{P}_r=\frac{\bar{\rho}_r}{3}. \quad (5.208)$$

By using these last two equations in the Friedmann equation (5.206) we obtain

$$a^{''}=\frac{2\pi Ga_{\rm eq}^3\bar{\rho}_{\rm eq}}{3}=2C_0. \tag{5.209}$$

The solution is immediately given by

$$a=C_0\eta^2+C_1\eta+C_2. \tag{5.210}$$

We find $C_2=0$ from the boundary condition $a(0)=0$. By substituting this solution in the Friedmann equation $\mathcal{H}^2=8\pi Ga^2\bar{\rho}/3$ or equivalently

$$a^{'2}=\frac{4\pi G}{3}\bar{\rho}_{\rm eq}(a_{\rm eq}^3a+a_{\rm eq}^4), \tag{5.211}$$

we obtain $C_1=\sqrt{4C_0a_{\rm eq}}$. The scale factor is therefore given by

$$a=a_{\rm eq}(\frac{\eta^2}{\eta_*^2}+2\frac{\eta}{\eta_*}). \tag{5.212}$$

The time η_* is related to the time of equality $\eta_{\rm eq}$ by $\eta_{\rm eq}=(\sqrt{2}-1)\eta_*$. In the radiation dominated universe corresponding to $\eta<<\eta_{\rm eq}$ we have $a\propto\eta$ whereas in the matter dominated universe corresponding to $\eta>>\eta_{\rm eq}$ we have $a\propto\eta^2$.

5.6 Hydrodynamical Adiabatic Scalar Perturbations

The Einstein's equation (5.201) for $i\neq j$ gives $\partial_i\partial_j(\Phi-\Psi)=0$. The only solutions consistent with Φ and Ψ being perturbations are $\Phi=\Psi$. The remaining Einstein's equations simplify therefore to

$$\partial_i^2\Phi_B-3\mathcal{H}\Phi_B^{'}-3\mathcal{H}^2\Phi_B=4\pi Ga^2\delta\hat{\rho}. \tag{5.213}$$

$$\partial_i(a\Phi_B)^{'}=-4\pi Ga^2(\bar{\rho}+\bar{P})\delta\hat{U}_i. \tag{5.214}$$

$$\Phi_B^{''}+3\mathcal{H}\Phi_B^{'}+\big(2\mathcal{H}^{'}+\mathcal{H}^2\big)\Phi_B=4\pi Ga^2\delta\hat{P}. \tag{5.215}$$

The first equation is the generalization of Poisson's equation for the Newtonian gravitational potential which is identified here with the Bardeen potential Φ_B. Recall that the sub-Hubble or sub-horizon scales correspond to comoving Fourier scales k^{-1} such that $k>\mathcal{H}=aH$. The second and third terms in (5.213) can be rewritten as $-3\mathcal{H}(a\Phi_B)^{'}/a$, i.e. they are suppressed by a factor $1/\mathcal{H}$ on sub-Hubble scales and thus can be neglected compared to the first term. Equation (5.213) reduces therefore to the usual Poisson's equation for the Newtonian gravitational potential in this limit. The combination $(a\Phi_B)^{'}$ is precisely the velocity potential which is given by equation (5.214).

Now we will split the pressure perturbation into an adiabatic (curvature) piece and an entropy (isocurvature) piece as follows

$$\delta\hat{P} = c_s^2\delta\hat{\rho} + \tau\delta S. \tag{5.216}$$

The first component $\delta\hat{P} = c_s^2\delta\hat{\rho}$ is the adiabatic perturbation and it corresponds to fluctuations in the energy density and thus induce inhomogeneities in the spatial curvature. The second component $\delta\hat{P} = \tau\delta S$ is the entropy perturbation and it corresponds to fluctuations in the form of the local equation of state of the system, i.e. fluctuations in the relative number densities of the different particle types present in the system. The two perturbations are orthogonal since any other perturbation can be written as a linear combination of the two. The coefficients c_s^2 and τ are given by

$$c_s^2 = (\frac{\partial P}{\partial \rho})_S \ , \ \tau = (\frac{\partial p}{\partial S})_\rho. \tag{5.217}$$

In particular c_s^2 is the speed of sound as we now show. We combine the two Einstein's equations (5.213) and (5.215) as follows

$$c_s^2\bigg(\partial_i^2\Phi_B - 3\mathcal{H}\Phi_B^{'} - 3\mathcal{H}^2\Phi_B\bigg) - \bigg(\Phi_B^{''} + 3\mathcal{H}\Phi_B^{'} + \big(2\mathcal{H}^{'} + \mathcal{H}^2\big)\Phi_B\bigg) = 4\pi G a^2(c_s^2\delta\hat{\rho} - \delta\hat{P}). \tag{5.218}$$

We get then the general relativistic Poisson's equation for the Newtonian gravitational potential given by

$$\Phi_B^{''} + 3\mathcal{H}(1+c_s^2)\Phi_B^{'} - c_s^2\partial_i^2\Phi_B + \big(2\mathcal{H}^{'} + \mathcal{H}^2(1+3c_s^2)\big)\Phi_B = 4\pi G a^2\tau\delta S. \tag{5.219}$$

Adiabatic Perturbations: We will only concentrate here on adiabatic perturbations. The case of entropy perturbations is treated in the excellent book [2].

In this case we set

$$\delta S = 0. \tag{5.220}$$

Equivalently

$$\begin{aligned} c_s^2 = \frac{\delta\hat{P}}{\delta\hat{\rho}} &= (\frac{\partial P}{\partial\rho})_S \\ &= (\frac{\partial\eta}{\partial\rho}\frac{\partial P}{\partial\eta})_S \\ &= \frac{\bar{P}^{'}}{\bar{\rho}^{'}}. \end{aligned} \tag{5.221}$$

The above general relativistic Poisson's equation equation can be simplified by introducing the variable u defined by

$$\begin{aligned} u &= \exp\left(\frac{3}{2}\int(1+c_s^2)\mathcal{H}d\eta\right)\Phi_B \\ &= \exp\left(-\frac{1}{2}\int(1+\frac{\bar{P}'}{\bar{\rho}'})\frac{\bar{\rho}'}{\bar{\rho}+\bar{P}}d\eta\right)\Phi_B \\ &= \frac{1}{\sqrt{\bar{\rho}+\bar{P}}}\Phi_B. \end{aligned} \tag{5.222}$$

We rewrite this as

$$u = \frac{a}{\sqrt{\bar{\rho}}}\theta\Phi_B \ , \ \theta = \frac{1}{a\sqrt{1+\frac{\bar{P}}{\bar{\rho}}}}. \tag{5.223}$$

We remark that

$$\theta = \frac{1}{a\sqrt{\frac{2}{3}(1-\frac{\mathcal{H}'}{\mathcal{H}^2})}}. \tag{5.224}$$

We compute immediately

$$\Phi_B'' + 3\mathcal{H}(1+c_s^2)\Phi_B' = \exp\big(-\tfrac{3}{2}\textstyle\int(1+c_s^2)\mathcal{H}d\eta\big)\Big[u'' - \big[-\tfrac{3}{2}(1+c_s^2)\mathcal{H}\big]^2 u + \big[-\tfrac{3}{2}(1+c_s^2)\mathcal{H}\big]' u\Big]. \tag{5.225}$$

We observe that the friction term cancels exactly. Also

$$\big(2\mathcal{H}' + \mathcal{H}^2(1+3c_s^2)\big)\Phi_B = \exp\big(-\frac{3}{2}\int(1+c_s^2)\mathcal{H}d\eta\big)\Big[-\frac{3\mathcal{H}^2}{a^2\theta^2} + 3(1+c_s^2)\mathcal{H}^2\Big]u. \tag{5.226}$$

We use

$$1-\frac{\mathcal{H}'}{\mathcal{H}^2} = \frac{3}{2a^2\theta^2}. \tag{5.227}$$

$$1+c_s^2 = \frac{1}{a^2\theta^2} + \frac{2}{3} + \frac{2}{3\mathcal{H}}\frac{\theta'}{\theta}. \tag{5.228}$$

Thus

$$\big(2\mathcal{H}' + \mathcal{H}^2(1+3c_s^2)\big)\Phi_B = \exp\big(-\frac{3}{2}\int(1+c_s^2)\mathcal{H}d\eta\big)\Big[2\mathcal{H}^2 + 2\mathcal{H}\frac{\theta'}{\theta}\Big]u. \tag{5.229}$$

After some calculation we get

$$\Big(2\mathcal{H}^{'}+\mathcal{H}^2(1+3c_s^2)+\big(2\mathcal{H}^{'}+\mathcal{H}^2(1+3c_s^2)\big)\Big)\Phi_B = \exp\big(-\frac{3}{2}\int(1+c_s^2)\mathcal{H}d\eta\big) \times \Big[u^{''}+\Big(-\frac{3\mathcal{H}^{'}}{2a^2\theta^2}-\mathcal{H}^{'}+\mathcal{H}^2-\frac{9\mathcal{H}^2}{4a^4\theta^4}-\frac{\theta^{''}}{\theta}\Big)u\Big]. \tag{5.230}$$

After some inspection we get

$$\Big(2\mathcal{H}^{'}+\mathcal{H}^2(1+3c_s^2)+\big(2\mathcal{H}^{'}+\mathcal{H}^2(1+3c_s^2)\big)\Big)\Phi_B = \exp\big(-\frac{3}{2}\int(1+c_s^2)\mathcal{H}d\eta\big) \times \Big[u^{''}-\frac{\theta^{''}}{\theta}u\Big]. \tag{5.231}$$

Poisson's equation reduces therefore to

$$u^{''}-c_s^2\partial_i^2u-\frac{\theta^{''}}{\theta}u=0. \tag{5.232}$$

We look for plane wave solutions of the form

$$u=u_{\vec{k}}(\vec{x},\eta)=\exp(i\vec{k}\vec{x})\chi_{\vec{k}}(\eta). \tag{5.233}$$

We need to solve

$$\chi_{\vec{k}}^{''}+(c_s^2\vec{k}^2-\frac{\theta^{''}}{\theta})\chi_{\vec{k}}=0. \tag{5.234}$$

Let us first assume that $\theta^{''}/\theta$ is a constant, viz

$$\frac{\theta^{''}}{\theta}=\sigma^2. \tag{5.235}$$

The above differential equation becomes

$$\chi_{\vec{k}}^{''}+\omega_{\vec{k}}^2\chi_{\vec{k}}=0\ ,\ \omega_{\vec{k}}=\sqrt{c_s^2\vec{k}^2-\sigma^2}. \tag{5.236}$$

We define the so-called Jeans length by

$$\lambda_J=\frac{2\pi}{k_J}\ ,\ k_J=\frac{\sigma}{c_s}. \tag{5.237}$$

In other words,

$$\omega_{\vec{k}}=c_s\sqrt{\vec{k}^2-\vec{k}_J^2}. \tag{5.238}$$

The behavior of the perturbation depends therefore crucially on its spatial size given by the Jeans length. Two interesting limiting cases emerge immediately:

- Large scales corresponding to long-wavelengths where gravity dominates given by $k << k_J$, $\lambda >> \lambda_J$: In this case we get the solutions

$$\chi_{\vec{k}} \sim \exp(\pm|\omega_{\vec{k}}|\eta). \tag{5.239}$$

  The plus sign describes exponentially fast growth of inhomogeneities whereas the negative sign describes a decaying solution. We have when $k \longrightarrow 0$ the behavior

$$|\omega_{\vec{k}}|\eta \longrightarrow c_s k_J \eta = \frac{\eta}{\eta_{\rm gr}} \ , \ \eta_{\rm gr} = \frac{1}{\sigma}. \tag{5.240}$$

  From this we can deduce that gravity is very efficient in amplifying adiabatic perturbations. As an example, if the initial adiabatic perturbation is extremely small of the order of 10^{-100}, gravity will only need $\eta = 230\eta_{\rm gr}$ to amplify it to order 1.

  We remark that this limit $k << k_J$ corresponds to $c_s k\eta << \eta/\eta_{\rm gr} = \lambda/\lambda_J$ where the Jeans length $\lambda_J = c_s t_{\rm gr}$ is the sound communication scale, i.e. the scale over which pressure can react to changes in the energy density due to gravity. Thus this limit can be characterized simply by $c_s k\eta << 1$.

- Small scales corresponding to short-wavelengths where gravity is negligible compared to pressure given by $k >> k_J$, $\lambda << \lambda_J$: In this case we get the solutions

$$\chi_{\vec{k}} \sim \exp(\pm i\omega_{\vec{k}}\eta). \tag{5.241}$$

  These are sound waves with phase velocity given by

$$c_{\rm phase} = \frac{\omega_{\vec{k}}}{k} = c_s\sqrt{1-\frac{k_J^2}{k^2}} \longrightarrow c_s. \tag{5.242}$$

We solve now the differential equation (5.234) more rigorously in these two limiting cases.

Large scales or long-wavelengths ($c_s k\eta << 1$): In this case we can neglect the spatial derivative in (5.234) and the equation reduces to

$$\chi_{\vec{k}}^{''} - \frac{\theta^{''}}{\theta}\chi_{\vec{k}} = 0. \tag{5.243}$$

The first solution is obviously $\chi_{\vec{k}} = C_1\theta$. The second linearly independent solution is

$$\chi_{\vec{k}} = C_2\theta\int_{\eta_0}^{\eta}\frac{d\eta^{'}}{\theta^2(\eta^{'})}. \tag{5.244}$$

This can be checked using the Wronskian. The most general solution is a linear combination which is also of the above form (5.244) with a different η_0. It is straightforward to compute

$$\int_{\eta_0}^{\eta}\frac{d\eta^{'}}{\theta^2(\eta^{'})} = \frac{2}{3}\bigg(\frac{a^2}{\mathcal{H}}-\int a^2 d\eta\bigg). \tag{5.245}$$

The gravitational potential is therefore given by

$$\begin{aligned}\Phi_B &= \frac{\sqrt{\bar{\rho}}}{a\theta}u\\ &= \sqrt{\bar{\rho}+\bar{P}}\exp(i\vec{k}\vec{x})\chi_{\vec{k}}\\ &= C_2\exp(i\vec{k}\vec{x})\frac{\sqrt{\bar{\rho}}}{a}\int_{\eta_0}^{\eta}\frac{d\eta^{'}}{\theta^2(\eta^{'})}\\ &= \frac{2}{3}C_2\sqrt{\frac{3}{8\pi G}}\exp(i\vec{k}\vec{x})\bigg(1-\frac{\mathcal{H}}{a^2}\int a^2 d\eta\bigg)\\ &= \frac{2}{3}C_2\sqrt{\frac{3}{8\pi G}}\exp(i\vec{k}\vec{x})\bigg(1-\frac{\dot{a}}{a^2}\int a dt\bigg)\\ &= \frac{2}{3}C_2\sqrt{\frac{3}{8\pi G}}\exp(i\vec{k}\vec{x})\frac{d}{dt}\bigg(\frac{1}{a}\int a dt\bigg).\end{aligned} \tag{5.246}$$

Since we are interested in long-wavelengths, i.e. $k\longrightarrow 0$, we can set the plane wave equal 1. The result is then

$$\Phi_B = A\frac{d}{dt}\bigg(\frac{1}{a}\int a dt\bigg). \tag{5.247}$$

We assume now that the universe is a mixture of radiation and matter in the form of say cold baryons. The scale factor is then given by (5.248), viz

$$a=a_{\rm eq}(\frac{\eta^2}{\eta_*^2}+2\frac{\eta}{\eta_*}). \tag{5.248}$$

We compute immediately (with $\xi=\eta/\eta_*$)

$$\begin{aligned}\Phi_B &= \frac{A}{\xi(\xi+2)}\frac{d}{d\xi}\bigg(\frac{1}{\xi+2}\bigg(\frac{1}{5}\xi^4+\xi^3+\frac{4}{3}\xi^2\bigg)+\frac{A^{'}}{\xi(\xi+2)}\bigg)\\ &= \frac{A(\xi+1)}{(\xi+2)^3}\bigg(\frac{3}{5}\xi^2+3\xi+\frac{13}{3}+\frac{1}{\xi+1}\bigg)+\frac{B(\xi+1)}{\xi^3(\xi+2)^3}.\end{aligned} \tag{5.249}$$

The A term is the term corresponding to the growth of inhomogeneities whereas the B term is the decaying mode which we can neglect.

By using the Friedmann equation $\mathcal{H}^2=8\pi Ga^2\bar{\rho}/3$ and the Einstein equation (5.213) we obtain an expression for the energy density perturbation given by

$$\frac{\delta\hat{\rho}}{\bar{\rho}}=-2\Phi_B-\frac{2}{\mathcal{H}}\Phi_B^{'}+\frac{2}{3\mathcal{H}^2}\partial_i^2\Phi_B. \tag{5.250}$$

We use the results

$$\mathcal{H} = \frac{2(\xi+1)}{\eta_*\xi(\xi+2)} \ , \ \frac{d\Phi_B}{d\xi} = -\frac{4A(\xi+5)}{15(\xi+2)^4}. \tag{5.251}$$

Thus

$$\frac{\delta\hat{\rho}}{\bar{\rho}} = -2\Phi_B + \frac{4A\xi(\xi+5)}{15(\xi+1)(\xi+2)^3} - \frac{\vec{k}^2\eta^2(\xi+2)^2}{6(\xi+1)^2}\Phi_B. \tag{5.252}$$

The last term is of course negligible for long-wavelengths $k \longrightarrow 0$. At early times compared to $\eta_{\rm eq} \sim \eta_*$ we have $\xi \longrightarrow 0$ and $\Phi_B \longrightarrow 2A/3$, $\delta\hat{\rho}/\bar{\rho} \longrightarrow -4A/3$, whereas at late times compared to $\eta_{\rm eq} \sim \eta_*$ we have $\xi \longrightarrow \infty$ and $\Phi_B \longrightarrow 3A/5$, $\delta\hat{\rho}/\bar{\rho} \longrightarrow -6A/5$. Thus Φ_B and $\delta\hat{\rho}/\bar{\rho}$ are both constants during radiation-dominated (early times) and matter-dominated (late times) epochs with the amplitude decreasing by a factor of 9/10 at the time of radiation-matter equality. In the matter dominated epoch the gravitational potential remains always a constant whereas the energy density fluctuation starts to increase as η^2 at the time of horizon crossing around $\eta \sim k^{-1}$.

Small scales or short-wavelengths ($c_s k\eta >> 1$): In this case we can neglect the last term (gravity effect) in (5.234) and the equation reduces to

$$\chi_{\vec{k}}^{''} + c_s^2\vec{k}^2\chi_{\vec{k}} = 0. \tag{5.253}$$

This is a wave equation for sound perturbations with time-dependent amplitude which can be solved explicitly in the WKB approximation for slowly varying speed of sound.

5.7 Quantum Cosmological Scalar Perturbations

5.7.1 Slow-Roll Revisited

We consider a flat universe filled with a scalar field ϕ with an action

$$S = \int \sqrt{-g}d^4x\mathcal{P}(X,\phi) \ , \ X = -\frac{1}{2}g^{\alpha\beta}\nabla_\alpha\phi\nabla_\beta\phi. \tag{5.254}$$

A canonical scalar field is given by

$$\mathcal{P}(X,\phi) = X - V(\phi). \tag{5.255}$$

The energy-momentum tensor is defined by

$$\begin{aligned} T_{\mu\nu} &= -\frac{2}{\sqrt{-g}}\frac{\delta S}{\delta g^{\mu\nu}} \\ &= 2X\frac{\partial\mathcal{P}}{\partial X}u_\mu u_\nu + \mathcal{P}g_{\mu\nu} \ , \ u_\mu = -\frac{1}{\sqrt{2X}}\nabla_\mu\phi. \end{aligned} \tag{5.256}$$

We observe that $g^{\mu\nu}u_\mu u_\nu=-1$. Since $T_{00}=\rho a^2$ we deduce

$$\rho = 2X\frac{\partial\mathcal{P}}{\partial X}-\mathcal{P}. \tag{5.257}$$

Thus

$$T_{\mu\nu} = (\rho+\mathcal{P})u_\mu u_\nu+\mathcal{P}g_{\mu\nu}. \tag{5.258}$$

In other words, $\mathcal{P}$ plays the role of pressure.

The unperturbed system consists of the usual scale factor $a(\eta)$ and a homogeneous field $\phi_0(\eta)$. The equations of motion of the scale factor are the Friedmann equations

$$\mathcal{H}^2=\frac{8\pi Ga^2}{3}\rho\ ,\ \mathcal{H}^2-\mathcal{H}^{'}=4\pi Ga^2(\rho+\mathcal{P}). \tag{5.259}$$

Also we note the continuity equation

$$\rho^{'}=-3\mathcal{H}(\rho+\mathcal{P})=\frac{\partial\rho}{\partial\phi}\phi_0^{'}+\frac{\partial\rho}{\partial X}X_0^{'}. \tag{5.260}$$

The equation of motion of a canonical scalar field ϕ is given by

$$\begin{aligned}\frac{\delta S}{\delta\phi} &= \frac{1}{\sqrt{-g}}\partial_\alpha\big(\sqrt{-g}g^{\alpha\beta}\frac{\partial\mathcal{P}}{\partial X}\partial_\beta\phi\big)+\frac{\partial\mathcal{P}}{\partial\phi}\\ &= \frac{1}{\sqrt{-g}}\partial_\alpha\big(\sqrt{-g}g^{\alpha\beta}\partial_\beta\phi\big)-\frac{\partial V}{\partial\phi}\\ &= 0.\end{aligned} \tag{5.261}$$

For the background ϕ_0 this reads explicitly

$$\phi_0^{''}+2\mathcal{H}\phi_0^{'}+a^2\frac{\partial V}{\partial\phi}=0\ ,\ \ddot{\phi}_0+3H\dot{\phi}_0+\frac{\partial V}{\partial\phi}=0. \tag{5.262}$$

We consider scalar perturbation of the form

$$\phi=\phi_0+\delta\phi. \tag{5.263}$$

The gauge transformation of the scalar perturbation is computed as follows

$$\begin{aligned}\Delta\delta\phi &= \phi^{'}(x^{'}-\epsilon)-\phi(x)\\ &= -\epsilon^\lambda\frac{\partial}{\partial x^\lambda}\phi_0\\ &= -\alpha\dot{\phi}_0.\end{aligned} \tag{5.264}$$

Thus the gauge invariant scalar perturbation is given by

$$\delta\hat{\phi}=\delta\phi-(E^{'}-B){\phi_0}'\ ,\ \Delta\delta\hat{\phi}=0. \tag{5.265}$$

The above scalar perturbation induces scalar metric perturbation of the form

$$ds^2 = a^2\bigg(- (1+2\Phi)d\eta^2 + 2a\partial_i B d\eta dx^i + a^2[2\partial_i\partial_j E + (1-2\Psi)\delta_{ij}]dx^i dx^j\bigg). \tag{5.266}$$

Again we will work in the longitudinal (conformal-Newtonian) gauge, viz

$$ds^2 = a^2\big(- (1+2\Phi)d\eta^2 + (1-2\Psi)\delta_{ij}dx^i dx^j\big). \tag{5.267}$$

To linear order the equation of motion of the scalar field perturbation $\delta\phi$ reads

$$\partial_\alpha\bigg(- a^4\delta g^{\alpha\beta}\partial_\beta\phi_0 + a^4(\Phi-3\Psi)\bar{g}^{\alpha\beta}\partial_\beta\phi_0 + a^4\bar{g}^{\alpha\beta}\partial_\beta\delta\phi\bigg) - (\Phi-3\Psi)\partial_\alpha(a^4\bar{g}^{\alpha\beta}\partial_\beta\phi_0) - a^4\delta\phi\frac{\partial^2 V}{\partial\phi^2} = 0. \tag{5.268}$$

Or equivalently

$$\delta\phi^{''} + 2\mathcal{H}\delta\phi^{'} - (\Phi^{'}+3\Psi^{'})\phi_0^{'} + 2a^2\Phi\frac{\partial V}{\partial\phi} + a^2\delta\phi\frac{\partial^2 V}{\partial\phi^2} - \partial_i^2\delta\phi = 0. \tag{5.269}$$

The gauge invariant version of this equation is obtained by making the replacements $\delta\phi\longrightarrow\delta\hat{\phi}$, $\Phi\longrightarrow\Phi_B$ and $\Psi\longrightarrow\Psi_B$, viz

$$\delta\hat{\phi}^{''} + 2\mathcal{H}\delta\hat{\phi}^{'} - (\Phi_B^{'}+3\Psi_B^{'})\phi_0^{'} + 2a^2\Phi_B\frac{\partial V}{\partial\phi} + a^2\delta\hat{\phi}\frac{\partial^2 V}{\partial\phi^2} - \partial_i^2\delta\hat{\phi} = 0. \tag{5.270}$$

Small scales or short-wavelengths: This corresponds to wavelengths $\lambda << 1/H$ or equivalently wavenumbers $k >> aH$ where gravity can be neglected. Remember that $1/H$ is the Hubble distance and $1/aH$ is the Hubble length or radius. During inflation since $a\sim\exp(Ht)$ we have $aH = -1/\eta$. Thus this limit corresponds to $k\eta >> 1$. The last term in the above equation therefore dominates and we end up with a solution of the form $\delta\hat{\phi}\sim\exp(\pm ik\eta)$. By using equations (5.257) and (5.306) (see below) we find that the gravitational potential solves the equations

$$\Psi_B^{'} + \mathcal{H}\Phi_B = 4\pi G\phi_0^{'}\delta\hat{\phi}. \tag{5.271}$$

We must also have $\Phi_B = \Psi_B$ (see below). The gravitational potential therefore oscillates as

$$\Psi_B = \Phi_B \sim \frac{4\pi G}{k}\phi_0^{'}\delta\hat{\phi}. \tag{5.272}$$

The third and fourth terms can therefore be neglected. The fifth term can also be neglected since during inflation $\partial^2 V/\partial\phi^2 << V \sim H^2(\eta_V << 1)$. The equation (5.270) reduces therefore with $\delta\hat{\phi} = \exp(i\vec{k}\vec{x})\delta\hat{\phi}_k$ to

$$\delta\hat{\phi}_k^{''} + 2\mathcal{H}\delta\hat{\phi}_k^{'} + \vec{k}^2\delta\hat{\phi}_k = 0. \tag{5.273}$$

In terms of $u_k=a\delta\hat{\phi}_k$ this reads

$$u_k^{''}+(\vec{k}^2-\frac{a^{''}}{a})u_k=0. \quad (5.274)$$

Since $k\eta>>1$ the solution is of the form

$$\delta\hat{\phi}_k\simeq\frac{C_k}{a}\exp(\pm ik\eta). \quad (5.275)$$

We fix the constant of integration C_k by requiring that the initial scalar mode arises as vacuum quantum fluctuation.

The minimal vacuum fluctuations must satisfy Heisenberg uncertainty principle $\Delta X\Delta P\sim 1$. From (5.254) the action of the perturbation $\delta\phi$ starts as

$$S = \int dt\int dV\,[\frac{1}{2}\dot{\delta\phi}^2+...]. \quad (5.276)$$

Obviously $dV=a^3d^3x$. Thus in a finite volume $V=L^3$ the canonical field is $X=L^{3/2}\delta\phi$ while the conjugate field is $P=L^{3/2}\dot{\delta\phi}$. For a massless field we have the estimate $P=L^{1/2}\delta\phi$ and as a consequence the Heisenberg uncertainty principle yields $\Delta\delta\phi=1/L$. In other words, minimal quantum fluctuations of the scalar perturbation are of the order of $1/L$. However, quantum fluctuations of the Fourier mode $\delta\phi_k$ are related to quantum fluctuations of the scalar perturbations $\delta\phi$ by (see below for a derivation)

$$\delta\phi\sim\delta\phi_k k^{3/2}. \quad (5.277)$$

Since $k\sim a/L$ we conclude that $\delta\phi_k\sim L^{1/2}/a^{3/2}$ or equivalently $\delta\phi_k\sim 1/a\sqrt{k}$. Hence

$$\delta\hat{\phi}_k\simeq\frac{1}{\sqrt{k}a}. \quad (5.278)$$

In other words, $C_k=1/\sqrt{k}$. The evolution of the mode in this region is such that the vacuum spectrum is preserved. We observe that the amplitude of fluctuation is such that

$$\delta_\phi\sim\delta\phi_k k^{3/2}=\frac{k}{a}>>H. \quad (5.279)$$

Thus every mode will eventually be stretched to very large scales while new modes will be generated. The moment $\eta_k\sim 1/k$ at which the mode k leaves the horizon is called horizon crossing and is defined by

$$\delta_\phi\sim\delta\phi_k k^{3/2}=\frac{k}{a_k}=H_{k\sim Ha}. \quad (5.280)$$

If this mode was classical it will be completely washed out, i.e. becomes very small, after it is stretched out to galactic scales.

Large scales or long-wavelengths: In the slow-roll approximation the equation of motion (5.262) becomes

$$3H\dot{\phi}_0+\frac{\partial V}{\partial \phi}=0. \tag{5.281}$$

The equations of motion (5.270) and (5.271) in terms of the physical time read

$$\delta\ddot{\hat{\phi}}+3H\delta\dot{\hat{\phi}}-4\dot{\Phi}_B\dot{\phi}_0+2\Phi_B\frac{\partial V}{\partial \phi}+\delta\hat{\phi}\frac{\partial^2 V}{\partial \phi^2}-\frac{1}{a^2}\partial_i^2\delta\hat{\phi}=0. \tag{5.282}$$

$$\dot{\Phi}_B+H\Phi_B=4\pi G\dot{\phi}_0\delta\hat{\phi}. \tag{5.283}$$

For long-wavelengths $k<<aH$ we can drop the Laplacian term. As we will see the terms $\delta\ddot{\hat{\phi}}$ and $\dot{\Phi}_B$ are also negligible in this limit. The equations become

$$3H\delta\dot{\hat{\phi}}+2\Phi_B\frac{\partial V}{\partial \phi}+\delta\hat{\phi}\frac{\partial^2 V}{\partial \phi^2}=0. \tag{5.284}$$

$$H\Phi_B=4\pi G\dot{\phi}_0\delta\hat{\phi}. \tag{5.285}$$

We introduce the variable

$$y=\frac{\delta\hat{\phi}}{\frac{\partial V}{\partial \phi}}=-\frac{\delta\hat{\phi}}{3H\dot{\phi}_0}. \tag{5.286}$$

Thus

$$H\Phi_B=4\pi Gy(-3H\dot{\phi}_0^2)=4\pi Gy\dot{V}. \tag{5.287}$$

Also (by neglecting $\ddot{\phi}_0$ and $\partial^2V/\partial\phi^2$ and $\dot{H}$ during inflation)

$$3H\dot{y}+2\Phi_B=0. \tag{5.288}$$

By using also $3H^2=8\pi GV$ during inflation we have

$$\frac{d}{dt}(yV)=\frac{H}{8\pi G}(3H\dot{y}+2\Phi_B)=0. \tag{5.289}$$

The solutions are immediately given by

$$\delta\phi_k=C_k\frac{1}{V}\frac{\partial V}{\partial \phi}. \tag{5.290}$$

$$\begin{aligned}\Phi_B &= \frac{4\pi GC_k}{H}\frac{\dot{V}}{V}\\ &= -\frac{1}{2}C_k\big(\frac{1}{V}\frac{\partial V}{\partial \phi}\big)^2.\end{aligned} \tag{5.291}$$

We fix the constant of integration C_k by comparing with (5.280) at the instant of horizon crossing. We obtain

$$C_k = \frac{k^{-1/2}}{a_k}\left(\frac{V}{\frac{\partial V}{\partial \phi}}\right)_{k\sim Ha}. \tag{5.292}$$

The solutions (5.275) and (5.290) are sketched on figure 2COS, 1. After horizon crossing the short-wavelengths modes are stretched to galactic scales in such a way that they do not lose their amplitudes. Remember that inside the horizon gravity is negligible. Thus perturbations which are initially inside the horizon, will eventually exit the horizon, and then start feeling the curvature effects of gravity preserving therefore their amplitudes from decay. We say that the perturbation is frozen after horizon crossing. This is how we get the required amplitude $\Phi \sim 10^{-5}$ on large scales from initial quantum fluctuations.

At the end of inflation the slow-roll condition is violated since $V/(\partial V/\partial \phi)$ becomes of order 1 and the amplitude of fluctuations is

$$\begin{aligned}\delta_\phi(k)_{t_f} &\sim C_k k^{3/2}\\ &\sim \left(H\frac{V}{\frac{\partial V}{\partial \phi}}\right)_{k\sim Ha}\\ &\sim \left(\frac{V^{3/2}}{\frac{\partial V}{\partial \phi}}\right)_{k\sim Ha}.\end{aligned} \tag{5.293}$$

This depends only on quantities evaluated at the moment of horizon crossing. For a power-low potential $V = \lambda\phi^n/n$ we get

$$\delta_\phi(k)_{t_f} \sim \lambda^{1/2}\left(\phi^2_{k\sim Ha}\right)^{\frac{n+2}{4}}. \tag{5.294}$$

By using (5.74) we have

$$\begin{aligned}\phi^2_{k\sim Ha} &\sim \ln\frac{a(t_f)}{a(t_k)}\\ &\sim \ln\frac{1}{(aH)_k}H_k a(t_f)\\ &\sim \ln\lambda_{\rm ph}H_k.\end{aligned} \tag{5.295}$$

The physical wavelength is $\lambda_{\rm ph} = a(t_f)/k$. Thus the amplitude of fluctuations at the end of inflation is

$$\delta_\phi(k)_{t_f} \sim \lambda^{1/2}\left(\ln\lambda_{\rm ph}H_k\right)^{\frac{n+2}{4}}. \tag{5.296}$$

We can further make the approximation $H_k \sim H_f$ since curvature scale does not change very much during inflation which is essentially the defining property of inflation. We get finally the amplitude

$$\delta_\phi(k)_{t_f} \sim \lambda^{1/2}\bigg(\ln \lambda_{\rm ph} H_f\bigg)^{\frac{n+2}{4}}. \tag{5.297}$$

Inside the horizon $\lambda_{\rm ph} < 1/H_f$ or equivalently $k > H_f a_f$ the logarithm becomes negative and thus we should instead make the replacment $\phi^2_{k\sim Ha} = \phi_f^2$, i.e. the amplitude comes out proportional to $\lambda^{1/2}$ in this regime. This is the flat space result since gravity is neglected inside the horizon. This is sketched on figure $2{\rm COS}, 2$.

For a quadratic potential $V = m^2\phi^2/2$ we get the amplitude $\delta_\phi = m \ln \lambda_{\rm ph} H_f$. Galactic scales correspond to $L = 10^{25}$cm or equivalently $\ln \lambda_{\rm ph} H_f \sim 50$ and thus in order to get an amplitude of the gravitational potential around 10^{-5} the mass of the inflaton scalar field should be around 10^{-6} in Planck units, viz $m = 10^{-6}.m_{\rm pl} = 10^{-6}\sqrt{\hbar c}/\sqrt{8\pi G} = 10^{-6}10^{18}{\rm GeV} = 10^{12}{\rm GeV}$. At the end of inflation the scalar field is around 1 in Planck units, viz $\phi = 1.1/l_{\rm pl} = \sqrt{c^3}/\sqrt{\hbar G}$. The energy density at the end of inflation is therefore $\rho \sim m^2\phi^2 \sim 10^{-12}.\rho_{\rm pl}$.

5.7.2 Mukhanov Action

The equation (5.270) contains three unknown variables Φ_B, Ψ_B and $\delta\hat{\phi}$ which should also satisfy Einstein's equations. Thus we need to compute the energy-momentum tensor explicitly. We will drop in the following the subscript B and the hat for ease of notation.

We compute (with $u_0|_{\phi_0} = -a$, $u_i|_{\phi_0} = 0$, $X_0 = (\phi_0^{'})^2/2a^2$, etc)

$$\begin{aligned} \delta T^i_j &= 2\bar{\mathcal{P}}\Psi\delta_{ij} + \frac{1}{a^2}\delta T_{ij} \\ &= \delta\mathcal{P}\delta_{ij}. \end{aligned} \tag{5.298}$$

Thus from the Einstein's equation with $i \neq j$ we conclude, as before, that $\Phi_B = \Psi_B$. The other two Einstein's equations are therefore sufficient to determine Φ_B and $\delta\hat{\phi}$. We compute then

$$\begin{aligned} \delta T^0_i &= -\frac{1}{a^2}\delta T_{0i} \\ &= (\bar{\rho} + \bar{\mathcal{P}})u^0|_{\phi_0}\delta u_i. \end{aligned} \tag{5.299}$$

$$\begin{aligned} \delta T^0_0 &= 2\bar{\rho}\Phi - \frac{1}{a^2}\delta T_{00} \\ &= 2(\bar{\rho} + \bar{\mathcal{P}})\Phi + \frac{2}{a}(\bar{\rho} + \bar{\mathcal{P}})\delta u_0 - \delta\rho \\ &= -\delta\rho. \end{aligned} \tag{5.300}$$

In the above two equations we have used

$$\delta u_\mu = -\frac{a}{\phi_0^{'}}\partial_\mu\delta\phi + (\Phi - \frac{\delta\phi^{'}}{\phi_0^{'}})u_\mu|_{\phi_0}. \tag{5.301}$$

Further we compute

$$\begin{aligned}
-\delta T^0_0 &= \delta\rho\\
&= \frac{\partial\rho}{\partial X}\delta X+\frac{\partial\rho}{\partial\phi}\delta\phi\\
&= \frac{\partial\rho}{\partial X}\bigg(-\frac{\Phi\phi_0^{'2}}{a^2}+\frac{\phi_0^{'}\delta\phi^{'}}{a^2}\bigg)+\bigg(-\frac{\partial\rho}{\partial X}\frac{X_0^{'}}{\phi_0^{'}}-\frac{3\mathcal{H}(\bar{\rho}+\bar{\mathcal{P}})}{\phi_0^{'}}\bigg)\delta\phi\\
&= \frac{\partial\rho}{\partial X}\bigg(-\frac{\Phi\phi_0^{'2}}{a^2}+\frac{\phi_0^{'}\delta\phi^{'}}{a^2}-\frac{X_0^{'}}{\phi_0^{'}}\delta\phi\bigg)+\bigg(-\frac{3\mathcal{H}(\bar{\rho}+\bar{\mathcal{P}})}{\phi_0^{'}}\bigg)\delta\phi\\
&= \frac{\partial\rho}{\partial X}\bigg(-\frac{\Phi\phi_0^{'2}}{a^2}+\frac{\phi_0^{'}\delta\phi^{'}}{a^2}-\frac{\phi_0^{''}}{a^2}\delta\phi+\frac{\phi_0^{'}\mathcal{H}}{a^2}\delta\phi\bigg)+\bigg(-\frac{3\mathcal{H}(\bar{\rho}+\bar{\mathcal{P}})}{\phi_0^{'}}\bigg)\delta\phi\\
&= 2X_0\frac{\partial\rho}{\partial X}\bigg(-\Phi+\big(\frac{\delta\phi}{\phi_0^{'}}\big)^{'}+\frac{\mathcal{H}}{\phi_0^{'}}\delta\phi\bigg)+\bigg(-\frac{3\mathcal{H}(\bar{\rho}+\bar{\mathcal{P}})}{\phi_0^{'}}\bigg)\delta\phi\\
&= \frac{\bar{\rho}+\bar{\mathcal{P}}}{c_s^2}\bigg(-\Phi+\big(\frac{\delta\phi}{\phi_0^{'}}\big)^{'}+\frac{\mathcal{H}}{\phi_0^{'}}\delta\phi\bigg)+\bigg(-\frac{3\mathcal{H}(\bar{\rho}+\bar{\mathcal{P}})}{\phi_0^{'}}\bigg)\delta\phi. \qquad (5.302)
\end{aligned}$$

In the last equation we have introduced the speed of sound by the relation

$$c_s^2=\frac{\partial\mathcal{P}}{\partial X}\frac{\partial X}{\partial\rho}=\frac{\rho+\mathcal{P}}{2X}\frac{\partial X}{\partial\rho}. \qquad (5.303)$$

Also

$$\delta T^0_i = -\frac{\bar{\rho}+\bar{\mathcal{P}}}{\phi_0^{'}}\partial_i\delta\phi. \qquad (5.304)$$

The relevant Einstein's equations are now given explicitly by

$$\partial_i^2\Psi-3\mathcal{H}(\Psi^{'}+\mathcal{H}\Phi)=4\pi Ga^2(\bar{\rho}+\bar{\mathcal{P}})\bigg[\frac{1}{c_s^2}\big(-\Phi+\big(\frac{\delta\phi}{\phi_0^{'}}\big)^{'}+\frac{\mathcal{H}}{\phi_0^{'}}\delta\phi\big)-\frac{3\mathcal{H}}{\phi_0^{'}}\delta\phi\bigg]. \qquad (5.305)$$

$$\big(\Psi^{'}+\mathcal{H}\Phi\big)=4\pi Ga^2(\bar{\rho}+\bar{\mathcal{P}})\frac{\delta\phi}{\phi_0^{'}}. \qquad (5.306)$$

By substituting this last equation into the previous one we obtain

$$\begin{aligned}
\partial_i^2\Psi &= \frac{4\pi Ga^2}{\mathcal{H}c_s^2}(\bar{\rho}+\bar{\mathcal{P}})\bigg[\Psi^{'}-4\pi Ga^2(\bar{\rho}+\bar{\mathcal{P}})\frac{\delta\phi}{\phi_0^{'}}+\mathcal{H}\big(\frac{\delta\phi}{\phi_0^{'}}\big)^{'}+\frac{\mathcal{H}^2}{\phi_0^{'}}\delta\phi\bigg]\\
&= \frac{4\pi Ga^2}{\mathcal{H}c_s^2}(\bar{\rho}+\bar{\mathcal{P}})\big(\Psi+\mathcal{H}\frac{\delta\phi}{\phi_0^{'}}\big)^{'}. \qquad (5.307)
\end{aligned}$$

Further

$$(a^2\frac{\Psi}{\mathcal{H}})^{'}=a^2\big(\frac{\Psi^{'}}{\mathcal{H}}+\Psi+4\pi Ga^2\frac{\bar{\rho}+\bar{\mathcal{P}}}{\mathcal{H}^2}\Psi\big). \qquad (5.308)$$

We have then the two equations

$$\partial_i^2\Psi = \frac{4\pi G a^2}{\mathcal{H}c_s^2}(\bar{\rho}+\bar{\mathcal{P})}\big(\Psi+\mathcal{H}\frac{\delta\phi}{\phi_0^{'}}\big)^{'}. \tag{5.309}$$

$$(a^2\frac{\Psi}{\mathcal{H}})^{'} = \frac{4\pi G a^4}{\mathcal{H}^2}(\bar{\rho}+\bar{\mathcal{P}})\big(\Psi+\mathcal{H}\frac{\delta\phi}{\phi_0^{'}}\big). \tag{5.310}$$

We introduce the variables u and v and the parameters z and θ by

$$u=\frac{\Psi}{4\pi G\sqrt{\bar{\rho}+\bar{\mathcal{P}}}}\ ,\ v=\sqrt{\frac{\partial\rho}{\partial X}}a(\delta\phi+\frac{\phi_0^{'}}{\mathcal{H}}\Psi). \tag{5.311}$$

$$z=\frac{a^2\sqrt{\bar{\rho}+\bar{\mathcal{P}}}}{c_s\mathcal{H}}\ ,\ \theta=\frac{1}{c_s z}=\sqrt{\frac{8\pi G}{3}}\frac{1}{a}\frac{1}{\sqrt{1+\frac{\bar{\mathcal{P}}}{\bar{\rho}}}}. \tag{5.312}$$

The Einstein's equations in terms of these new variables take the simpler form

$$\begin{aligned}\partial_i^2 u &= \frac{1}{c_s}z\big(\Psi+\mathcal{H}\frac{\delta\phi}{\phi_0^{'}}\big)^{'}\\ &= \frac{1}{c_s}z\big(\frac{v}{z}\big)^{'}.\end{aligned} \tag{5.313}$$

$$(4\pi G\frac{u}{\theta})^{'}=\frac{4\pi G a^4}{\mathcal{H}^2}(\bar{\rho}+\bar{\mathcal{P}})\frac{v}{z}\Rightarrow(\frac{u}{\theta})^{'}=c_s\frac{v}{\theta}. \tag{5.314}$$

By substituting one of the equations into the other one we find the second order differential equation

$$u^{''}-c_s^2\partial_i^2u-\frac{\theta^{''}}{\theta}u=0. \tag{5.315}$$

This is precisely the Poisson equation (5.232). In fact the definitions of u and θ used here for the scalar field are essentially those used in the hydrodynamical fluid. A similar equation for v holds, viz

$$v^{''}-c_s^2\partial_i^2v-\frac{z^{''}}{z}v=0. \tag{5.316}$$

Since we are interested in quantizing the scalar metric perturbation we will have to quantize the fields u and v. Thus one must start from an appropriate action which gives as equations of motion of the fields u and v precisely the above Poisson equations. This is straightforward and one finds for the field v the action

$$S=\int d\eta d^3x\ \mathcal{L}=\frac{1}{2}\int d\eta d^3x\ \bigg(v^{'2}+c_s^2v\partial_i^2v+\frac{z^{''}}{z}v^2\bigg). \tag{5.317}$$

From this result we see that metric scalar perturbations are given by a massless scalar field in a de Sitter spacetime (see the chapter on QFT on curved backgrounds). This is the most fundamental result in our view and a direct derivation of this action using ADM formalism, which is a very complex calculation, is included in the next section for completeness.

Small scales or short-wavelengths: For a plane wave perturbation with a wavenumber k such that $c_s^2k^2 >> |\theta^{''}/\theta|$ we have the WKB (slowly varying speed of sound c_s) solution

$$u = \frac{C}{\sqrt{c_s}}\exp\big(\pm ik\int c_s d\eta\big). \tag{5.318}$$

The gravitational potential is immediately given

$$\Phi = 4\pi G\dot{\phi}_0 C\sqrt{\frac{\frac{\partial P}{\partial x}}{c_s}}\exp\big(\pm ik\int c_s d\eta\big). \tag{5.319}$$

On the other hand, we can determine the perturbation of the scalar field from (5.306). We get

$$\delta\phi = C\frac{1}{\sqrt{c_s\frac{\partial P}{\partial x}}}\big(\pm ic_s\frac{k}{a} + H + ...\big)\exp\big(\pm ik\int c_s d\eta\big). \tag{5.320}$$

The most important observation here is that both the gravitational potential and the scalar perturbation oscillate in this regime. The amplitude of the gravitational potential is proportional to $\dot{\phi}_0$ and thus will grow at the end of inflation while the amplitude of the scalar perturbation decays as $1/a$.

Large scales or long-wavelengths: These are characterized by $c_s^2k^2 << |\theta^{''}/\theta|$. The solution was found in previous sections and it is given by

$$\Phi = A\frac{d}{dt}\bigg(\frac{1}{a}\int adt\bigg) = A\bigg(1-\frac{H}{a}\int adt\bigg). \tag{5.321}$$

The perturbation of the scalar field from (5.306) is given by

$$\begin{aligned}4\pi Ga^2(\bar{\rho}+\bar{\mathcal{P}})\frac{\delta\phi}{\phi_0^{'}} &= \frac{d}{dt}\big(a\Phi\big)\\ &= -A\dot{H}\int adt\\ &= A4\pi G(\bar{\rho}+\bar{\mathcal{P}})\int adt.\end{aligned} \tag{5.322}$$

Thus

$$\delta\phi = A\dot{\phi}_0\frac{1}{a}\int adt. \tag{5.323}$$

During slow-roll inflation we can make the approximation

$$\begin{aligned}\frac{H}{a}\int adt &= 1-\frac{d}{dt}(\frac{1}{H})+\frac{H}{a}\int\frac{da}{H}\frac{d}{dt}\big(\frac{1}{H}\frac{d}{dt}(\frac{1}{H})\big)\\ &\simeq 1-\frac{d}{dt}(\frac{1}{H})+...\end{aligned} \tag{5.324}$$

Our results reduce therefore during inflation to

$$\Phi = -\frac{A}{H^2}\dot{H} \sim A(\frac{1}{V}\frac{\partial V}{\partial \phi})^2. \tag{5.325}$$

$$\delta\phi = \frac{A\dot{\phi}_0}{H} \sim A\frac{1}{V}\frac{\partial V}{\partial \phi}. \tag{5.326}$$

These are precisely the equations (5.291) and (5.290) respectively. At the end of inflation $V/(\partial V/\partial\phi)$ becomes of order 1 and thus

$$\Phi = A = \left(H\frac{\delta\phi}{\dot{\phi}_0}\right)_{c_s k \sim Ha}. \tag{5.327}$$

We evaluated the different quantities at the instant of horizon crossing.

After inflation the scale factor behaves as $a \propto t^p$. In this case we get the results

$$\Phi = \frac{A}{p+1}. \tag{5.328}$$

$$\delta\phi = \frac{A\dot{\phi}_0}{p+1}t. \tag{5.329}$$

Hence the amplitude of the gravitational field freezes out after inflation. In the radiation-dominated phase corresponding to $p = 1/2$ we get then

$$\begin{aligned}\Phi &= \frac{2A}{3}\\ &= \frac{2}{3}\left(H\frac{\delta\phi}{\dot{\phi}_0}\right)_{c_s k \sim Ha}.\end{aligned} \tag{5.330}$$

Thus the amplitudes at the end of inflation and in the radiation-dominated phase differ only by a numerical coefficient.

5.7.3 Quantization and Inflationary Spectrum

The canonical momentum is defined by the usual formula

$$\pi = \frac{\partial \mathcal{L}}{\partial v'} = v'. \tag{5.331}$$

In the quantum theory we replace v and π with operators $\hat{v}$ and $\hat{\pi}$ satisfying the equal-time commutation relations given by

$$[\hat{v}(\eta,\vec{x}), \hat{\pi}(\eta,\vec{y})] = i\delta^3(\vec{x}-\vec{y}). \tag{5.332}$$

$$[\hat{v}(\eta,\vec{x}),\hat{v}(\eta,\vec{y})]=[\hat{\pi}(\eta,\vec{x}),\hat{\pi}(\eta,\vec{y})]=0. \tag{5.333}$$

We expand the field as

$$\hat{v}(\eta,\vec{x})=\frac{1}{\sqrt{2}}\int\frac{d^3k}{(2\pi)^{3/2}}\bigg(\hat{a}_k v_k^*(\eta)e^{i\vec{k}\vec{x}}+\hat{a}_k^+ v_k(\eta)e^{-i\vec{k}\vec{x}}\bigg). \tag{5.334}$$

Thus

$$\hat{\pi}(\eta,\vec{x})=\frac{1}{\sqrt{2}}\int\frac{d^3k}{(2\pi)^{3/2}}\bigg(\hat{a}_k v_k^{*'}(\eta)e^{i\vec{k}\vec{x}}+\hat{a}_k^+ v_k^{'}(\eta)e^{-i\vec{k}\vec{x}}\bigg). \tag{5.335}$$

This field obeys the equation of motion

$$\hat{v}^{''}-c_s^2\partial_i^2\hat{v}-\frac{z^{''}}{z}\hat{v}=0. \tag{5.336}$$

Equivalently

$$v_k{''}+\omega_k^2(\eta)v_k=0\ ,\ \omega_k^2(\eta)=c_s^2\vec{k}^2-\frac{z^{''}}{z}. \tag{5.337}$$

The creation and annihilation operators are expected to satisfy the commutation relations

$$[\hat{a}_k,\hat{a}_p^+]=\delta^3(\vec{k}-\vec{p}). \tag{5.338}$$

$$[\hat{a}_k,\hat{a}_p]=[\hat{a}_k^+,\hat{a}_p^+]=0. \tag{5.339}$$

We compute then

$$[\hat{v}(\eta,\vec{x}),\hat{\pi}(\eta,\vec{y})]=\frac{1}{2}\int\frac{d^3k}{(2\pi)^3}e^{i\vec{k}(\vec{x}-\vec{y})}\big(v_k^*v_k^{'}-v_kv_k^{*'}\big)\big). \tag{5.340}$$

Thus we must have

$$v_k^*v_k^{'}-v_kv_k^{*'}=2i. \tag{5.341}$$

This is the condition for v_k to be a positive norm solution (see later for more detail). The negative norm solution is immediately given by v_k^*. Alternatively, the above condition is the Wronskian which expresses the linear independence of these two solutions.

The Hamiltonian is given by

$$\begin{aligned}\hat{H}&=\int d^3x(\hat{\pi}\hat{v}^{'}-\mathcal{L})\\&=\frac{1}{2}\int d^3x(\hat{\pi}^2-c_s^2\hat{v}\partial_i^2\hat{v}-\frac{z^{''}}{z}\hat{v}^2)\\&=\int d^3k\bigg(E_k(\hat{a}_k\hat{a}_k^++\hat{a}_k^+a_k)+F_k\hat{a}_k^+\hat{a}_{-k}^++F_k^*\hat{a}_k\hat{a}_{-k}\bigg),\end{aligned} \tag{5.342}$$

where

$$E_k = \frac{1}{2}(|v_k{}'|^2 + \omega_k^2|v_k|^2) \ , \ F_k = \frac{1}{2}((v_k{}')^2 + \omega_k^2(v_k)^2). \quad (5.343)$$

The choice of the vacuum state is a very subtle issue in a curved spacetime (see the chapter on QFT on curved backgrounds). Here, we will simply define the vacuum state as the state annihilated by all the $\hat{a}_k$, viz

$$\hat{a}_k|0>= 0. \quad (5.344)$$

Then

$$\begin{aligned} <0|\hat{H}|0> &= \int d^3k E_k \\ &= \frac{1}{2}\int d^3k(|v_k{}'|^2 + \omega_k^2|v_k|^2). \end{aligned} \quad (5.345)$$

We consider now the ansatz for v_k given by

$$v_k = r_k \exp(i\alpha_k). \quad (5.346)$$

The Wronskian condition becomes

$$r_k^2\alpha_k^{'} = 1. \quad (5.347)$$

The energy of the vacuum in this vacuum becomes

$$<0|\hat{H}|0> = \frac{1}{2}\int d^3k(r_k^{'2} + \frac{1}{r_k^2} + \omega_k^2 r_k^2). \quad (5.348)$$

This energy is minimized when $r_k^{'}(\eta) = 0$ and $r_k(\eta) = 1/\sqrt{\omega_k(\eta)}$. Thus at a given initial time η_0 the energy in the vacuum $|0>$ is minimum iff

$$v_k(\eta_0) = \frac{1}{\sqrt{\omega_k(\eta_0)}}\exp(i\alpha_k(\eta_0)) \ , \ v_k^{'}(\eta_0) = i\sqrt{\omega_k(\eta_0)}\exp(i\alpha_k(\eta_0)). \quad (5.349)$$

The phases $\eta(\eta_0)$ can clearly be set to zero. These are the initial conditions for v_k and $v_k^{'}$. These considerations are well defined for modes with $\omega_k^2 > 0$ or equivalently $c_s^2k^2 > (z^{''}/z)_{\eta_0}$. This is the sub-horizon or sub-Hubble regime. By allowing c_s to change only adiabatically the modes $c_s^2k^2 > (z^{''}/z)_{\eta_0}$ remain not exited and the above minimal fluctuations are well defined. In the case that ω_k is independent of time the vacuum state $|0>$ coincides precisely with the Minkowski vacuum and minimal fluctuations are obviously well defined.

On the other hand, the super-horizon or super-Hubble modes $c_s^2k^2 < (z^{''}/z)_{\eta_0}$ can not be well determined in the same way but fortunately they will be stretched to extreme unobservable distances subsequent to inflation.

We compute now the 2−point function

$$< 0|\hat{\Phi}(\eta,\vec{x})\hat{\Phi}(\eta,\vec{y})|0> \;=\; (4\pi G)^2(\bar{\rho}+\bar{\mathcal{P}}) < 0|\hat{u}(\eta,\vec{x})\hat{u}(\eta,\vec{y})|0> . \tag{5.350}$$

The expansion of the field operator $\hat{u}$ is similarly given by

$$\hat{u}(\eta,\vec{x}) = \frac{1}{\sqrt{2}}\int\frac{d^3k}{(2\pi)^{3/2}}\bigg(\hat{a}_k u_k^*(\eta)e^{i\vec{k}\vec{x}} + \hat{a}_k^+ u_k(\eta)e^{-i\vec{k}\vec{x}}\bigg). \tag{5.351}$$

Thus

$$\begin{aligned}< 0|\hat{\Phi}(\eta,\vec{x})\hat{\Phi}(\eta,\vec{y})|0> &= \frac{1}{2}(4\pi G)^2(\bar{\rho}+\bar{\mathcal{P}}) < 0|\int\frac{d^3p}{(2\pi)^{3/2}}\bigg(\hat{a}_p u_p^*(\eta)e^{i\vec{p}\vec{x}}\bigg).\int\frac{d^3k}{(2\pi)^{3/2}}\bigg(\hat{a}_k^+ u_k(\eta)e^{-i\vec{k}\vec{y}}\bigg)|0>\\ &= \frac{1}{2}(4\pi G)^2(\bar{\rho}+\bar{\mathcal{P}})\int\frac{d^3k}{(2\pi)^3}|u_k(\eta)|^2 e^{i\vec{k}(\vec{x}-\vec{y})}\\ &= 4G^2(\bar{\rho}+\bar{\mathcal{P}})\int k^3|u_k(\eta)|^2\frac{\sin kr}{kr}\frac{dk}{k}.\end{aligned} \tag{5.352}$$

In this equation $r=|\vec{x}-\vec{y}|$. This can be related to the variance σ_k^2 of the gravitational potential Φ as follows. First we write the gravitational potential in the form

$$\hat{\Phi}(\eta,\vec{x}) = \int\frac{d^3k}{(2\pi)^{3/2}}\hat{\Phi}(\eta,\vec{k})e^{i\vec{k}\vec{x}}\;,\;\Phi(\eta,\vec{k}) = \frac{4\pi G\sqrt{\bar{\rho}+\bar{\mathcal{P}}}}{\sqrt{2}}\bigg(\hat{a}_k u_k^*(\eta)+\hat{a}_{-k}^+ u_{-k}(\eta)\bigg). \tag{5.353}$$

Then we have

$$\begin{aligned}< 0|\hat{\Phi}(\eta,\vec{x})\hat{\Phi}(\eta,\vec{y})|0> &= \int\frac{d^3k}{(2\pi)^{3/2}}\int\frac{d^3p}{(2\pi)^{3/2}} < 0|\hat{\Phi}(\eta,\vec{k})\hat{\Phi}(\eta,\vec{p})|0> e^{i\vec{k}\vec{x}}e^{i\vec{p}\vec{y}}\\ &= \int\frac{d^3k}{(2\pi)^{3/2}}\int\frac{d^3p}{(2\pi)^{3/2}}\sigma_k^2\delta^3(\vec{k}+\vec{p})e^{i\vec{k}\vec{x}}e^{i\vec{p}\vec{y}}\\ &= \int\frac{k^3\sigma_k^2}{2\pi^2}\frac{\sin kr}{kr}\frac{dk}{k}.\end{aligned} \tag{5.354}$$

σ_k^2 is precisely the variance of the gravitational potential Φ given by

$$\sigma_k^2 = 8\pi^2 G^2(\bar{\rho}+\bar{\mathcal{P}})|u_k(\eta)|^2. \tag{5.355}$$

The dimensionless variance or power spectrum is defined by

$$\delta_\Phi^2(k) = \frac{k^3\sigma_k^2}{2\pi^2}. \tag{5.356}$$

Short-wavelengths: From the equations of motion $c_s\partial_i^2 u = z(v/z)'$ and $(u/\theta)' = c_s v/\theta$ we find

$$u_k = -\frac{1}{c_s|\vec{k}|^2}({v_k}' - \frac{z'}{z}v_k) \Rightarrow u_k(\eta_0) = -\frac{i}{\sqrt{c_s}|\vec{k}|^{\frac{3}{2}}}\big[1-\frac{1}{c_s^2|\vec{k}|^2}\frac{z''}{z}\big]^{\frac{1}{4}} + \frac{z'}{z}\frac{1}{c_s^{\frac{3}{2}}|\vec{k}|^{\frac{5}{2}}}\big[1-\frac{1}{c_s^2|\vec{k}|^2}\frac{z''}{z}\big]^{-\frac{1}{4}}. \tag{5.357}$$

$$u_k^{'}=c_sv_k-(\frac{c_s^{'}}{c_s}+\frac{z^{'}}{z})u_k\Rightarrow u_k^{'}(\eta_0)=\frac{\sqrt{c_s}}{|\vec{k}|^{\frac{1}{2}}}\big[1-\frac{1}{c_s^2|\vec{k}|^2}\frac{z^{''}}{z}\big]^{-\frac{1}{4}}-(\frac{c_s^{'}}{c_s}+\frac{z^{'}}{z})u_k(\eta_0). \tag{5.358}$$

All functions are of course evaluated at the initial time $\eta=\eta_0$. In the relevant regime of short-wavelengths where $c_s^2|\vec{k}|^2>>(z^{''}/z)_{\eta_0}$ or equivalently $c_s^2|\vec{k}|^2>>(\theta^{''}/\theta)_{\eta_0}$ we can neglect the gravity terms in equations (5.315) and (5.316) and we obtain the initial conditions

$$u_k(\eta_0)=-\frac{i}{\sqrt{c_s}|\vec{k}|^{\frac{3}{2}}}. \tag{5.359}$$

$$u_k^{'}(\eta_0)=\frac{\sqrt{c_s}}{|\vec{k}|^{\frac{1}{2}}}. \tag{5.360}$$

In this regime, the WKB solution of equation (5.315) is therefore given by

$$u_k(\eta)=-\frac{i}{\sqrt{c_s}|\vec{k}|^{\frac{3}{2}}}\exp\big(ik\int_{\eta_0}^{\eta}c_s(\eta')d\eta'\big). \tag{5.361}$$

During inflation $|\dot{H}/H^2|<<1$ and thus in this regime θ behaves as $\theta\sim 1/a$ while a behaves as $a\sim -1/\eta H$. Thus $|\theta''/\theta|\sim|\dot{H}/\eta^2H^2|<<1/\eta^2$. The short-wavelengths regime is given by $c_s^2|\vec{k}|^2>>(\theta^{''}/\theta)_\eta$ or equivalently, with $c_s<<1$ during inflation, by

$$|\eta|>>\frac{1}{k}\sqrt{|\frac{\dot{H}}{H^2}|}. \tag{5.362}$$

Remember that at the end of inflation $\dot{H}/H^2$ becomes of order 1. Equivalently short-wavelengths regime is given by $|\eta|>>1/c_sk$ which is much larger than the previous estimate. On the other hand, horizon crossing is given by $c_sk|\eta|\sim 1$ and long-wavelengths regime is given by $|\eta|<<1/c_sk$. Hence there is a short time interval outside the horizon given by

$$\frac{1}{c_sk}>|\eta|>\frac{1}{k}\sqrt{|\frac{\dot{H}}{H^2}|}, \tag{5.363}$$

in which the solution (5.361) is still valid. Since the above time interval is very narrow the solution (5.361) in this range is effectively a constant, i.e. the gravitational potential freezes at horizon crossing.

In this case the power spectrum is given by

$$\sigma_k^2=8\pi^2G^2(\bar{\rho}+\bar{\mathcal{P}})\frac{1}{c_sk^3}\Rightarrow\delta_\Phi^2(k,t)=4G^2\frac{\bar{\rho}+\bar{\mathcal{P}}}{c_s}\ ,\ c_sk>>Ha. \tag{5.364}$$

Long-wavelengths: In this case the solution is given by (5.321), viz

$$u_k=\frac{\Phi_k}{4\pi G\sqrt{\bar{\rho}+\bar{\mathcal{P}}}}=\frac{A_k}{4\pi G\sqrt{\bar{\rho}+\bar{\mathcal{P}}}}\bigg(1-\frac{H}{a}\int adt\bigg). \tag{5.365}$$

During inflation, and using $\dot{H}/4\pi G=-(\bar{\rho}+\bar{\mathcal{P}})$, we have

$$\begin{aligned} u_k &= \frac{A_k}{4\pi G\sqrt{\bar{\rho}+\bar{\mathcal{P}}}}\frac{d}{dt}(\frac{1}{H}) \\ &= A_k\frac{\sqrt{\bar{\rho}+\bar{\mathcal{P}}}}{H^2}. \end{aligned} \tag{5.366}$$

This is constant in the time interval (5.363). This should be compared with the solution (5.361) which holds in the time interval (5.363). Since both η_0 and η_0 are in this short time interval they can be taken both to be equal to the moment of horizon crossing. This allows us to fix A_k as

$$A_k=-\frac{i}{k^{\frac{3}{2}}}\bigg(\frac{H^2}{\sqrt{c_s(\bar{\rho}+\bar{\mathcal{P}})}}\bigg)_{c_sk\sim Ha}. \tag{5.367}$$

In this case the power spectrum is given by

$$\begin{aligned} &\sigma_k^2=\frac{1}{2k^3}\bigg(\frac{H^4}{c_s(\bar{\rho}+\bar{\mathcal{P}})}\bigg)_{c_sk\sim Ha}\bigg(1-\frac{H}{a}\int adt\bigg)^2 \\ &\Rightarrow \delta_\Phi^2(k,t)=\frac{16}{9}G^2\bigg(\frac{\bar{\rho}}{c_s(1+\frac{\bar{\mathcal{P}}}{\bar{\rho}})}\bigg)_{c_sk\sim Ha}\bigg(1-\frac{H}{a}\int adt\bigg)^2\;,\;(Ha/c_s)_i<k<<Ha/c_s. \end{aligned} \tag{5.368}$$

This formula gives the time evolution of long-wavelength perturbations even after inflation. After inflation the universe is radiation-dominated (where CMB originated) and hence $a\sim t^{1/2}$. In this case we get the power spectrum

$$\delta_\Phi^2=\frac{64}{81}G^2\bigg(\frac{\bar{\rho}}{c_s(1+\frac{\bar{\mathcal{P}}}{\bar{\rho}})}\bigg)_{c_sk\sim Ha}\;,\;(Ha/c_s)_i<k<(Ha/c_s)_f. \tag{5.369}$$

This results applies for large scales which includes the whole universe. This depends on the energy density $\bar{\rho}$ and the deviation of the equation of state from the vacuum given by $\Delta w=1+\bar{\mathcal{P}}/\bar{\rho}$ at the instant of horizon crossing. We know that $\delta_\Phi\sim 10^{-5}$ on galactic scales while Δw is estimated as 10^{-2} thus the energy density at horizon crossing must be of the order of $10^{-12}G^{-2}$, i.e. 10^{-12} of the Planck density. This is the same estimate obtained previously.

The above spectrum depends on the scale slightly. The requirement that inflation must have a graceful exist means that the energy density decreases slowly while the deviation of the equation of state from the vacuum increases slowly at the end of inflation, and as a consequence, the perturbations which cross the horizon earlier have larger amplitudes than those which cross the horizon later. A flat (scale-invariant) spectrum is characterized by a spectral index $n_s = 1$ where n_s is defined through the power law

$$\delta_\Phi^2 \sim k^{n_s - 1}. \tag{5.370}$$

Obviously

$$n_s - 1 = \frac{d \ln \delta_\Phi^2}{d \ln k}. \tag{5.371}$$

On the other hand,

$$\begin{aligned} n_s - 1 &= \frac{1}{H}\frac{\dot{\bar{\rho}}}{\bar{\rho}} - \frac{1}{H}\frac{d}{dt}\big(\ln c_s + \ln(1 + \frac{\bar{\mathcal{P}}}{\bar{\rho}})\big) \\ &= 2\frac{\dot{H}}{H^2} - \frac{1}{H}\frac{d}{dt}\big(\ln c_s + \ln(1 + \frac{\bar{\mathcal{P}}}{\bar{\rho}})\big) \\ &= -3(1 + \frac{\bar{\mathcal{P}}}{\bar{\rho}}) - \frac{1}{H}\frac{d}{dt}\big(\ln c_s + \ln(1 + \frac{\bar{\mathcal{P}}}{\bar{\rho}})\big). \end{aligned} \tag{5.372}$$

In the above equation we have used the approximation $d \ln k = d \ln a_k = H dt$. Since all correction terms are negative we have $n_s < 1$ and thus the amplitude increases slightly for small k corresponding to larger scales. We say that the spectrum is red-tilted. This tilt can be traced to the requirement that inflation must have a graceful exit.

An estimation for n_s can be given as follows. Galactic scales cross the horizon at 50 e-folds before the end of inflation. At this time the deviation of the equation of state from the vacuum is around 10^{-2} and the second term in (5.372) is also around 10^{-2} and hence $n_s = 0.96$. This should be compared with the 2013 Planck result $n_s = 0.9603 \pm 0.0073$.

For inflation with a potential V the above formula becomes

$$\begin{aligned} n_s - 1 &= -\frac{1}{8\pi G}\big(\frac{1}{V}\frac{\partial V}{\partial \phi}\big)^2 - \frac{2}{H}\frac{d}{dt}\ln\frac{1}{V}\frac{\partial V}{\partial \phi} \\ &= -\frac{1}{8\pi G}\big(\frac{1}{V}\frac{\partial V}{\partial \phi}\big)^2 + \frac{1}{4\pi G}\left(\frac{1}{V}\frac{\partial^2 V}{\partial \phi^2} - \big(\frac{1}{V}\frac{\partial V}{\partial \phi}\big)^2\right) \\ &= -\frac{3}{8\pi G}\big(\frac{1}{V}\frac{\partial V}{\partial \phi}\big)^2 + \frac{1}{4\pi G}\frac{1}{V}\frac{\partial^2 V}{\partial \phi^2} \\ &= -6\epsilon_V + 2\eta_V. \end{aligned} \tag{5.373}$$

5.8 Rederivation of the Mukhanov Action

5.8.1 Mukhanov Action from ADM

The action of interest here is of course

$$S=\frac{1}{2}\int d^4x\sqrt{-\det g}\,\bigg[R-g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi-2V(\phi)\bigg]. \tag{5.374}$$

By going now through the steps of the famous ADM (Arnowitt, Deser and Misner) formalism we can express this action in terms of 3−dimensional quantities which is very useful if one is interested in canonical quantization. The ADM formalism starts with the metric put in the form

$$ds^2=-N^2dt^2+\gamma_{ij}(dx^i+N^idt)(dx^j+N^jdt). \tag{5.375}$$

In other words we slice spacetime into 3−dimensional spatial hypersurfaces. Indeed γ_{ij} is the metric on the spatial 3−dimensional slices of constant t. The function N and the vector N_i are called lapse function and shift vector. We have

$$g_{00}=\gamma_{ij}N^iN^j-N^2\ ,\ g_{0j}=\gamma_{ij}N^i\ ,\ g_{i0}=\gamma_{ij}N^j\ ,\ g_{ij}=\gamma_{ij}. \tag{5.376}$$

A straightforward calculation shows that

$$\sqrt{-\det g}=N\sqrt{\det\gamma}. \tag{5.377}$$

$$g^{00}=-\frac{1}{N^2}\ ,\ g^{0j}=\frac{1}{N^2}N^j\ ,\ g^{i0}=\frac{1}{N^2}N^i\ ,\ g^{ij}=\gamma^{ij}-\frac{1}{N^2}N^iN^j. \tag{5.378}$$

The variables γ_{ij} , N_i and N contain the same information as the original spacetime metric $g_{\mu\nu}$. As it turns out N and N_i are only Lagrange multipliers.

We get after some calculation the action

$$S=\frac{1}{2}\int d^4x\sqrt{\det\gamma}\,\bigg[NR_{(3)}+N^{-1}(E_{ij}E^{ij}-E^2)+N^{-1}(\partial_t\phi-N^i\partial_i\phi)^2-N\gamma^{ij}\partial_i\phi\partial_j\phi-2NV\bigg]. \tag{5.379}$$

The extrinsic curvature of the three-dimensional spatial slices is $K_{ij}=N^{-1}E_{ij}$ where

$$E_{ij}=\frac{1}{2}(\partial_t\gamma_{ij}-\nabla_iN_j-\nabla_jN_i)\ ,\ E=\gamma^{ij}E_{ij}. \tag{5.380}$$

Recall that

$$\nabla_iN_j=\partial_iN_j-\Gamma^k{}_{ij}N_k\ ,\ \Gamma^k{}_{ij}=\frac{1}{2}\gamma^{kl}(\partial_i\gamma_{lj}+\partial_j\gamma_{li}-\partial_l\gamma_{ij}). \tag{5.381}$$

By varying the above action with respect to N and N^i we obtain the equations of motion

$$R_{(3)}-N^{-2}(E_{ij}E^{ij}-E^2)-N^{-2}(\partial_t\phi-N^i\partial_i\phi)^2-\gamma^{ij}\partial_i\phi\partial_j\phi-2V=0. \tag{5.382}$$

$$-N^{-1}\partial_i\phi(\partial_t\phi - N^i\partial_i\phi) + \nabla_j\big(N^{-1}(E_i^j - \gamma_i^j E)\big) = 0. \quad (5.383)$$

These are constraints equations for the lapse function N and the shift vector N^i. In the comoving gauge we will choose $\delta\phi = 0$ and hence $\phi = \bar{\phi}$ where the unperturbed configuration $\bar{\phi}$ is uniform. Hence the above equations of motion reduce to

$$R_{(3)} - N^{-2}(E_{ij}E^{ij} - E^2) - N^{-2}(\partial_t\bar{\phi})^2 - 2\bar{V} = 0. \quad (5.384)$$

$$\nabla_j\big(N^{-1}(E_i^j - \gamma_i^j E)\big) = 0. \quad (5.385)$$

In the comoving gauge we also choose

$$\gamma_{ij} = a^2(1-2\mathcal{R})\delta_{ij} + h_{ij}\ ,\ h^i{}_i = \partial^i h_{ij} = 0. \quad (5.386)$$

In most of the following we will set $h = 0$. Then

$$R_{(3)} = \frac{4}{a^2}\vec{\nabla}^2\mathcal{R}. \quad (5.387)$$

We resolve the shift vector N_i into the sum of a total derivative (irrotational scalar) and a divergenceless vector (incompressible vector) as

$$N_i = \partial_i\psi + \tilde{N}_i\ ,\ \partial^i\tilde{N}_i = 0. \quad (5.388)$$

We also introduce the lapse perturbation α as

$$N = 1 + \alpha. \quad (5.389)$$

We expand ψ, $\tilde{N}_i$ and α in powers of $\mathcal{R}$ as follows

$$\psi = \psi_1 + \psi_2 + ... \quad (5.390)$$

$$\alpha = \alpha_1 + \alpha_2 + ... \quad (5.391)$$

$$\tilde{N}_i = \tilde{N}_i^{(1)} + \tilde{N}_i^{(2)} + ... \quad (5.392)$$

We have

$$\gamma^{ij} = \frac{1}{a^2(1-2\mathcal{R})}\delta_{ij}. \quad (5.393)$$

Since we are only going to keep the first order in powers of $\mathcal{R}$ we can approximate E_{ij} by

$$\begin{aligned} E_{ij} &= \frac{1}{2}(\partial_t\gamma_{ij} - \partial_i N_j - \partial_j N_i) \\ &= a^2H[1 - 2\mathcal{R} - \frac{\dot{\mathcal{R}}}{H}]\delta_{ij} - \frac{1}{2}(\partial_i N_j + \partial_j N_i). \end{aligned} \quad (5.394)$$

Thus we compute

$$E_{ij}E^{ij} = \frac{1}{a^4(1-2\mathcal{R})^2}E_{ij}E_{ij}$$
$$\simeq 3H^2[1-2\frac{\dot{\mathcal{R}}}{H}] - \frac{2H}{a^2}\partial_i N_i. \quad (5.395)$$

$$E \simeq 3H[1-\frac{\dot{\mathcal{R}}}{H}] - \frac{1}{a^2}\partial_i N_i \Rightarrow E^2 \simeq 9H^2[1-2\frac{\dot{\mathcal{R}}}{H}] - \frac{6H}{a^2}\partial_i N_i. \quad (5.396)$$

The constraints become (by using the first Friedmann equation in the form $6H^2=(\partial_t\bar{\phi})^2+2\bar{V}$ and $8\pi G=1$)

$$\frac{4}{a^2}\vec{\nabla}^2\mathcal{R} - 12H\dot{\mathcal{R}} - \frac{4H}{a^2}\partial_i N_i - 12\alpha H^2 + 2\alpha(\partial_t\bar{\phi})^2 = 0. \quad (5.397)$$

$$\partial_j\bigg(-2H[1-\alpha-\frac{\dot{\mathcal{R}}}{H}]\delta_{ij} + \frac{1}{a^2}\partial_k N_k\delta_{ij} - \frac{1}{2a^2}(\partial_i N_j + \partial_j N_i)\bigg) = 0. \quad (5.398)$$

Equivalently (with $\vec{\nabla}_i^2 = \partial^i\partial_i$)

$$\frac{4}{a^2}\vec{\nabla}^2\mathcal{R} - 12H\dot{\mathcal{R}} - 4H\vec{\nabla}^2\psi_1 - 4\alpha_1\bar{V} = 0. \quad (5.399)$$

$$2H\partial_i(\alpha_1 + \frac{\dot{\mathcal{R}}}{H}) - \frac{1}{2}\vec{\nabla}^2\tilde{N}_i^{(1)} = 0. \quad (5.400)$$

From the second constraint we obtain

$$\alpha_1 = -\frac{\dot{\mathcal{R}}}{H}\ ,\ \tilde{N}_i^{(1)} = 0. \quad (5.401)$$

The first constraint gives then

$$\vec{\nabla}^2\psi_1 = \frac{\vec{\nabla}^2\mathcal{R}}{a^2H} + \dot{\mathcal{R}}(\frac{\bar{V}}{H^2} - 3). \quad (5.402)$$

Recall that the slow-roll parameter ϵ is given by

$$\epsilon = \frac{(\partial_t\bar{\phi})^2}{2H^2} = 3 - \frac{\bar{V}}{H^2}. \quad (5.403)$$

Hence we obtain

$$\psi_1 = \frac{\mathcal{R}}{a^2H} - \epsilon(\vec{\nabla}^2)^{-1}\dot{\mathcal{R}}. \quad (5.404)$$

We compute

$$\begin{aligned}\sqrt{\det\gamma} &= a^3\sqrt{\det\gamma_{(3)}}\\ &= a^3[1-3\mathcal{R}+\frac{3}{2}\mathcal{R}^2+...].\end{aligned} \quad (5.405)$$

$$\begin{aligned}\mathcal{L} &= NR_{(3)}+N^{-1}(E_{ij}E^{ij}-E^2)+N^{-1}(\partial_t\phi-N^i\partial_i\phi)^2-N\gamma^{ij}\partial_i\phi\partial_j\phi-2NV\\ &= \mathcal{L}_0+\mathcal{L}_1+\mathcal{L}_2+...\end{aligned} \quad (5.406)$$

$$\mathcal{L}_0 = (\partial_t\bar{\phi})^2-2\bar{V}+(E_{ij}E^{ij}-E^2)^{(0)}. \quad (5.407)$$

$$\mathcal{L}_1 = \mathcal{R}_{(3)}-\alpha_1(\partial_t\bar{\phi})^2-2\alpha_1\bar{V}+(E_{ij}E^{ij}-E^2)^{(1)}-\alpha_1(E_{ij}E^{ij}-E^2)^{(0)}. \quad (5.408)$$

$$\begin{aligned}\mathcal{L}_2 &= \alpha_1\mathcal{R}_{(3)}+(-\alpha_2+\alpha_1^2)(\partial_t\bar{\phi})^2-2\alpha_2\bar{V}+(E_{ij}E^{ij}-E^2)^{(2)}-\alpha_1(E_{ij}E^{ij}-E^2)^{(1)}\\ &+ (-\alpha_2+\alpha_1^2)(E_{ij}E^{ij}-E^2)^{(0)}.\end{aligned} \quad (5.409)$$

A more precise formula for $E_{ij}E^{ij}-E^2$ is

$$\begin{aligned}E_{ij}E^{ij}-E^2 &= -6H^2\big[1-\frac{2\dot{\mathcal{R}}}{H}+\frac{\dot{\mathcal{R}}^2}{H^2}-\frac{4\mathcal{R}\dot{\mathcal{R}}}{H}\big]+\frac{4H}{a^2}\big[1-\frac{\dot{\mathcal{R}}}{H}+2\mathcal{R}\big]\nabla_iN_i\\ &+ \frac{1}{4a^4}(\nabla_iN_j+\nabla_jN_i)^2-\frac{1}{a^4}(\nabla_iN_i)^2.\end{aligned} \quad (5.410)$$

The last term is already of order 2 and thus we can set $\nabla_iN_j=\partial_iN_j$. By partial integration we can see that this term actually cancels. Further we compute

$$\nabla_iN_j=\partial_iN_j+\partial_i\mathcal{R}.N_j+\partial_j\mathcal{R}.N_i-\delta_{ij}\partial_k\mathcal{R}.N_k. \quad (5.411)$$

By using the equation of motion $(\partial_t\bar{\phi})^2+2\bar{V}+(E_{ij}E^{ij}-E^2)^{(0)}=0$ we obtain

$$\mathcal{L}_0 = -4\bar{V}. \quad (5.412)$$

$$\begin{aligned}\mathcal{L}_1 &= \mathcal{R}_{(3)}+(E_{ij}E^{ij}-E^2)^{(1)}\\ &= \frac{8}{a^2}\vec{\nabla}^2\mathcal{R}+12H\dot{\mathcal{R}}-4\epsilon H\dot{\mathcal{R}}.\end{aligned} \quad (5.413)$$

$$\begin{aligned}\mathcal{L}_2 &= \alpha_1\mathcal{R}_{(3)}+\alpha_1^2(\partial_t\bar{\phi})^2+(E_{ij}E^{ij}-E^2)^{(2)}-\alpha_1(E_{ij}E^{ij}-E^2)^{(1)}+\alpha_1^2(E_{ij}E^{ij}-E^2)^{(0)}\\ &= -\frac{4}{a^2H}\dot{\mathcal{R}}\vec{\nabla}^2\mathcal{R}+\frac{\dot{\mathcal{R}}^2}{H^2}(\partial_t\bar{\phi})^2+24H\mathcal{R}\dot{\mathcal{R}}+\frac{12}{a^2}\mathcal{R}\vec{\nabla}^2\mathcal{R}-12\epsilon H\mathcal{R}\dot{\mathcal{R}}.\end{aligned} \quad (5.414)$$

The first term in $\mathcal{L}_1$ is a boundary term by Stokes theorem and thus it will be neglected, viz

$$\begin{aligned}\int d^4x\sqrt{\det\gamma}\,\frac{8}{a^2}\vec{\nabla}^2\mathcal{R} &= 8\int dt\ a(t)\int d^3x\sqrt{\det\gamma_{(3)}}\ \partial^i\partial_i\mathcal{R}\\ &= 8\int dt\ a(t)\int d^2x\sqrt{\det\gamma_{(2)}}\ n^i\partial_i\mathcal{R}\\ &= 0.\end{aligned} \tag{5.415}$$

The quadratic contribution coming from $\mathcal{L}_0$ is

$$(\frac{3}{2}\mathcal{R}^2)\mathcal{L}_0 = -6\mathcal{R}^2\bar{V}. \tag{5.416}$$

The quadratic contribution coming from $\mathcal{L}_1$ is

$$(-3\mathcal{R})\mathcal{L}_1 = -36H\mathcal{R}\dot{\mathcal{R}} + 12\epsilon H\mathcal{R}\dot{\mathcal{R}}. \tag{5.417}$$

Thus

$$(\frac{3}{2}\mathcal{R}^2)\mathcal{L}_0 + (-3\mathcal{R})\mathcal{L}_1 + \mathcal{L}_2 = -\frac{4}{a^2H}\dot{\mathcal{R}}\vec{\nabla}^2\mathcal{R} + \frac{\dot{\mathcal{R}}^2}{H^2}(\partial_t\bar{\phi})^2 - 12H\mathcal{R}\dot{\mathcal{R}} + \frac{12}{a^2}\mathcal{R}\vec{\nabla}^2\mathcal{R} - 6\mathcal{R}^2\bar{V}. \tag{5.418}$$

This must be multiplied by a^3. Integration by parts gives

$$\begin{aligned}(\frac{3}{2}\mathcal{R}^2)\mathcal{L}_0 + (-3\mathcal{R})\mathcal{L}_1 + \mathcal{L}_2 &= -\frac{4}{a^2H}\dot{\mathcal{R}}\vec{\nabla}^2\mathcal{R} + \frac{\dot{\mathcal{R}}^2}{H^2}(\partial_t\bar{\phi})^2 + \frac{12}{a^2}\mathcal{R}\vec{\nabla}^2\mathcal{R}\\ &= \frac{(\partial_t\bar{\phi})^2}{H^2}\bigg(\dot{\mathcal{R}}^2 - \frac{1}{a^2}\partial^i\mathcal{R}\partial_i\mathcal{R}\bigg) + \frac{14}{a^2}\mathcal{R}\vec{\nabla}^2\mathcal{R}.\end{aligned} \tag{5.419}$$

Since we are only keeping quadratic terms in the curvature perturbation $\mathcal{R}$ the last term in the above equation vanishes by (5.415). Indeed we have

$$\begin{aligned}14\int dt\ a(t)\int d^3x\mathcal{R}\vec{\nabla}^2\mathcal{R} &\simeq -\frac{14}{3}\int dt\ a(t)\int d^3x\sqrt{\det\gamma_{(3)}}\ \partial^i\partial_i\mathcal{R}\\ &= 0\end{aligned} \tag{5.420}$$

We obtain the final action

$$S = \frac{1}{2}\int d^4x a^3\frac{(\partial_t\bar{\phi})^2}{H^2}\bigg(\dot{\mathcal{R}}^2 - \frac{1}{a^2}\partial^i\mathcal{R}\partial_i\mathcal{R}\bigg). \tag{5.421}$$

There is also a linear term in $\mathcal{R}$ which we must discuss. This is given by

$$(-3\mathcal{R})\mathcal{L}_0 + (1)\mathcal{L}_1 = \frac{4\bar{V}}{H}(\dot{\mathcal{R}} + 3H\mathcal{R}). \tag{5.422}$$

Again this must be multiplied by a^3. After integration by parts we obtain

$$(-3\mathcal{R})\mathcal{L}_0 + (1)\mathcal{L}_1 = -4\epsilon\mathcal{R}(\bar{V} + 2H\frac{\delta\bar{V}}{\delta\bar{\phi}}). \tag{5.423}$$

This can be neglected in the slow-roll limit $\epsilon \longrightarrow 0$.

The conformal time is defined by $dt = ad\eta$. We introduce also Mukhanov variables

$$v = z\mathcal{R} \ , \ z = a\frac{\partial_t\bar{\phi}}{H}. \tag{5.424}$$

Now a really straightforward calculation gives the action

$$S = \frac{1}{2}\int d\eta d^3x\bigg((v^{'})^2 + \frac{z^{''}}{z}v^2 - \partial^i v\partial_i v\bigg). \tag{5.425}$$

5.8.2 Power Spectra and Tensor Perturbations

The equation of motion derived from the above action reads

$$v^{''} - \partial_i\partial^i v - \frac{z^{''}}{z}v = 0. \tag{5.426}$$

A solution is given by $u_k = \exp(i\vec{k}\vec{x})\chi_k$ (with $\vec{k}\vec{x} = k^i x_i$) provided

$$\chi_k^{''} + (k^2 - \frac{z^{''}}{z})\chi_k = 0. \tag{5.427}$$

These solutions are positive norm solutions, viz $(u_k, u_l) = \delta_{kl}$ if and only if

$$iV(\chi_k^*\dot{\chi}_k - \chi_k\dot{\chi}_k^*) = 1. \tag{5.428}$$

The negative norm solutions are u_k^*. As before we will choose $\chi_k = v_k^*/\sqrt{2}$. The field v can then be expanded as

$$v = \int\frac{d^3k}{(2\pi)^3}\frac{1}{\sqrt{2}}\bigg[a_k v_k^*(\eta)e^{i\vec{k}\vec{x}} + a_k^* v_k(\eta)e^{-i\vec{k}\vec{x}}\bigg]. \tag{5.429}$$

In the quantum theory a_k and a_k^+ become operators $\hat{a}_k$ and $\hat{a}_k^+$ satisfying $[\hat{a}_k, \hat{a}_l^+] = V\delta_{kl}$. The field operator is

$$\hat{v} = \int\frac{d^3k}{(2\pi)^3}\frac{1}{\sqrt{2}}\bigg[\hat{a}_k v_k^*(\eta)e^{i\vec{k}\vec{x}} + \hat{a}_k^+ v_k(\eta)e^{-i\vec{k}\vec{x}}\bigg]. \tag{5.430}$$

We are interested in the 2−point function

$$\begin{aligned} <\hat{\mathcal{R}}(x_1)\hat{\mathcal{R}}(x_2)> &= \frac{H^2}{a^2(\partial_t\bar{\phi})^2} <\hat{v}(t_1)\hat{v}(t_2)> \\ &= \frac{H^2}{2a^2(\partial_t\bar{\phi})^2}\int\frac{d^3k}{(2\pi)^3}v_k^*(t_1)v_k(t_2). \end{aligned} \tag{5.431}$$

We define the Fourier transform of $\hat{\mathcal{R}}(x)$ by

$$\hat{\mathcal{R}}(x) = \int\frac{d^3k}{(2\pi)^3}\hat{\mathcal{R}}_k(t)e^{i\vec{k}\vec{x}}. \tag{5.432}$$

We define the power spectrum $P_{\mathcal{R}}(k)$ of the curvature perturbation $\mathcal{R}$ by

$$< \hat{\mathcal{R}}_{k_1}(t_1)\hat{\mathcal{R}}_{k_2}(t_2) > \ = \ (2\pi)^3\delta^3(k_1+k_2)P_{\mathcal{R}}(k_1). \tag{5.433}$$

We compute then

$$< \hat{\mathcal{R}}(x_1)\hat{\mathcal{R}}(x_2) > \ = \ \int \frac{d^3k}{(2\pi)^3}P_{\mathcal{R}}(k_1)e^{i\vec{k}_1(\vec{x}_1-\vec{x}_2)}. \tag{5.434}$$

Let us now consider the de Sitter limit $\epsilon \longrightarrow 0$ in which H can be treated as a constant and $a \simeq e^{Ht}$ or equivalently $a \simeq -1/(H\eta)$. We compute

$$\frac{z^{''}}{z} = a\dot{a}\frac{\dot{z}}{z} + a^2\frac{\ddot{z}}{z}. \tag{5.435}$$

$$\dot{z} = \dot{a}\frac{\partial_t\bar{\phi}}{H} - a\frac{\partial_t\bar{\phi}}{H^2}\dot{H} + a\frac{\ddot{a}}{H}. \tag{5.436}$$

In the de Sitter limit case we can make the approximations

$$\dot{z} \simeq \dot{a}\frac{\partial_t\bar{\phi}}{H} \Rightarrow \frac{\dot{z}}{z} \simeq \frac{\dot{a}}{a} \ , \ \frac{\ddot{z}}{z} \simeq \frac{\ddot{a}}{a}. \tag{5.437}$$

Thus

$$\begin{aligned} \frac{z^{''}}{z} &\simeq \frac{a^{''}}{a} \\ &\simeq \frac{2}{\eta^2}. \end{aligned} \tag{5.438}$$

The equation of motion becomes

$$\chi_k^{''} + (k^2 - \frac{2}{\eta^2})\chi_k = 0. \tag{5.439}$$

In the limit $\eta \longrightarrow -\infty$ the frequency approaches the flat space result and hence we can choose the vacuum state to be given by the Minkowski vacuum. This is the Bunch-Davies vacuum given by equation (6.128). We have then

$$v_k = -\frac{e^{ik\eta}}{\sqrt{k}}(1+\frac{i}{k\eta}). \tag{5.440}$$

We can then compute in the de Sitter limit the real space variance

$$< \hat{\mathcal{R}}(x)\hat{\mathcal{R}}(x) >= \int_0^\infty d\ln k \Delta_{\mathcal{R}}^2(k). \tag{5.441}$$

The dimensionless power spectrum $\Delta_{\mathcal{R}}^2(k)$ is given by

$$\Delta_{\mathcal{R}}^2(k) = \frac{H^2}{(2\pi)^2}\frac{H^2}{(\partial_t\bar{\phi})^2}(1+k^2\eta^2). \tag{5.442}$$

For super-horizon scales ($|k\eta| << 1$ or equivalently $k << aH$) this dimensionless power spectrum becomes constant. This is precisely the statement that $\mathcal{R}$ remains constant outside the horizon. We may then restrict the calculation to the instant of horizon crossing given by

$$|k\eta_*| = 1 \Leftrightarrow k = a(t_*)H(t_*). \tag{5.443}$$

The dimensionless power spectrum $\Delta^2_{\mathcal{R}}(k)$ and the power spectrum $P_{\mathcal{R}}(k)$ at horizon crossing are given respectively by [2]

$$\Delta^2_{\mathcal{R}}(k) = \frac{H_*^2}{2\pi^2}\frac{H_*^2}{(\partial_t\bar{\phi})^2_*}. \tag{5.444}$$

$$\begin{aligned} P_{\mathcal{R}}(k) &= \frac{2\pi^2}{k^3}\Delta^2_{\mathcal{R}}(k) \\ &= \frac{H_*^2}{k^3}\frac{H_*^2}{(\partial_t\bar{\phi})^2_*}. \end{aligned} \tag{5.445}$$

In summary the primordial power spectrum of comoving curvature perturbation $\mathcal{R}$ at horizon crossing is found to be given by

$$P_{\mathcal{R}}(k) = \frac{H_*^2}{k^3}\frac{H_*^2}{(\partial_t\bar{\phi})^2_*} \Leftrightarrow \Delta^2_{\mathcal{R}}(k) = \frac{H_*^2}{2\pi^2}\frac{H_*^2}{(\partial_t\bar{\phi})^2_*}. \tag{5.446}$$

This is the scalar power spectrum, viz

$$P_s(k) \equiv P_{\mathcal{R}}(k) \Leftrightarrow \Delta^2_s(k) \equiv \Delta^2_{\mathcal{R}}(k) = \frac{1}{4\pi^2}\frac{H^2}{M^2_{\rm pl}}\frac{1}{\epsilon}|_{k=aH}. \tag{5.447}$$

As seen from the gauge fixing condition (5.386) there is extra degrees of freedom (2 polarizations) encoded in the symmetric traceless and divergenceless tensor h_{ij} which we have not considered at all until now. These degrees of freedom correspond to gravitational waves. In order to determine the primordial power spectrum $\Delta^2_t(k)$ of the tensor perturbation h at horizon crossing $k = aH$ we go back to (5.386) and set $\mathcal{R} = 0$ and then go through some very similar calculations to those which led to $\Delta^2_s(k)$. We find at the end the result

$$\Delta^2_t(k) \equiv 2\Delta^2_h(k) = \frac{2}{\pi^2}\frac{H^2}{M^2_{\rm pl}}|_{k=aH}. \tag{5.448}$$

The scalar-to-tensor ratio is defined by

$$\begin{aligned} r &\equiv \frac{\Delta^2_t(k)}{\Delta^2_s(k)} \\ &= 8\epsilon_*. \end{aligned} \tag{5.449}$$

[2]These two formulas differ by a factor of 1/2 compared with reference [8].

The scale-dependence of the power spectrum $\Delta_s^2(k)$ can be given by the so-called spectral index n_s defined by

$$n_s = 1 + \frac{d\ln \Delta_s^2}{d\ln k}. \tag{5.450}$$

Obviously scale invariance corresponds to $n_s = 1$. We may approximate $\Delta_s^2(k)$ by a power law as follows

$$\Delta_s^2(k) = A_s(k_*)(\frac{k}{k_*})^{n_s(k_*)-1+\frac{1}{2}\alpha_s(k_*)\ln\frac{k}{k_*}}\ ,\ \alpha_s(k) = \frac{dn_s}{d\ln k}. \tag{5.451}$$

Similarly we define

$$n_t = \frac{d\ln \Delta_s^2}{d\ln k}. \tag{5.452}$$

In terms of the Hubble slow-roll parameters ϵ and η the indices n_s and n_t are given by

$$n_s = 1 + 2\eta_* - 4\epsilon_*. \tag{5.453}$$

$$n_t = -2\epsilon_*. \tag{5.454}$$

In the slow-roll limit with $m^2\phi^2$ potential we obtain the predictions

$$n_s = 0.96\ ,\ r = 0.05. \tag{5.455}$$

Let us summarize our results so far. During inflation the comoving horizon $1/(aH)$ decreases while after inflation it increases. In this inflationary universe fluctuation are created quantum mechanically on all scales with a spectrum of wave numbers k. The comoving scales k^{-1} are constant during and after inflation. The physically relevant fluctuations are created at sub-horizon scales $k > aH$. Any given fluctuation with a wave number k starts thus inside the horizon and at some point it will exit the horizon (during inflation) and then it will re-enter again the horizon at a later time (after inflation during the hot big bang). All fluctuations after they exit the horizon (corresponding to super-horizon scales $k < aH$) are frozen until they re-enter the horizon in the sense that they are not affected by and they can not affect the physics inside the horizon. This is the statement that the curvature perturbation $\mathcal{R}$ is constant outside the horizon which allows us to concentrate on the value of $\mathcal{R}$ at the time of exit (crossing) since that value will not change until re-entry. See figure 3COS.

5.8.3 CMB Temperature Anisotropies

The remaining question we would like to discuss is how to relate the power spectrum P_s to CMB temperature anisotropies. The CMB temperature fluctuations $\Delta T(\hat{n})$ relative to the background temperature $T = 2.7K$ is given by

$$\frac{\Delta T(\vec{n})}{T} = \sum_{lm} a_{lm} \hat{Y}_{lm}(\hat{n}). \tag{5.456}$$

$$a_{lm} = \int d\Omega Y^*_{lm}(\vec{n}) \frac{\Delta T(\vec{n})}{T}. \tag{5.457}$$

The two-point correlator $< a_{l_1 m_1} a_{l_2 m_2} >$ must behave (by rotational invariance) as

$$< a^*_{l_1 m_1} a_{l_2 m_2} >= C_l^{TT} \delta_{l_1 l_2} \delta_{m_1 m_2}. \tag{5.458}$$

The rotationally invariant angular power spectrum C_l^{TT} is given by

$$C_l^{TT} = \frac{1}{2l+1} \sum_m < a^*_{l_1 m_1} a_{l_2 m_2} > . \tag{5.459}$$

For values of the tensor-to-scalar ratio $r < 0.3$ the CMB temperature fluctuations are dominated by the scalar curvature perturbation $\mathcal{R}$. We have already computed the curvature perturbation at horizon crossing (exit) which then remains constant (freeze at a constant value) until the time of re-entry. From the time of re-entry until the time of CMB recombination the curvature perturbation will evolve in time causing a temperature fluctuation. The temperature fluctuation we observe today as a remnant of last scattering (CMB recombination) is encoded in the multipole moments a_{lm} and is related to the scalar curvature perturbation $\mathcal{R}_k$ at the time of horizon crossing $k = a(t_*)H(t_*)$ through a transfer function $\Delta_{Tl}(k)$ as follows [3]

$$a_{lm} = 4\pi(-i)^l \int \frac{d^3k}{(2\pi)^3} \Delta_{Tl}(k) \mathcal{R}_k Y_{lm}(\vec{k}). \tag{5.460}$$

In the quantum theory $\mathcal{R}_k$ become operators and hence a_{lm} become operators. We compute immediately (with $\mathcal{R}^*_k = \mathcal{R}_{-k}$)

$$\begin{aligned}
\sum_m < \hat{a}^+_{lm} \hat{a}_{lm} > &= (4\pi)^2 \sum_m \int \frac{d^3k}{(2\pi)^3} \Delta_{Tl}(k) Y^*_{lm}(\vec{k}) \int \frac{d^3k'}{(2\pi)^3} \Delta_{Tl}(k') Y_{lm}(\vec{k}') < \hat{\mathcal{R}}^+_k \hat{\mathcal{R}}_{k'} > \\
&= (4\pi)^2 \int \frac{d^3k}{(2\pi)^3} \Delta^2_{Tl}(k) P_{\mathcal{R}}(k) \sum_m Y^*_{lm}(\vec{k}) Y_{lm}(\vec{k}) \\
&= (4\pi)^2 \int \frac{d^3k}{(2\pi)^3} \Delta^2_{Tl}(k) P_{\mathcal{R}}(k) \frac{2l+1}{4\pi} P_l(\hat{k}^2) \\
&= \frac{2}{\pi}(2l+1) \int k^2 dk \Delta^2_{Tl}(k) P_{\mathcal{R}}(k).
\end{aligned} \tag{5.461}$$

Hence

$$C_l^{TT} = \frac{2}{\pi} \int k^2 dk \Delta^2_{Tl}(k) P_{\mathcal{R}}(k). \tag{5.462}$$

[3] Exercise: Derive this equation. Very difficult. A considerable amount of reading is required.

The term $\Delta^2_{Tl}(k)$ is the anisotropies term.

For large scales, i.e. large k^{-1} we can safely assume that the modes were still outside the horizon at the time of recombination. As a consequence the large scale CMB spectrum is only affected by the geometric projection from recombination to our current epoch and is not affected by sub-horizon evolution. This is the so-called Sachs-Wolf regime in which the transfer function is a Bessel function, viz [4]

$$\Delta^2_{Tl}(k) = \frac{1}{3} j_l(k(\eta_0 - \eta_{\rm rec})) + ... \tag{5.463}$$

This term is the monopole contribution to the transfer function. We have neglected a dipole term and the so-called integrated Sachs-Wolfe (ISW) terms.

The Bessel function essentially projects the linear scales with wavenumber k onto angular scales with angular wavenumber l. The angular power spectrum C_l^{TT} on large scale (corresponding to small l or large angles) is therefore

$$\begin{aligned} C_l^{TT} &= \frac{2}{9\pi}\int k^2 dk j_l^2(k(\eta_0 - \eta_{\rm rec}))P_s(k) \\ &= \frac{4\pi}{9}\int \frac{dk}{k} j_l^2(k(\eta_0 - \eta_{\rm rec}))\Delta_s^2(k). \end{aligned} \tag{5.464}$$

The Bessel function for large l acts effectively as a delta function since it is peaked around

$$l = k(\eta_0 - \eta_{\rm rec}). \tag{5.465}$$

We approximate the dimensionless power spectrum $\Delta_s^2(k)$ by the following power law (where n_s is the spectral index evaluated at some reference point k_*)

$$\Delta_s^2(k) = A_s k^{n_s - 1}. \tag{5.466}$$

We obtain then

$$\begin{aligned} C_l^{TT} &= \frac{4\pi}{9}A_s \int \frac{dk}{k^{2-n_s}} j_l^2(k(\eta_0 - \eta_{\rm rec})) \\ &= \frac{4\pi}{9}A_s(\eta_0 - \eta_{\rm rec})^{1-n_s}\int \frac{dx}{x^{2-n_s}} j_l^2(x) \\ &= 2^{n_s - 4}\frac{4\pi^2}{9}A_s(\eta_0 - \eta_{\rm rec})^{1-n_s}\frac{\Gamma(l + \frac{n_s}{2} - \frac{1}{2})}{\Gamma(l - \frac{n_s}{2} + \frac{5}{2})}\frac{\Gamma(3 - n_s)}{\Gamma^2(2 - \frac{n_s}{2})}. \end{aligned} \tag{5.467}$$

For a scale-invariant spectrum we have $n_s = 1$. In this case

$$\begin{aligned} C_l &\equiv \frac{l(l+1)}{2\pi} C_l^{TT} \\ &= \frac{A_s}{9}. \end{aligned} \tag{5.468}$$

[4] Exercise: Derive this equation. This is related to the previous question.

The modified power spectrum C_l is therefore independent of l for small values of l corresponding to the largest scales (largest angles). This is what is observed in the real world. See figure 8.12 of [12]. Thus we conclude that n_s must be indeed very close to 1.

The situation is more involved for intermediate scales where acoustic peaks dominate and for small scales where damping dominates which is an effect due to photon diffusion. The acoustic peaks arise because the early universe was a plasma of photons and baryons forming a single fluid which can oscillate due to the competing forces of radiation pressure and gravitational compression. This struggle between gravity and radiation pressure is what sets up longitudinal acoustic oscillations in the photon-baryon fluid. At recombination the pattern of acoustic oscillations became frozen into the CMB which is what we see today as peaks and troughs in the power spectrum of temperature fluctuations. A proper study of the acoustic peaks seen at intermediate scales and also of the damping seen at small scales is beyond our means at this point.

In conclusion the predictions of cosmological scalar perturbation theory for the angular power spectrum of CMB temperature anisotropies agrees very well with observations. See for example figure 10 of [21].

Chapter 6

QFT on Curved Backgrounds and Vacuum Energy

6.1 Dark Energy

It is generally accepted now that there is a positive dark energy in the universe which affects in measurable ways the physics of the expansion. The characteristic feature of dark energy is that it has a negative pressure (tension) smoothly distributed in spacetime so it was proposed that a name like "smooth tension" is more appropriate to describe it (see reference [11]). The most dramatic consequence of a non zero value of Ω_Λ is the observation that the universe appears to be accelerating.

From an observational point of view astronomical evidence for dark energy comes from various measurements. Here we concentrate, and only briefly, on the the two measurements of CMB anisotropies and type Ia supernovae.

- CMB Anisotropies: This point will be discussed in more detail later from a theoretical point of view. The main point is as follows. The temperature anisotropies are given by the power spectrum C_l. At intermediate scales (angular scales subtended by $H_{\rm CMB}^{-1}$ where $H_{\rm CMB}$ is the Hubble radius at the time of the formation of the cosmic microwave background (decoupling, recombination, last scattering)) we observe peaks in C_l due to acoustic oscillations in the early universe. The first peak is tied directly to the geometry of the universe. In a negatively curved universe photon paths diverge leading to a larger apparent angular size compared to flat space whereas in a positively curved universe photon paths converge leading to a smaller apparent angular size compared to flat space. The spatial curvature as measured by Ω is related to the first peak in the CMB power spectrum by

$$l_{\rm peak} \sim \frac{220}{\sqrt{\Omega}}. \tag{6.1}$$

  The observation indicates that the first peak occurs around $l_{\rm peak} \sim 200$ which means that the universe is spatially flat. The Boomerang experiment gives (at the 68 per cent confidence level) the measurement

$$0.85 \leq \Omega \leq 1.25. \tag{6.2}$$

Since $\Omega = \Omega_M + \Omega_\Lambda$ this is a constraint on the sum of Ω_M and Ω_Λ. The constraints from the CMB in the $\Omega_M - \Omega_\Lambda$ plane using models with different values of Ω_M and Ω_Λ is shown on figure 3 of reference [26]. The best fit is a marginally closed model with

$$\Omega_{\rm CDM} = 0.26\ ,\ \Omega_{\rm B} = 0.05\ ,\ \Omega_\Lambda = 0.75. \tag{6.3}$$

- **Type Ia Supernovae:** This relies on the measurement of the distance modulus $m - M$ of type Ia supernovae where m is the apparent magnitude of the source and M is the absolute magnitude defined by

$$m - M = 5\log_{10}[(1+z)d_M(Mpc)] + 25. \tag{6.4}$$

The d_M is the proper distance which is given between any two sources at redshifts z_1 and z_2 by the formula

$$d_M(z_1, z_2) = \frac{1}{H_0\sqrt{|\Omega_{k0}|}} S_k\bigg(H_0\sqrt{|\Omega_{k0}|}\int_{1/(1+z_1)}^{1/(1+z_2)} \frac{da}{a^2 H(a)}\bigg). \tag{6.5}$$

Type Ia supernovae are rare events which thought of as standard candles. They are very bright events with almost uniform intrinsic luminosity with absolute brightness comparable to the host galaxies. They result from exploding white dwarfs when they cross the Chandrasekhar limit.

Constraints from type Ia supernovae in the $\Omega_M - \Omega_\Lambda$ plane are consistent with the results obtained from the CMB measurements although the data used is completely independent. In particular these observations strongly favors a positive cosmological constant.

6.2 The Cosmological Constant

The cosmological constant was introduce by Einstein in 1917 in order to produce a static universe. To see this explicitly let us rewrite the Friedmann equations (**??**) and (**??**) as

$$H^2 = \frac{8\pi G\rho}{3} - \frac{\kappa}{a^2}. \tag{6.6}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P). \tag{6.7}$$

The first equation is consistent with a static universe ($\dot{a} = 0$) if $\kappa > 0$ and $\rho = 3\kappa/(8\pi G a^2)$ whereas the second equation can not be consistent with a static universe ($\ddot{a} = 0$) containing only ordinary matter and energy which have non negative pressure.

Einstein solved this problem by modifying his equations as follows

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R + \Lambda g_{\mu\nu} = 8\pi G T_{\mu\nu}. \tag{6.8}$$

The new free parameter Λ is precisely the cosmological constant. This new equations of motion will entail a modification of the Friedmann equations. To find the modified Friedmann equations we rewrite the modified Einstein's equations as

$$R_{\mu\nu} - \frac{1}{2}g_{\mu\nu}R = 8\pi G(T_{\mu\nu} + T^{\Lambda}_{\mu\nu}). \tag{6.9}$$

$$T^{\Lambda}_{\mu\nu} = -\rho_{\Lambda} g_{\mu\nu}\ ,\ \rho_{\Lambda} = \frac{\Lambda}{8\pi G}. \tag{6.10}$$

The modified Friedmann equations are then given by (with the substitution $\rho \longrightarrow \rho + \rho_{\Lambda}$, $P \longrightarrow P - \rho_{\Lambda}$ in the original Friedmann equations)

$$H^2 = \frac{8\pi G(\rho + \rho_{\Lambda})}{3} - \frac{\kappa}{a^2} = \frac{8\pi G\rho}{3} - \frac{\kappa}{a^2} + \frac{\Lambda}{3}. \tag{6.11}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho - 2\rho_{\Lambda} + 3P) = -\frac{4\pi G}{3}(\rho + 3P) + \frac{\Lambda}{3}. \tag{6.12}$$

The Einstein static universe corresponds to $\kappa > 0$ (a $3-$sphere S^3) and $\Lambda > 0$ (in the range $\kappa/a^2 \leq \Lambda \leq 3\kappa/a^2$) with positive mass density and pressure given by

$$\rho = \frac{3\kappa}{8\pi G a^2} - \frac{\Lambda}{8\pi G} > 0\ ,\ P = \frac{\Lambda}{8\pi G} - \frac{\kappa}{8\pi G a^2} > 0. \tag{6.13}$$

The universe is in fact expanding and thus this solution is of no physical interest. The cosmological constant is however of fundamental importance to cosmology as it might be relevant to dark energy.

It is not difficult to verify that the modified Einstein's equations (6.8) can be derived from the action

$$S = \frac{1}{16\pi G}\int d^4x\sqrt{-\det g}\ (R - 2\Lambda) + \int d^4x\sqrt{-\det g}\ \hat{\mathcal{L}}_M. \tag{6.14}$$

Thus the cosmological constant Λ is just a constant term in the Lagrangian density. We call Λ the bare cosmological constant. The effective cosmological constant $\Lambda_{\rm eff}$ will in general be different from Λ due to possible contribution from matter. Consider for example a scalar field with Lagrangian density

$$\hat{\mathcal{L}}_M = -\frac{1}{2}g^{\mu\nu}\nabla_{\mu}\phi\nabla_{\nu}\phi - V(\phi). \tag{6.15}$$

The stress-energy-momentum tensor is calculated to be given by

$$T_{\mu\nu} = \nabla_\mu \phi \nabla_\nu \phi - \frac{1}{2} g_{\mu\nu} g^{\rho\sigma} \nabla_\rho \phi \nabla_\sigma \phi - g_{\mu\nu} V(\phi). \tag{6.16}$$

The configuration ϕ_0 with lowest energy density (the vacuum) is the contribution which minimizes separately the kinetic and potential terms and as a consequence $\partial_\mu \phi_0 = 0$ and $V^{'}(\phi_0) = 0$. The corresponding stress-energy-momentum tensor is therefore $T_{\mu\nu}^{(\phi)} = -g_{\mu\nu} V(\phi_0)$. In other words the stress-energy-momentum tensor of the vacuum acts precisely like the stress-energy-momentum tensor of a cosmological constant. We write (with $T_{\mu\nu}^{(\phi_0)} \equiv T_{\mu\nu}^{\rm vac}$, $V(\phi_0) \equiv \rho_{\rm vac}$)

$$T_{\mu\nu}^{\rm vac} = -\rho_{\rm vac} g_{\mu\nu}. \tag{6.17}$$

The vacuum ϕ_0 is therefore a perfect fluid with pressure given by

$$P_{\rm vac} = -\rho_{\rm vac}. \tag{6.18}$$

Thus the vacuum energy acts like a cosmological constant Λ_ϕ given by

$$\Lambda_\phi = 8\pi G \rho_{\rm vac}. \tag{6.19}$$

In other words the cosmological constant and the vacuum energy are completely equivalent. We will use the two terms "cosmological constant" and "vacuum energy" interchangeably.

The effective cosmological constant $\Lambda_{\rm eff}$ is therefore given by

$$\Lambda_{\rm eff} = \Lambda + \Lambda_\phi. \tag{6.20}$$

In other words

$$\Lambda_{\rm eff} = \Lambda + 8\pi G \rho_{\rm vac}. \tag{6.21}$$

This calculation is purely classical.

Quantum mechanics will naturally modify this result. We follow a semi-classical approach in which the gravitational field is treated classically and the scalar field (matter fields in general) are treated quantum mechanically. Thus we need to quantize the scalar field in a background metric $g_{\mu\nu}$ which is here the Robertson-Walker metric. In the quantum vacuum state of the scalar field (assuming that it exists) the expectation value of the stress-energy-momentum tensor $T_{\mu\nu}$ must be, by Lorentz invariance, of the form

$$< T_{\mu\nu} >_{\rm vac} = - < \rho >_{\rm vac} g_{\mu\nu}. \tag{6.22}$$

The Einstein's equations in the vacuum state of the scalar field are

$$R_{\mu\nu} - \frac{1}{2} g_{\mu\nu} R + \Lambda g_{\mu\nu} = 8\pi G < T_{\mu\nu} >_{\rm vac} . \tag{6.23}$$

The effective cosmological constant $\Lambda_{\rm eff}$ must therefore be given by

$$\Lambda_{\rm eff}=\Lambda+8\pi G<\rho>_{\rm vac}. \tag{6.24}$$

The energy density of empty space $<\rho>_{\rm vac}$ is the sum of zero-point energies associated with vacuum fluctuations together with other contributions resulting from virtual particles (higher order vacuum fluctuations) and vacuum condensates.

We will assume from simplicity that the bare cosmological constant Λ is zero. Thus the effective cosmological constant is entirely given by vacuum energy, viz

$$\Lambda_{\rm eff}=8\pi G<\rho>_{\rm vac}. \tag{6.25}$$

We drop now the subscript "eff" without fear of confusion. The relation between the density ρ_Λ of the cosmological constant and the density $<\rho>_{\rm vac}$ of the vacuum is then simply

$$\rho_\Lambda=<\rho>_{\rm vac}. \tag{6.26}$$

From the concordance model we know that the favorite estimate for the value of the density parameter of dark energy at this epoch is $\Omega_\Lambda=0.7$. We recall $G=6.67\times 10^{-11}m^3kg^{-1}s^{-2}$ and $H_0=70\ kms^{-1}Mpc^{-1}$ with $Mpc=3.09\times 10^{24}cm$. We compute then the density

$$\begin{aligned}\rho_\Lambda &= \frac{3H_0^2}{8\pi G}\Omega_\Lambda\\ &= 9.19\times 10^{-27}\Omega_\Lambda kg/m^3.\end{aligned} \tag{6.27}$$

We convert to natural units ($1GeV=1.8\times 10^{-27}kg$, $1GeV^{-1}=0.197\times 10^{-15}m$, $1GeV^{-1}=6.58\times 10^{-25}s$) to obtain

$$\rho_\Lambda = 39\Omega_\Lambda(10^{-12}GeV)^4. \tag{6.28}$$

To get a theoretical order-of-magnitude estimate of $<\rho>_{\rm vac}$ we use the flat space Hamiltonian operator of a free scalar field given by

$$\hat{H} = \int\frac{d^3p}{(2\pi)^3}\omega(\vec{p})\bigg[\hat{a}(\vec{p})^+\hat{a}(\vec{p})+\frac{1}{2}(2\pi)^3\delta^3(0)\bigg]. \tag{6.29}$$

The vacuum state is defined in this case unambiguously by $\hat{a}(\vec{p})|0>=0$. We get then in the vacuum state the energy $E_{\rm vac}=<0|\hat{H}|0>$ where

$$E_{\rm vac} = \frac{1}{2}(2\pi)^3\delta^3(0)\int\frac{d^3p}{(2\pi)^3}\omega(\vec{p}). \tag{6.30}$$

If we use box normalization then $(2\pi)^3\delta^3(\vec{p}-\vec{q})$ will be replaced with $V\delta_{\vec{p},\vec{q}}$ where V is spacetime volume. The vacuum energy density is therefore given by (using also $\omega(\vec{p})=\sqrt{\vec{p}^2+m^2}$)

$$< \rho >_{\rm vac} = \frac{1}{2}\int \frac{d^3p}{(2\pi)^3}\sqrt{\vec{p}^2+m^2}. \tag{6.31}$$

This is clearly divergent. We introduce a cutoff λ and compute

$$\begin{aligned} < \rho >_{\rm vac} &= \frac{1}{4\pi^2}\int_0^{\lambda} dpp^2\sqrt{p^2+m^2} \\ &= \frac{1}{4\pi^2}\bigg[\bigg(\frac{1}{4}\lambda^3+\frac{m^2}{8}\lambda\bigg)\sqrt{\lambda^2+m^2}-\frac{m^4}{8}\ln\bigg(\frac{\lambda}{m}+\sqrt{1+\frac{\lambda^2}{m^2}}\bigg)\bigg]. \end{aligned} \tag{6.32}$$

In the massless limit (the mass is in any case much smaller than the cutoff λ) we obtain the estimate

$$< \rho >_{\rm vac} = \frac{\lambda^4}{16\pi^2}. \tag{6.33}$$

By assuming that quantum field theory calculations are valid up to the Planck scale $M_{\rm pl} = 1/\sqrt{8\pi G} = 2.42\times 10^{18}GeV$ then we can take $\lambda = M_{\rm pl}$ and get the estimate

$$< \rho >_{\rm vac} = 0.22(10^{18}GeV)^4. \tag{6.34}$$

By taking the ratio of the value (6.28) obtained from cosmological observations and the theoretical value (6.34) we get

$$(\frac{\rho_\Lambda}{< \rho >_{\rm vac}})^{1/4} = 3.65\times\Omega_\Lambda^{1/4}\times 10^{30}. \tag{6.35}$$

For the observed value $\Omega_\Lambda = 0.7$ we see that there is a discrepancy of 30 orders of magnitude between the theoretical and observational mass scales of the vacuum energy which is the famous cosmological constant problem.

Let us note that in flat spacetime we can make the vacuum energy vanishes by the usual normal ordering procedure which reflects the fact that only differences in energy have experimental consequences in this case. In curved spacetime this is not however possible since general relativity is sensitive to the absolute value of the vacuum energy. In other words the gravitational effect of vacuum energy will curve spacetime and the above problem of the cosmological constant is certainly genuine.

6.3 Calculation of Vacuum Energy in Curved Backgrounds

6.3.1 Elements of Quantum Field Theory in Curved Spacetime

Let us start by writing Friedmann equations with a cosmological constant Λ which are given by (with $H = \dot{a}/a$)

$$H^2=\frac{8\pi G\rho}{3}-\frac{\kappa}{a^2}+\frac{\Lambda}{3}. \tag{6.36}$$

$$\frac{\ddot{a}}{a}=-\frac{4\pi G}{3}(\rho+3P)+\frac{\Lambda}{3}. \tag{6.37}$$

We will assume that ρ and P are those of a real scalar field coupled to the metric minimally with action given by

$$S_M=\int d^4x\sqrt{-\mathrm{det}g}\,\bigg(-\frac{1}{2}g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi-V(\phi)\bigg). \tag{6.38}$$

If we are interested in an action which is at most quadratic in the scalar field then we must choose $V(\phi)=m^2\phi^2/2$. In curved spacetime there is another term we can add which is quadratic in ϕ namely $R\phi^2$ where R is the Ricci scalar. The full action should then read (in arbitrary dimension n)

$$S_M=\int d^nx\sqrt{-\mathrm{det}g}\,\bigg(-\frac{1}{2}g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi-\frac{1}{2}m^2\phi^2-\frac{1}{2}\xi R\phi^2\bigg). \tag{6.39}$$

The choice $\xi=(n-2)/(4(n-1))$ is called conformal coupling. At this value the action with $m^2=0$ is invariant under conformal transformations defined by [1]

$$g_{\mu\nu}\longrightarrow\bar{g}_{\mu\nu}=\Omega^2(x)g_{\mu\nu}(x)\ ,\ \phi\longrightarrow\bar{\phi}=\Omega^{\frac{2-n}{2}}(x)\phi(x). \tag{6.40}$$

The equation of motion derived from this action are (we will keep in the following the metric arbitrary as long as possible)

$$\big(\nabla_\mu\nabla^\mu-m^2-\xi R\big)\phi=0. \tag{6.41}$$

Let ϕ_1 and ϕ_2 be two solutions of this equation of motion. We define their inner product by

$$(\phi_1,\phi_2)=-i\int_\Sigma\big(\phi_1\partial_\mu\phi_2^*-\partial_\mu\phi_1.\phi_2^*\big)d\Sigma n^\mu. \tag{6.42}$$

$d\Sigma$ is the volume element in the space like hypersurface Σ and n^μ is the time like unit vector which is normal to this hypersurface. This inner product is independent of the hypersurface Σ. Indeed let Σ_1 and Σ_2 be two non intersecting hypersurfaces and let V be the four-volume bounded by Σ_1, Σ_2 and (if necessary) time like boundaries on which $\phi_1=\phi_2=0$. We have from one hand

$$\begin{aligned}i\int_V\nabla^\mu\big(\phi_1\partial_\mu\phi_2^*-\partial_\mu\phi_1.\phi_2^*\big)dV &= i\oint_{\partial V}\big(\phi_1\partial_\mu\phi_2^*-\partial_\mu\phi_1.\phi_2^*\big)d\Sigma^\mu\\ &= (\phi_1,\phi_2)_{\Sigma_1}-(\phi_1,\phi_2)_{\Sigma_2}.\end{aligned} \tag{6.43}$$

[1] Exercise: Show this result.

From the other hand

$$i\int_V \nabla^\mu\big(\phi_1\partial_\mu\phi_2^* - \partial_\mu\phi_1.\phi_2^*\big)dV = i\int_V \big(\phi_1\nabla^\mu\partial_\mu\phi_2^* - \nabla^\mu\partial_\mu\phi_1.\phi_2^*\big)dV$$
$$= i\int_V \big(\phi_1(m^2+\xi R)\phi_2^* - (m^2+\xi R)\phi_1.\phi_2^*\big)dV$$
$$= 0. \quad (6.44)$$

Hence

$$(\phi_1,\phi_2)_{\Sigma_1} - (\phi_1,\phi_2)_{\Sigma_2} = 0. \quad (6.45)$$

There is always a complete set of solutions u_i and u_i^* of the equation of motion (6.41) which are orthonormal in the inner product (6.42), i.e. satisfying

$$(u_i,u_j) = \delta_{ij}\ ,\ (u_i^*,u_j^*) = -\delta_{ij}\ ,\ (u_i,u_j^*) = 0. \quad (6.46)$$

We can then expand the field as

$$\phi = \sum_i (a_iu_i + a_i^*u_i^*). \quad (6.47)$$

We now canonically quantize this system. We choose a foliation of spacetime into space like hypersurfaces. Let Σ be a particular hypersurface with unit normal vector n^μ corresponding to a fixed value of the time coordinate $x^0=t$ and with induced metric h_{ij}. We write the action as $S_M = \int dx^0 L_M$ where $L_M = \int d^{n-1}x\sqrt{-\text{det}g}\ \mathcal{L}_M$. The canonical momentum π is defined by [2]

$$\pi = \frac{\delta L_M}{\delta(\partial_0\phi)} = -\sqrt{-\text{det}g}\ g^{\mu 0}\partial_\mu\phi$$
$$= -\sqrt{-\text{det}h}\ n^\mu\partial_\mu\phi. \quad (6.48)$$

We promote ϕ and π to hermitian operators $\hat{\phi}$ and $\hat{\pi}$ and then impose the equal time canonical commutation relations

$$[\hat{\phi}(x^0,x^i),\hat{\pi}(x^0,y^i)] = i\delta^{n-1}(x^i-y^i). \quad (6.49)$$

The delta function satisfies the property

$$\int \delta^{n-1}(x^i-y^i)d^{n-1}y = 1. \quad (6.50)$$

The coefficients a_i and a_i^* become annihilation and creation operators $\hat{a}_i$ and $\hat{a}_i^+$ satisfying the commutation relations [3]

$$[\hat{a}_i,\hat{a}_j^+] = \delta_{ij}\ ,\ [\hat{a}_i,\hat{a}_j] = [\hat{a}_i^+,\hat{a}_j^+] = 0. \quad (6.51)$$

[2]Exercise: Show the second line of this equation.

[3]Exercise: Show this explicitly.

The vacuum state is given by a state $|0>_u$ defined by

$$\hat{a}_i|0_u>=0. \tag{6.52}$$

The entire Fock basis of the Hilbert space can be constructed from the vacuum state by repeated application of the creation operators $\hat{a}_i^+$.

The solutions u_i, u_i^* are not unique and as a consequence the vacuum state $|0>_u$ is not unique. Let us consider another complete set of solutions v_i and v_i^* of the equation of motion (6.41) which are orthonormal in the inner product (6.42). We can then expand the field as

$$\phi=\sum_i(b_iv_i+b_i^*v_i^*). \tag{6.53}$$

After canonical quantization the coefficients b_i and b_i^* become annihilation and creation operators $\hat{b}_i$ and $\hat{b}_i^+$ satisfying the standard commutation relations with a vacuum state given by $|0>_v$ defined by

$$\hat{b}_i|0_v>=0. \tag{6.54}$$

We introduce the so-called Bogolubov transformation as the transformation from the set $\{u_i,u_i^*\}$ (which are the set of modes seen by some observer) to the set $\{v_i,v_i^*\}$ (which are the set of modes seen by another observer) as

$$v_i=\sum_j(\alpha_{ij}u_j+\beta_{ij}u_j^*). \tag{6.55}$$

By using orthonormality conditions we find that

$$\alpha_{ij}=(v_i,u_j)\ ,\ \beta_{ij}=-(v_i,u_j^*). \tag{6.56}$$

We can also write

$$u_i=\sum_j(\alpha_{ji}^*v_j+\beta_{ji}v_j^*). \tag{6.57}$$

The Bogolubov coefficients α and β satisfy the normalization conditions

$$\sum_k(\alpha_{ik}\alpha_{jk}-\beta_{ik}\beta_{jk})=\delta_{ij}\ ,\ \sum_k(\alpha_{ik}\beta_{jk}^*-\beta_{ik}\alpha_{jk}^*)=0. \tag{6.58}$$

The Bogolubov coefficients α and β transform also between the creation and annihilation operators $\hat{a}$, $\hat{a}^+$ and $\hat{b}$, $\hat{b}^+$. We find

$$\hat{a}_k=\sum_i(\alpha_{ik}\hat{b}_i+\beta_{ik}^*\hat{b}_i^+)\ ,\ \hat{b}_k=\sum_i(\alpha_{ki}^*\hat{a}_i+\beta_{ki}^*\hat{a}_i^+). \tag{6.59}$$

Let N_u be the number operator with respect to the u-observer, viz $N_u=\sum_k\hat{a}_k^+\hat{a}_k$. Clearly

$$<0_u|N_u|0_u>=0. \tag{6.60}$$

We compute

$$< 0_v|\hat{a}_k^+\hat{a}_k|0_v >=\sum_i \beta_{ik}\beta_{ik}^*. \tag{6.61}$$

Thus

$$< 0_v|N_u|0_v >= tr\beta\beta^+. \tag{6.62}$$

In other words with respect to the v-observer the vacuum state $|0_u>$ is not empty but filled with particles. This opens the door to the possibility of particle creation by a gravitational field.

6.3.2 Quantization in FLRW Universes

We go back to the equation of motion (6.41), viz

$$\big(\nabla_\mu\nabla^\mu - m^2 - \xi R\big)\phi = 0. \tag{6.63}$$

The flat FLRW universes are given by

$$ds^2 = -dt^2 + a^2(t)(d\rho^2 + \rho^2 d\Omega^2). \tag{6.64}$$

The conformal time is denoted here by

$$\eta = \int^t \frac{dt_1}{a(t_1)}. \tag{6.65}$$

In terms of η the FLRW universes are manifestly conformally flat, viz

$$ds^2 = a^2(\eta)(-d\eta^2 + d\rho^2 + \rho^2 d\Omega^2). \tag{6.66}$$

The d'Alembertian in FLRW universes is

$$\begin{aligned}\nabla_\mu\nabla^\mu\phi &= \frac{1}{\sqrt{-\det g}}\partial_\mu(\sqrt{-\det g}\partial^\mu\phi)\\ &= \partial_\mu\partial^\mu\phi + \frac{1}{2}g^{\alpha\beta}\partial_\mu g_{\alpha\beta}\partial^\mu\phi\\ &= -\ddot{\phi} + \frac{1}{a^2}\partial_i^2\phi - 3\frac{\dot{a}}{a}\dot{\phi}.\end{aligned} \tag{6.67}$$

The Klein-Gordon equation of motion becomes

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} - \frac{1}{a^2}\partial_i^2\phi + (m^2 + \xi R)\phi = 0. \tag{6.68}$$

In terms of the conformal time this reads (where $d/d\eta$ is denoted by primes)

$$\phi^{''} + 2\frac{a^{'}}{a}\phi^{'} - \partial_i^2\phi + a^2(m^2 + \xi R)\phi = 0. \tag{6.69}$$

The positive norm solutions are given by

$$u_k(\eta, x^i) = \frac{e^{i\vec{k}\vec{x}}}{a(\eta)}\chi_k(\eta). \tag{6.70}$$

Indeed we check that $\phi \equiv u_k(\eta, x^i)$ is a solution of the Klein-Gordon equation of motion provided that χ_k is a solution of the equation of motion (using also $R = 6(\ddot{a}/a + \dot{a}^2/a^2) = 6a^{''}/a^3$)

$$\chi_k^{''} + \omega_k^2(\eta)\chi_k = 0. \tag{6.71}$$

$$\omega_k^2(\eta) = k^2 + m^2a^2 - (1-6\xi)\frac{a^{''}}{a}. \tag{6.72}$$

In the case of conformal coupling $m = 0$ and $\xi = 1/6$ this reduces to a time independent harmonic oscillator. This is similar to flat spacetime and all effects of the curvature are included in the factor $a(\eta)$ in equation (6.70). Thus calculation in a conformally invariant world is very easy.

The condition $(u_k, u_l) = \delta_{kl}$ becomes (with $n^\mu = (1,0,0,0)$, $d\Sigma = \sqrt{-\det h}\ d^3x$ and using box normalization $(2\pi)^3\delta^3(\vec{k}-\vec{p}) \longrightarrow V\delta_{\vec{k},\vec{p}}$ the Wronskian condition

$$iV(\chi_k^*\chi_k^{'} - \chi_k^{*'}\chi_k) = 1. \tag{6.73}$$

The negative norm solutions correspond obviously to u_k^*. Indeed we can check that $(u_k^*, \bar{u}_l) = -\delta_{kl}$ and $(u_k^*, u_l) = 0$.

The modes u_k and $\bar{u}_k$ provide a Fock space representation for field operators. The quantum field operator $\hat{\phi}$ can be expanded in terms of creation and annhiliation operators as

$$\hat{\phi} = \sum_k (\hat{a}_k u_k + \hat{a}_k^+ u_k^*). \tag{6.74}$$

Alternatively the mode functions satisfy the differential equations (with $\chi_k = v_k^*/\sqrt{2V}$)

$$v_k^{''} + \omega_k^2(\eta)v_k = 0 \tag{6.75}$$

They must satisfy the normalization condition

$$\frac{1}{2i}(v_k^{'}v_k^* - v_k v_k^{*'}) = 1. \tag{6.76}$$

The scalar field operator is given by $\hat{\phi} = \hat{\chi}/a(\eta)$ where (with $[\bar{a}_k, \bar{a}_{k'}^+] = V\delta_{k,k'}$, etc)

$$\hat{\chi} = \frac{1}{V}\sum_k \frac{1}{\sqrt{2}}\bigg(\bar{a}_k v_k^* e^{i\vec{k}\vec{x}} + \bar{a}_k^+ v_k e^{-i\vec{k}\vec{x}}\bigg). \tag{6.77}$$

The stress-energy-momentum tensor in minimal coupling $\xi = 0$ is given by

$$T_{\mu\nu} = \nabla_\mu\phi\nabla_\nu\phi - \frac{1}{2}g_{\mu\nu}g^{\rho\sigma}\nabla_\rho\phi\nabla_\sigma\phi - g_{\mu\nu}V(\phi). \tag{6.78}$$

We compute immediately in the conformal metric $ds^2 = a^2(-d\eta^2 + dx^i dx^i)$ the component

$$\begin{aligned} T_{00} &= \frac{1}{2}(\partial_\eta\phi)^2 + \frac{1}{2}(\partial_i\phi)^2 + \frac{1}{2}a^2m^2\phi^2 \\ &= \frac{1}{2a^2}\big[\chi^{'2} - 2\frac{a^{'}}{a}\chi\chi^{'} + \frac{a^{'2}}{a^2}\chi^2\big] + \frac{1}{2a^2}(\partial_i\chi)^2 + \frac{1}{2}m^2\chi^2. \end{aligned} \tag{6.79}$$

The conjugate momentum (6.48) in our case is $\pi = a^2\partial_\eta\phi$. The Hamiltonian is therefore

$$\begin{aligned} H &= \int d^{n-1}x\ \pi\partial_0\phi - L_M \\ &= \int d^{n-1}x\sqrt{-\det g}\ \frac{1}{a^2}T_{00} \\ &= -\int d^{n-1}x\sqrt{-\det g}\ T_0{}^0. \end{aligned} \tag{6.80}$$

In the quantum theory the stress-energy-momentum tensor in minimal coupling $\xi = 0$ is given by

$$\hat{T}_{00} = \frac{1}{2a^2}\big[\hat{\chi}^{'2} - \frac{a^{'}}{a}(\hat{\chi}\hat{\chi}^{'} + \hat{\chi}^{'}\hat{\chi}) + \frac{a^{'2}}{a^2}\hat{\chi}^2\big] + \frac{1}{2a^2}(\partial_i\hat{\chi})^2 + \frac{1}{2}m^2\hat{\chi}^2. \tag{6.81}$$

We assume the existence of a vacuum state $|0>$ with the properties $a|0>=0$, $<0|a^+=0$ and $<0|0>=1$. We compute

$$\begin{aligned} <\hat{\chi}^{'2}> &= \frac{1}{2V^2}\sum_k\sum_p v_k^{*'}v_p^{'}e^{i\vec{k}\vec{x}}e^{-i\vec{p}\vec{x}}<0|\bar{a}_k\bar{a}_p^+|0> \\ &= \frac{1}{2V}\sum_k|v_k^{'}|^2. \end{aligned} \tag{6.82}$$

$$\begin{aligned} <\hat{\chi}^2> &= \frac{1}{2V^2}\sum_k\sum_p v_k^*v_pe^{i\vec{k}\vec{x}}e^{-i\vec{p}\vec{x}}<0|\bar{a}_k\bar{a}_p^+|0> \\ &= \frac{1}{2V}\sum_k|v_k|^2. \end{aligned} \tag{6.83}$$

$$\begin{aligned} <(\partial_i\hat{\chi})^2> &= \frac{1}{2V^2}\sum_k\sum_p v_k^*v_p(k_ip_i)e^{i\vec{k}\vec{x}}e^{-i\vec{p}\vec{x}}<0|\bar{a}_k\bar{a}_p^+|0> \\ &= \frac{1}{2V}\sum_k k^2|v_k|^2. \end{aligned} \tag{6.84}$$

We get then

$$< \hat{T}_{00} > = \frac{1}{2a^2}\frac{1}{2V}\sum_k\bigg[|v_k^{'}|^2-\frac{a^{'}}{a}(v_k^*v_k^{'}+v_k^{'*}v_k)+\frac{a^{'2}}{a^2}|v_k|^2+k^2|v_k|^2+a^2m^2|v_k|^2\bigg]$$
$$= \frac{1}{4a^2}\frac{1}{V}\sum_k\bigg[|v_k^{'}|^2+(k^2+\frac{a^{''}}{a}+a^2m^2)|v_k|^2-\partial_\eta(\frac{a^{'}}{a}|v_k|^2)\bigg]. \quad (6.85)$$

The mass density is therefore given by

$$\rho=\frac{1}{a^2}< \hat{T}_{00} > = \frac{1}{4a^4}\int\frac{d^3k}{(2\pi)^3}\bigg[|v_k^{'}|^2+(k^2+\frac{a^{''}}{a}+a^2m^2)|v_k|^2-\partial_\eta(\frac{a^{'}}{a}|v_k|^2)\bigg]. \quad (6.86)$$

6.3.3 Instantaneous Vacuum

Let us do the calculation in a slightly different way. The comoving scalar field $\chi=a\phi$ satisfies the equation of motion

$$\chi^{''}+m_{\rm eff}^2\chi-\partial_i^2\chi=0\ ,\ m_{\rm eff}^2=a^2m^2-\frac{a^{''}}{a}. \quad (6.87)$$

This can be derived from the action

$$S=\frac{1}{2}\int d\eta d^3x\big[\chi^{'2}-(\partial_i\chi)^2-m_{\rm eff}^2\chi^2\big]. \quad (6.88)$$

We quantize this system now. The conjugate momentum is $\pi=\chi^{'}$. The Hamiltonian is

$$H=\frac{1}{2}\int d^3x\big[\chi^{'2}+(\partial_i\chi)^2+m_{\rm eff}^2\chi^2\big]. \quad (6.89)$$

This is different from the Hamiltonian written down in the previous section. The rest is now the same. For example the field operator can be expanded as (with $[\bar{a}_k,\bar{a}_{k'}^+]=V\delta_{k,k'}$, etc and $v_k^{'}v_k^*-v_kv_k^{*'}=2i$)

$$\hat{\chi}=\frac{1}{V}\sum_k\frac{1}{\sqrt{2}}\bigg(\bar{a}_kv_k^*e^{i\vec{k}\vec{x}}+\bar{a}_k^+v_ke^{-i\vec{k}\vec{x}}\bigg). \quad (6.90)$$

We compute the Hamiltonian operator (assuming isotropic mode functions,viz $v_k=v_{-k}$)

$$\hat{H}=\frac{1}{4V}\sum_k\bigg[F_k^*\bar{a}_k\bar{a}_{-k}+F_k\bar{a}_k^+\bar{a}_{-k}^++E_k(\bar{a}_k\bar{a}_k^++\bar{a}_k^+\bar{a}_k)\bigg]. \quad (6.91)$$

$$F_k=(v_k^{'})^2+\omega_k^2v_k^2\ ,\ E_k=|v_k^{'}|^2+\omega_k^2|v_k|^2. \quad (6.92)$$

Let $|0_v>$ be the vacuum state corresponding to the mode functions v_k. Then

$$< 0_v|\hat{H}|0_v > = \frac{1}{4}\sum_k E_k$$
$$= \frac{V}{4}\int\frac{d^3k}{(2\pi)^3}\bigg[|v_k^{'}|^2+\omega_k^2|v_k|^2\bigg]. \quad (6.93)$$

The vacuum energy density is

$$\rho = \frac{1}{4}\int \frac{d^3k}{(2\pi)^3}\bigg[|v_k^{'}|^2+\omega_k^2|v_k|^2\bigg]. \tag{6.94}$$

This clearly depends on the conformal time η. The instantaneous vacuum at a conformal time $\eta=\eta_0$ is the state $|0_{\eta_0}>$ which is the lowest energy eigenstate of the instantaneous Hamiltonian $H(\eta_0)$. Equivalently the instantaneous vacuum at a conformal time $\eta=\eta_0$ is the state in which the vacuum expectation value $<0_v|\hat{H}(\eta_0)|0_v>$ is minimized with respect to all possible choices of $v_k=v_k(\eta_0)$. The minimization of the energy density ρ corresponds to the minimization of each mode v_k separately. For a given value of $\vec{k}$ we choose $v_k(\eta)$ by imposing at $\eta=\eta_0$ the initial conditions

$$v_k(\eta_0)=q\ ,\ v_k^{'}(\eta_0)=p. \tag{6.95}$$

The normalization condition $v_k^{'}v_k^*-v_kv_k^{*'}=2i$ reads therefore

$$q^*p-p^*q=2i. \tag{6.96}$$

The corresponding energy is $E_k=|p|^2+\omega_k^2(\eta_0)|q|^2$. By using the symmetry $q\longrightarrow e^{i\lambda}q$ and $p\longrightarrow e^{i\lambda}p$ we can choose q real. If we write $p=p_1+ip_2$ then the above condition gives immediately $q=1/p_2$. The energy becomes

$$E_k(\eta_0)=p_1^2+p_2^2+\frac{\omega_k^2(\eta_0)}{p_2^2}. \tag{6.97}$$

The minimum of this energy with respect to p_1 is $p_1=0$ whereas its minimum with respect to p_2 is $p_2=\sqrt{\omega_k(\eta_0)}$. The initial conditions become

$$v_k(\eta_0)=\frac{1}{\sqrt{\omega_k(\eta_0)}}\ ,\ v_k^{'}(\eta_0)=i\omega_k(\eta_0)v_k(\eta_0). \tag{6.98}$$

In Minkowski spacetime we have $a=1$ and thus $\omega_k=\sqrt{k^2+m^2}$. We obtain (with $\eta_0=0$) the usual result $v_k(\eta)=e^{i\omega_k\eta}/\sqrt{\omega_k}$.

The energy in this minimum reads

$$E_k(\eta_0)=2\omega_k(\eta_0). \tag{6.99}$$

The vacuum energy density is therefore

$$\rho = \frac{1}{2}\int \frac{d^3k}{(2\pi)^3}\omega_k(\eta_0). \tag{6.100}$$

This is the usual formula which is clearly divergent so we may proceed in the usual way to perform regularization and renormalization. The problem (which is actually quite severe) is that this energy density is time dependent.

6.3.4 Quantization in de Sitter Spacetime and Bunch-Davies Vacuum

During inflation and also in the limit $a \longrightarrow \infty$ (the future) it is believed that vacuum dominates and thus spacetime is approximately de Sitter spacetime.

An interesting solution of the Friedmann equations (6.36) and (6.37) is precisley the maximally symmetric de Sitter space with positive curvature $\kappa > 0$ and positive cosmological constant $\Lambda > 0$ and no matter content $\rho = P = 0$ given by the scale factor

$$a(t) = \frac{\alpha}{R_0} \cosh \frac{t}{\alpha}. \tag{6.101}$$

$$\alpha = \sqrt{\frac{3}{\Lambda}} \ , \ R_0 = \frac{1}{\sqrt{\kappa}}. \tag{6.102}$$

At large times the Hubble parameter becomes a constant

$$H \simeq \frac{1}{\alpha} = \sqrt{\frac{\Lambda}{3}}. \tag{6.103}$$

The behavior of the scale factor at large times becomes thus

$$a(t) \simeq a_0 e^{Ht} \ \ a_0 = \frac{\alpha}{2R_0}. \tag{6.104}$$

Thus the scale factor on de Sitter space can be given by $a(t) \simeq a_0 \exp(Ht)$. In this case the curvature is computed to be zero and thus the coordinates t, x, y and z are incomplete in the past. The metric is given explicitly by

$$ds^2 = -dt^2 + a_0^2 e^{2Ht} dx^i dx^i. \tag{6.105}$$

In this flat patch (upper half of) de Sitter space is asymptotically static with respect to conformal time η in the past. This can be seen as follows. First we can compute in closed form that $\eta = -e^{-Ht}/(a_0 H)$ and $a(t) = a(\eta) = -1/(H\eta)$ and thus η is in the interval $]-\infty, 0]$ (and hence the coordinates t, x, y and z are incomplete). We then observe that $H_\eta = a^{'}/a = -1/\eta \longrightarrow 0$ when $\eta \longrightarrow -\infty$ which means that de Sitter is asymptotically static.

de Sitter space is characterized by the existence of horizons. As usual null radial geodesics are characterized by $a^2(t)\dot{r}^2 = 1$. The solution is explicitly given by

$$r(t) - r(t_0) = \frac{1}{a_0 H}(e^{-Ht_0} - e^{-Ht}). \tag{6.106}$$

Thus photons emitted at the origin $r(t_0) = 0$ at time t_0 will reach the sphere $r_h = e^{-Ht_0}/(a_0 H)$ at time $t \longrightarrow \infty$ (asymptotically). This sphere is precisely the horizon for the observer at the origin in the sense that signal emitted at the origin can not reach any point beyond the horizon and similarly any signal emitted at time t_0 at a point $r > r_h$ can not reach the observer at the origin.

The horizon scale at time t_0 is defined as the proper distance of the horizon from the observer at the origin, viz $a^2(t_0)r_h = 1/H$. This is clearly the same at all times.

The effective frequencies of oscillation in de Sitter space are

$$\begin{aligned}\omega_k^2(\eta) &= k^2 + m^2a^2 - (1-6\xi)\frac{a^{''}}{a} \\ &= k^2 + [\frac{m^2}{H^2} - 2(1-6\xi)]\frac{1}{\eta^2}.\end{aligned} \tag{6.107}$$

These may become imaginary. For example $\omega_0^2(\eta) < 0$ if $m^2 < 2(1-6\xi)H^2$. We will take $\xi = 0$ and assume that $m << H$.

From the previous section we know that the mode functions must satisfy the differential equations (with $\chi_k = v_k^*/\sqrt{2V}$)

$$v_k^{''} + \left(k^2 + [\frac{m^2}{H^2} - 2]\frac{1}{\eta^2}\right)v_k = 0 \tag{6.108}$$

The solution of this equation is given in terms of Bessel functions J_n and Y_n by [4]

$$v_k = \sqrt{k|\eta|}\left[AJ_n(k|\eta|) + BY_n(k|\eta|)\right] \ , \ n = \sqrt{\frac{9}{4} - \frac{m^2}{H^2}}. \tag{6.109}$$

The normalization condition (6.76) becomes (with $s = k|\eta|$)

$$ks(A^*B - AB^*)(\frac{d}{ds}J_n(s).Y_n(s) - \frac{d}{ds}Y_n(s).J_n(s)) = 2i. \tag{6.110}$$

We use the result [5]

$$\frac{d}{ds}J_n(s).Y_n(s) - \frac{d}{ds}Y_n(s).J_n(s) = -\frac{2}{\pi s}. \tag{6.111}$$

We obtain the constraint

$$AB^* - A^*B = \frac{i\pi}{k}. \tag{6.112}$$

We consider now two limits of interest.

The early time regime $\eta \longrightarrow -\infty$: This corresponds to $\omega_k^2 \longrightarrow k^2$ or equivalently

$$k^2 >> (2 - \frac{m^2}{H^2})\frac{1}{\eta^2}. \tag{6.113}$$

This is a high energy (short distance) limit. The effect of gravity on the modes v_k is therefore negligible and we obtain the Minkowski solutions

$$v_k = \frac{1}{\sqrt{k}}e^{ik\eta} \ , \ k|\eta| >> 1. \tag{6.114}$$

The normalization is chosen in accordance with (6.76).

[4]Exercise: Verify this result. See for example [15].

[5]Exercise: Show this result.

The late time regime $\eta \longrightarrow 0$: In this limit $\omega_k^2 \longrightarrow (m^2/H^2-2)1/\eta^2 < 0$ or equivalently

$$k^2 << (2-\frac{m^2}{H^2})\frac{1}{\eta^2}. \tag{6.115}$$

The differential equation becomes

$$v_k^{''} - (2-\frac{m^2}{H^2})\frac{1}{\eta^2}v_k = 0. \tag{6.116}$$

The solution is immediately given by $v_k = A|\eta|^{n_1} + B|\eta|^{n_2}$ with $n_{1,2} = \pm n + 1/2$. In the limit $\eta \longrightarrow 0$ the dominant solution is obviously associated with the exponent $-n+1/2$. We have then

$$v_k \sim |\eta|^{\frac{1}{2}-n} \ , \ k|\eta| << 1. \tag{6.117}$$

Any mode with momentum k is a wave with a comoving wave length $L \sim 1/k$ and a physical wave length $L_p = a(\eta)L$ and hence

$$k|\eta| = \frac{H^{-1}}{L_p}. \tag{6.118}$$

Thus modes with $k|\eta| >> 1$ corresponds to modes with $L_p << H^{-1}$. These are the sub-horizon modes with physical wave lengths much shorter than the horizon scale and which are unaffected by gravity. Similarly the modes with $k|\eta| << 1$ or equivalently $L_p >> H^{-1}$ are the super-horizon modes with physical wave lengths much larger than the horizon scale. These are the modes which are affected by gravity.

A mode with momentum k which is sub-horizon at early times will become super-horizon at a later time η_k defined by the requirement that $L_p = H^{-1}$ or equivalently $k|\eta_k| = 1$. The time η_k is called the time of horizon crossing of the mode with momentum k.

The behavior $a(\eta) \longrightarrow 0$ when $\eta \longrightarrow -\infty$ allows us to pick a particular vacuum state known as the Bunch-Davies or the Euclidean vacuum. The Bunch-Davies vacuum is a de Sitter invariant state and is the initial state used in cosmology.

In the limit $\eta \longrightarrow -\infty$ the frequency approaches the flat space result, i.e. $\omega_k(\eta) \longrightarrow k$ and hence we can choose the vacuum state to be given by the Minkowski vacuum. More precisely the frequency $\omega_k(\eta)$ is a slowly-varying function for some range of the conformal time η in the limit $\eta \longrightarrow -\infty$. This is called the adiabatic regime of $\omega_k(\eta)$ where it is also assumed that $\omega_k(\eta) > 0$. By applying the Minkowski vacuum prescription in the limit $\eta \longrightarrow -\infty$ we must have

$$v_k = \frac{N}{\sqrt{k}}e^{ik\eta} \ , \ \eta \longrightarrow -\infty. \tag{6.119}$$

From the other hand by using $J_n(s) = \sqrt{2/(\pi s)}\cos\lambda$, $Y_n(s) = \sqrt{2/(\pi s)}\sin\lambda$ with $\lambda = s - n\pi/2 - \pi/4$ we can compute the asymptotic behavior

$$v_k = \sqrt{\frac{2}{\pi}}[A\cos\lambda + B\sin\lambda] \ , \ \eta \longrightarrow -\infty. \tag{6.120}$$

By choosing $B=-iA$ and employing the normalization condition (6.112) we obtain

$$B=-iA\ ,\ A=\sqrt{\frac{\pi}{2k}}. \tag{6.121}$$

Thus we have the solution

$$v_k=\frac{1}{\sqrt{k}}e^{i(k\eta+\frac{n\pi}{2}+\frac{\pi}{4})}\ ,\ \eta\longrightarrow-\infty. \tag{6.122}$$

The Bunch-Davies vacuum corresponds to the choice $N=\exp(i\frac{n\pi}{2}+i\frac{\pi}{4})$. The full solution using this choice becomes

$$v_k=\sqrt{\frac{\pi|\eta|}{2}}\bigg[J_n(k|\eta|)-iY_n(k|\eta|)\bigg]\ ,\ n=\sqrt{\frac{9}{4}-\frac{m^2}{H^2}}. \tag{6.123}$$

The mass density in FLRW spacetime was already computed in equation (6.86). We have

$$\rho = \frac{1}{4a^4}\int\frac{d^3k}{(2\pi)^3}\bigg[|v_k^{'}|^2+(k^2+\frac{a^{''}}{a}+a^2m^2)|v_k|^2-\partial_\eta(\frac{a^{'}}{a}|v_k|^2)\bigg]. \tag{6.124}$$

For de Sitter space we have $a=-1/(\eta H)$ and thus

$$\rho = \frac{\eta^4H^4}{4}\int\frac{d^3k}{(2\pi)^3}\bigg[|v_k^{'}|^2+(k^2+\frac{2}{\eta^2}+\frac{m^2}{H^2\eta^2})|v_k|^2+\partial_\eta(\frac{1}{\eta}|v_k|^2)\bigg]. \tag{6.125}$$

For $m=0$ we have the solutions

$$v_k=\sqrt{\frac{\pi|\eta|}{2}}\bigg[J_{\frac{3}{2}}(k|\eta|)-iY_{\frac{3}{2}}(k|\eta|)\bigg]. \tag{6.126}$$

We use the results ($x=k|\eta|$)

$$J_{3/2}(x)=\sqrt{\frac{2}{\pi x}}\bigg(\frac{\sin x}{x}-\cos x\bigg)\ ,\ Y_{3/2}(x)=\sqrt{\frac{2}{\pi x}}\bigg(-\frac{\cos x}{x}-\sin x\bigg). \tag{6.127}$$

We obtain then

$$v_k=-\frac{i}{k^{\frac{3}{2}}}\frac{e^{ik\eta}}{\eta}-\frac{1}{k^{\frac{1}{2}}}e^{ik\eta}. \tag{6.128}$$

In other words

$$|v_k|^2=\frac{1}{k^3}\frac{1}{\eta^2}+\frac{1}{k}\ ,\ |v_k^{'}|^2=-\frac{1}{k}\frac{1}{\eta^2}+\frac{1}{k^3}\frac{1}{\eta^4}+k. \tag{6.129}$$

We obtain then (using also a hard cutoff Λ)

$$\begin{aligned}\rho &= \frac{\eta^4H^4}{4}\int\frac{d^3k}{(2\pi)^3}[2k+\frac{1}{k\eta^2}]\\ &= \frac{\eta^4H^4}{16\pi^2}(\Lambda^4+\frac{\Lambda^2}{\eta^2}).\end{aligned} \tag{6.130}$$

This goes to zero in the limit $\eta \longrightarrow 0$. However if we take $\Lambda = \Lambda_0 a$ where Λ_0 is a proper momentum cutoff then the energy density becomes independent of time and we are back to the same problem. We get

$$\rho = \frac{1}{16\pi^2}(\Lambda_0^4 + H^2\Lambda_0^2). \tag{6.131}$$

We observe that

$$\begin{aligned}\rho_{\rm deSitter} - \rho_{\rm Minkowski} &= \frac{H^2}{\Lambda_0^2}\frac{\Lambda_0^4}{16\pi^2}\\ &= \frac{H^2}{\Lambda_0^2}\rho_{\rm Minkowski}.\end{aligned} \tag{6.132}$$

We take the value of the Hubble parameter at the current epoch as the value of the Hubble parameter of de Sitter space, viz

$$H = H_0 = \frac{7 \times 6.58}{3.09} 10^{-43} GeV. \tag{6.133}$$

We get then

$$\begin{aligned}\rho_{\rm deSitter} - \rho_{\rm Minkowski} &= 0.38(10^{-30})^4.0.22(10^{18}GeV)^4\\ &= 0.084(10^{-12}GeV)^4.\end{aligned} \tag{6.134}$$

6.3.5 QFT on Curved Background with a Cutoff

In [30] a proposal for quantum field theories on curved backgrounds with a plausible cutoff is put forward.

6.3.6 The Conformal Limit $\xi \longrightarrow 1/6$

The mode functions χ_k satisfy

$$\chi_k^{''} + \omega_k^2(\eta)\chi_k = 0\ ,\ \omega_k^2 = k^2 + m^2a^2 - (1-6\xi)\frac{a^{''}}{a}. \tag{6.135}$$

$$V(\chi_k\chi_k^{*'} - \chi_k^*\chi_k^{'}) = i. \tag{6.136}$$

We will consider in this section $m^2 = 0$. We assume now that the universe is Minkowski in the past $\eta \longrightarrow -\infty$. In other words in the limit $\eta \longrightarrow -\infty$ the frequency ω_k tends to $\bar{\omega}_k = \sqrt{k^2}$. The corresponding mode function is therefore

$$\chi_k = \chi_k^{(in)} = \frac{1}{\sqrt{2V\bar{\omega}_k}} e^{-i\bar{\omega}_k\eta}. \tag{6.137}$$

We will also assume that the universe is Minkowski in the future $\eta \longrightarrow +\infty$. The frequency in the limit $\eta \longrightarrow +\infty$ is again given by $\bar{\omega}_k = \sqrt{k^2}$. The corresponding mode function is therefore

$$\chi_k = \chi_k^{(out)} = \frac{\alpha_k}{\sqrt{2V\bar{\omega}_k}} e^{-i\bar{\omega}_k\eta} + \frac{\beta_k}{\sqrt{2V\bar{\omega}_k}} e^{i\bar{\omega}_k\eta}. \tag{6.138}$$

We determine α_k and β_k from solving the equation of motion (6.135) with the initial condition (6.137). We remark that

$$\chi_k^{(out)} = \alpha_k \chi_k^{(in)} + \beta_k \chi_k^{(in)*}. \tag{6.139}$$

We imagine that the out state is the limit $\eta \longrightarrow +\infty$ of some v mode function while the in state is the limit $\eta \longrightarrow -\infty$ of some u mode function. More precisely we are assuming that

$$\begin{aligned} u_i &\longrightarrow \chi_i^{(in)} \ , \ \eta \longrightarrow -\infty \\ v_i &\longrightarrow \chi_i^{(out)} \ , \ \eta \longrightarrow +\infty. \end{aligned} \tag{6.140}$$

The relation between the u and the v mode functions is given in terms of Bogolubov coefficients by equation (6.55). By comparing with the above relation (6.139) we deduce that

$$\alpha_{ij} = \alpha_i \delta_{ij} \ , \ \beta_{ij} = \beta_i \delta_{ij}. \tag{6.141}$$

Let $N_u = \sum_k \hat{a}_k^+ \hat{a}_k$ be the number operator corresponding to the u modes. If $|0_u>$ is the vacuum state corresponding to the u modes then $<0_u|N_u|0_u>= 0$. The number of particles created by the gravitational field in the limit $\eta \longrightarrow +\infty$ is precisely $<0_v|N_u|0_v>$ where $|0_v>$ is the vacuum state corresponding to the v modes. The number density of created particles is then given by

$$\mathcal{N} = \frac{<0_v|N_u|0_v>}{V} = \int \frac{d^3k}{(2\pi)^3} |\beta_k|^2. \tag{6.142}$$

The corresponding energy density is

$$\rho = \int \frac{d^3k}{(2\pi)^3} \bar{\omega}_k |\beta_k|^2. \tag{6.143}$$

The initial differential equation (6.135) can be rewritten as

$$\chi_k^{''} + \bar{\omega}_k^2 \chi_k = j_k(\eta) \ , \ j_k(\eta) = (1-6\xi)\frac{a^{''}}{a}\chi_k. \tag{6.144}$$

We can write down immediately the solution as

$$\begin{aligned} \chi_k &= \chi_k^{(in)} + \frac{1}{\bar{\omega}_k}\int_{-\infty}^{\eta} d\eta^{'} \sin\bar{\omega}_k(\eta-\eta^{'}) j_k(\eta^{'}) \\ &= \chi_k^{(in)} + \frac{1-6\xi}{\bar{\omega}_k}\int_{-\infty}^{\eta} d\eta^{'} \frac{a^{''}(\eta^{'})}{a(\eta^{'})} \sin\bar{\omega}_k(\eta-\eta^{'}) \chi_k(\eta^{'}). \end{aligned} \tag{6.145}$$

To lowest order in $1-6\xi$ this solution becomes

$$\chi_k = \chi_k^{(in)}+\frac{1-6\xi}{\bar{\omega}_k}\int_{-\infty}^{\eta}d\eta^{'}\frac{a^{''}(\eta^{'})}{a(\eta^{'})}\sin\bar{\omega}_k(\eta-\eta^{'})\chi_k^{(in)}(\eta^{'}). \tag{6.146}$$

From this formula we obtain immediately

$$\chi_k^{(out)} = \chi_k^{(in)}+\frac{1-6\xi}{\bar{\omega}_k}\int_{-\infty}^{+\infty}d\eta^{'}\frac{a^{''}(\eta^{'})}{a(\eta^{'})}\sin\bar{\omega}_k(\eta-\eta^{'})\chi_k^{(in)}(\eta^{'}). \tag{6.147}$$

By comparing with (6.139) and using (6.137) we get after few more lines (with $a^2R=6a^{''}/a$)

$$\alpha_k=1+\frac{i}{2\bar{\omega}_k}(\frac{1}{6}-\xi)\int_{-\infty}^{+\infty}d\eta^{'}a^2(\eta^{'})R(\eta^{'})\ ,\ \beta_k=-\frac{i}{2\bar{\omega}_k}(\frac{1}{6}-\xi)\int_{-\infty}^{+\infty}d\eta^{'}a^2(\eta^{'})R(\eta^{'})e^{-2i\bar{\omega}_k\eta^{'}}. \tag{6.148}$$

The number density is given by

$$\begin{aligned}\mathcal{N} &= \frac{1}{4}(\frac{1}{6}-\xi)^2\int_{-\infty}^{+\infty}d\eta_1\int_{-\infty}^{+\infty}d\eta_2a^2(\eta_1)R(\eta_1)a^2(\eta_2)R(\eta_2)\int\frac{d^3k}{(2\pi)^3}\frac{1}{\bar{\omega}_k^2}e^{-2i\bar{\omega}_k(\eta_1-\eta_2)}\\ &= \frac{1}{4}(\frac{1}{6}-\xi)^2\int_{-\infty}^{+\infty}d\eta_1\int_{-\infty}^{+\infty}d\eta_2a^2(\eta_1)R(\eta_1)a^2(\eta_2)R(\eta_2)\frac{1}{2\pi}\int_0^{\infty}\frac{dk}{2\pi}e^{-ik(\eta_1-\eta_2)}\\ &= \frac{1}{4}(\frac{1}{6}-\xi)^2\int_{-\infty}^{+\infty}d\eta_1\int_{-\infty}^{+\infty}d\eta_2a^2(\eta_1)R(\eta_1)a^2(\eta_2)R(\eta_2)\frac{1}{4\pi}\int_0^{\infty}\frac{dk}{2\pi}e^{-ik(\eta_1-\eta_2)}\\ &= \frac{1}{16\pi}(\frac{1}{6}-\xi)^2\int_{-\infty}^{+\infty}d\eta a^4(\eta)R^2(\eta).\end{aligned} \tag{6.149}$$

The energy density is given by (with the assumption that $a^2(\eta)R(\eta)\longrightarrow 0$ when $\eta\longrightarrow\pm\infty$)

$$\begin{aligned}\rho &= \frac{1}{4}(\frac{1}{6}-\xi)^2\int_{-\infty}^{+\infty}d\eta_1\int_{-\infty}^{+\infty}d\eta_2a^2(\eta_1)R(\eta_1)a^2(\eta_2)R(\eta_2)\int\frac{d^3k}{(2\pi)^3}\frac{1}{\bar{\omega}_k}e^{-2i\bar{\omega}_k(\eta_1-\eta_2)}\\ &= \frac{1}{4}(\frac{1}{6}-\xi)^2\int_{-\infty}^{+\infty}d\eta_1\int_{-\infty}^{+\infty}d\eta_2a^2(\eta_1)R(\eta_1)a^2(\eta_2)R(\eta_2)\frac{1}{8\pi^2}\int_0^{\infty}kdke^{-ik(\eta_1-\eta_2)}\\ &= \frac{1}{4}(\frac{1}{6}-\xi)^2\int_{-\infty}^{+\infty}d\eta_1\int_{-\infty}^{+\infty}d\eta_2a^2(\eta_1)R(\eta_1)a^2(\eta_2)R(\eta_2)\frac{1}{8\pi^2}\frac{d^2}{d\eta_1d\eta_2}\int_0^{\infty}\frac{dk}{k}e^{-ik(\eta_1-\eta_2)}\\ &= \frac{1}{4}(\frac{1}{6}-\xi)^2\int_{-\infty}^{+\infty}d\eta_1\int_{-\infty}^{+\infty}d\eta_2\frac{d}{d\eta_1}(a^2(\eta_1)R(\eta_1))\frac{d}{d\eta_2}(a^2(\eta_2)R(\eta_2))\frac{1}{2\pi}\int_0^{\infty}\frac{dk}{2\pi}\frac{1}{2k}e^{-ik(\eta_1-\eta_2)}.\end{aligned} \tag{6.150}$$

The last factor is precisley one half the Feynamn propagator in $1+1$ dimension for $r=0$ (see equation (4) of [24]). We have then

$$\begin{aligned}\rho &= \frac{1}{4}(\frac{1}{6}-\xi)^2\int_{-\infty}^{+\infty}d\eta_1\int_{-\infty}^{+\infty}d\eta_2\frac{d}{d\eta_1}(a^2(\eta_1)R(\eta_1))\frac{d}{d\eta_2}(a^2(\eta_2)R(\eta_2))\frac{1}{2\pi}\frac{-1}{4\pi}\ln|\eta_1-\eta_2|\\ &= -\frac{1}{32\pi^2}(\frac{1}{6}-\xi)^2\int_{-\infty}^{+\infty}d\eta_1\int_{-\infty}^{+\infty}d\eta_2\frac{d}{d\eta_1}(a^2(\eta_1)R(\eta_1))\frac{d}{d\eta_2}(a^2(\eta_2)R(\eta_2))\ln|\eta_1-\eta_2|.\end{aligned} \tag{6.151}$$

At the end of inflation the universe transits from a de Sitter spacetime (which is asymptotically static in the infinite past) to a radiation dominated Robertson-Walker universe (which is asymptotically flat in the infinite future) in a very short time interval. Let us assume that the transition occurs abruptly at a time $\eta_0 < 0$. In de Sitter space ($\eta < \eta_0$) we have $a = -1/(\eta H)$ and $R = 12H^2$. In the radiation dominated phase ($\eta > \eta_0$) we may assume that $R = 0$. We get immediately

$$\begin{aligned}\mathcal{N} &= \frac{1}{16\pi}(\frac{1}{6}-\xi)^2\int_{-\infty}^{\eta_0} d\eta a^4(\eta)R^2(\eta)\\ &= \frac{H^3}{12\pi}(1-6\xi)^2 a^3(\eta_0).\end{aligned} \quad (6.152)$$

This is the number density of created particles (via gravitational interaction) just after the transition, i.e. during reheating.

To compute the energy density we will assume that the transition from de sitter spacetime to radiation dominated spacetime is smother given by the scale factor

$$a^2(\eta) = f(\eta H). \quad (6.153)$$

$$\begin{aligned}f &= \frac{1}{\eta^2 H^2}\ ,\ \eta < -H^{-1}\\ &= a_0 + a_1 H\eta + a_2 H^2\eta^2 + a_3 H^3\eta^3\ ,\ -H^{-1} < \eta < (x_0-1)H^{-1}\\ &= b_0(H\eta + b_1)^2\ ,\ \eta > (x_0-1)H^{-1}.\end{aligned} \quad (6.154)$$

In this model the time $\eta = -H^{-1}$ corresponding to $t = 0$ marks the end of the inflationary (de Sitter) phase and the transition to radiation dominated phase occurs on a time scale given by $\Delta\eta = H^{-1}x_0$. By requiring that f, $f^{'}$ and $f^{''}$ are continuous at $\eta = -H^{-1}$ and $\eta = (x_0-1)H^{-1}$ we can determine the coefficients a_i and b_i uniquely. We compute immediately

$$a^2 R = 3H^2 V\ ,\ V = f^{-2}\bigg[f^{''}f - \frac{1}{2}(f^{'})^2\bigg]. \quad (6.155)$$

We can then compute in a straightforward manner [6]

$$\begin{aligned}V &= \frac{4}{x^2}\ ,\ x < -1\\ &\simeq -\frac{4}{x_0}\ ,\ -1 < x < x_0 - 1\ ,\ x_0 << 1\\ &= 0\ ,\ x > x_0 - 1.\end{aligned} \quad (6.156)$$

[6] Exercise: Show this result explicitly.

The energy density is then given by [7]

$$\begin{aligned}
\rho &= -\frac{H^4}{128\pi^2}(1-6\xi)^2\int_{-\infty}^{x_0-1}dx_1\int_{-\infty}^{x_0-1}dx_2V^{'}(x_1)V^{'}(x_2)\ln\frac{|x_1-x_2|}{H}\\
&= -\frac{H^4}{128\pi^2}(1-6\xi)^2.16\ln x_0\\
&= -\frac{H^4}{8\pi^2}(1-6\xi)^2\ln x_0. \qquad (6.157)
\end{aligned}$$

In the above model we have chosen the transition time to be $\eta=-H^{-1}$ and thus $a=-1/(\eta H)=+1$ and as a consequence $\Delta\eta=-H\eta\Delta t=\Delta t$. From the other hand the transition from de Sitter spacetime to radiation dominated phase occurs on a time scale given by $\Delta\eta=H^{-1}x_0$. From these two facts we obtain $x_0=H\Delta t$ and hence the energy density becomes

$$\rho = -\frac{H^4}{8\pi^2}(1-6\xi)^2\ln H\Delta t. \qquad (6.158)$$

This is the energy density of the created particles after the end of inflation. The factor $1-6\xi$ is small whereas the factor $\ln H\Delta t$ is large and it is not obvious how they should balance without an extra input.

6.4 Is Vacuum Energy Real?

6.4.1 The Casimir Force

We consider two large and perfectly conducting plates of surface area A at a distance L apart with $\sqrt{A}>>L$ so that we can ignore edge contributions. The plates are in the xy plane at $x=0$ and $x=L$. In the volume AL the electromagnetic standing waves take the form

$$\psi_n(t,x,y,z)=e^{-i\omega_nt}e^{ik_xx+ik_yy}\sin k_nz. \qquad (6.159)$$

They satisfy the Dirichlet boundary conditions

$$\psi_n|_{z=0}=\psi_n|_{z=L}=0. \qquad (6.160)$$

Thus we must have

$$k_n=\frac{n\pi}{L}\ ,\ n=1,2,.... \qquad (6.161)$$

$$\omega_n=\sqrt{k_x^2+k_y^2+\frac{n^2\pi^2}{L^2}}. \qquad (6.162)$$

[7]Exercise: Derive the second line.

These modes are transverse and thus each value of n is associated with two degrees of freedom. There is also the possibility of

$$k_n = 0. \tag{6.163}$$

In this case there is a corresponding single degree of freedom.

The zero point energy of the electromagnetic field between the plates is

$$\begin{aligned} E &= \frac{1}{2}\sum_n \omega_n \\ &= \frac{1}{2}A\int \frac{d^2k}{(2\pi)^2}\bigg[k + 2\sum_{n=1}^{\infty}(k^2+\frac{n^2\pi^2}{L^2})^{1/2}\bigg]. \end{aligned} \tag{6.164}$$

The zero point energy of the electromagnetic field in the same volume in the absence of the plates is

$$\begin{aligned} E_0 &= \frac{1}{2}\sum_n \omega_n \\ &= \frac{1}{2}A\int \frac{d^2k}{(2\pi)^2}\bigg[2L\int\frac{dk_n}{2\pi}(k^2+k_n^2)^{1/2}\bigg]. \end{aligned} \tag{6.165}$$

After the change of variable $k = n\pi/L$ we obtain

$$E_0 = \frac{1}{2}A\int \frac{d^2k}{(2\pi)^2}\bigg[2\int_0^{\infty}dn(k^2+\frac{n^2\pi^2}{L^2})^{1/2}\bigg]. \tag{6.166}$$

We have then

$$\mathcal{E} = \frac{E-E_0}{A} = \int \frac{d^2k}{(2\pi)^2}\bigg[\frac{1}{2}k + \sum_{n=1}^{\infty}(k^2+\frac{n^2\pi^2}{L^2})^{1/2} - \int_0^{\infty}dn(k^2+\frac{n^2\pi^2}{L^2})^{1/2}\bigg]. \tag{6.167}$$

This is obvioulsy a UV divergent quantity. We regularize this energy density by introducing a cutoff function $f_\Lambda(k)$ which is equal to 1 for $k << \Lambda$ and 0 for $k >> \Lambda$. We have then (with the change of variables $k = \pi x/L$ and $x^2 = t$)

$$\begin{aligned} \mathcal{E}_\Lambda &= \int \frac{d^2k}{(2\pi)^2}\bigg[\frac{1}{2}f_\Lambda(k)k + \sum_{n=1}^{\infty}f_\Lambda(\sqrt{k^2+\frac{n^2\pi^2}{L^2}})(k^2+\frac{n^2\pi^2}{L^2})^{1/2} - \int_0^{\infty}dnf_\Lambda(\sqrt{k^2+\frac{n^2\pi^2}{L^2}})(k^2+\frac{n^2\pi^2}{L^2})^{1/2}\bigg] \\ &= \frac{\pi^2}{4L^3}\int dt\bigg[\frac{1}{2}f_\Lambda(\frac{\pi}{L}\sqrt{t})t^{1/2} + \sum_{n=1}^{\infty}f_\Lambda(\frac{\pi}{L}\sqrt{t+n^2})(t+n^2)^{1/2} - \int_0^{\infty}dnf_\Lambda(\frac{\pi}{L}\sqrt{t+n^2})(t+n^2)^{1/2}\bigg] \end{aligned} \tag{6.168}$$

This is an absolutely convergent quantity and thus we can exchange the sums and the integrals. We obtain

$$\mathcal{E}_\Lambda = \frac{\pi^2}{4L^3}\bigg[\frac{1}{2}F(0) + F(1) + F(2).... - \int_0^{\infty}dnF(n)\bigg]. \tag{6.169}$$

The function $F(n)$ is defined by

$$F(n)=\int_0^{\infty}dtf_{\Lambda}(\frac{\pi}{L}\sqrt{t+n^2})(t+n^2)^{1/2}. \tag{6.170}$$

Since $f(k)\longrightarrow 0$ when $k\longrightarrow\infty$ we have $F(n)\longrightarrow 0$ when $n\longrightarrow\infty$.

We use the Euler-MacLaurin formula

$$\frac{1}{2}F(0)+F(1)+F(2)....-\int_0^{\infty}dnF(n)=-\frac{1}{2!}B_2F^{'}(0)-\frac{1}{4!}B_4F^{'''}(0)+.... \tag{6.171}$$

The Bernoulli numbers B_i are defined by

$$\frac{y}{e^y-1}=\sum_{i=0}^{\infty}B_i\frac{y^i}{i!}. \tag{6.172}$$

For example

$$B_0=\frac{1}{6}\ ,\ B_4=-\frac{1}{30}\ ,\ \text{etc.} \tag{6.173}$$

Thus

$$\mathcal{E}_{\Lambda}\ =\ \frac{\pi^2}{4L^3}\bigg[-\frac{1}{12}F^{'}(0)+\frac{1}{720}F^{'''}(0)+....\bigg]. \tag{6.174}$$

We can write

$$F(n)=\int_{n^2}^{\infty}dtf_{\Lambda}(\frac{\pi}{L}\sqrt{t})(t)^{1/2}. \tag{6.175}$$

We assume that $f(0)=1$ while all its derivatives are zero at $n=0$. Thus

$$F^{'}(n)=-\int_{n^2}^{n^2+2n\delta n}dtf_{\Lambda}(\frac{\pi}{L}\sqrt{t})(t)^{1/2}=-2n^2f_{\Lambda}(\frac{\pi}{L}n)\Rightarrow F^{'}(0)=0. \tag{6.176}$$

$$F^{''}(n)=-4nf_{\Lambda}(\frac{\pi}{L}n)-\frac{2\pi}{L}n^2f_{\Lambda}^{'}(\frac{\pi}{L}n)\Rightarrow F^{''}(0)=0. \tag{6.177}$$

$$F^{'''}(n)=-4f_{\Lambda}(\frac{\pi}{L}n)-\frac{8\pi}{L}nf_{\Lambda}^{'}(\frac{\pi}{L}n)-\frac{2\pi^2}{L^2}n^2f_{\Lambda}^{''}(\frac{\pi}{L}n)\Rightarrow F^{'''}(0)=-4. \tag{6.178}$$

We can check that all higher derivatives of F are actually 0[8]. Hence

$$\mathcal{E}_{\Lambda}\ =\ \frac{\pi^2}{4L^3}\bigg[-\frac{4}{720}\bigg]=-\frac{\pi^2}{720L^3}. \tag{6.179}$$

This is the Casimir energy. It corresponds to an attractive force which is the famous Casimir force.

[8]Exercise: Convince yourself of this fact.

6.4.2 The Dirichlet Propagator

We define the propagator by

$$D_F(x,x^{'})=<0|T\hat{\phi}(x)\hat{\phi}(x^{'})|0>. \tag{6.180}$$

It satisfies the inhomogeneous Klein-Gordon equation

$$(\partial_t^2-\partial_i^2)D_F(x,x^{'})=i\delta^4(x-x^{'}). \tag{6.181}$$

We introduce Fourier transform in the time direction by

$$D_F(\omega,\vec{x},\vec{x}^{'})=\int dt e^{-i\omega(t-t^{'})}D_F(x,x^{'})\ ,\ D_F(x,x^{'})=\int\frac{d\omega}{2\pi}e^{i\omega(t-t^{'})}D_F(\omega,\vec{x},\vec{x}^{'}). \tag{6.182}$$

We have

$$(\partial_i^2+\omega^2)D_F(\omega,\vec{x},\vec{x}^{'})=-i\delta^3(\vec{x}-\vec{x}^{'}). \tag{6.183}$$

We expand the reduced Green's function $D_F(\omega,\vec{x},\vec{x}^{'})$ as

$$D_F(\omega,\vec{x},\vec{x}^{'})=-i\sum_n\frac{\phi_n(\vec{x})\phi_n^*(\vec{x}^{'})}{\omega^2-k_n^2}. \tag{6.184}$$

The eigenfunctions $\phi_n(\vec{x})$ satisfy

$$\begin{aligned}\partial_i^2\phi_n(\vec{x})&=-k_n^2\phi_n(\vec{x})\\ \delta^3(\vec{x}-\vec{x}^{'})&=\sum_n\phi_n(\vec{x})\phi_n^*(\vec{x}^{'}).\end{aligned} \tag{6.185}$$

In infinite space we have

$$\phi_i(\vec{x})\longrightarrow\phi_{\vec{k}}(\vec{x})=e^{-i\vec{k}\vec{x}}\ ,\ \sum_i\longrightarrow\int\frac{d^3k}{(2\pi)^3}. \tag{6.186}$$

Thus

$$D_F(\omega,\vec{x},\vec{x}^{'})=i\int\frac{d^3k}{(2\pi)^3}\frac{e^{-i\vec{k}(\vec{x}-\vec{x}^{'})}}{\vec{k}^2-\omega^2}. \tag{6.187}$$

We can compute the closed form [9]

$$D_F(\omega,\vec{x},\vec{x}^{'})=\frac{i}{4\pi}\frac{e^{i\omega|\vec{x}-\vec{x}^{'}|}}{|\vec{x}-\vec{x}^{'}|}. \tag{6.188}$$

[9] Exercise: derive this result.

Equivalently we have

$$D_F(x,x^{'}) = i\int \frac{d^4k}{(2\pi)^4}\frac{e^{-ik(x-x^{'})}}{k^2}. \tag{6.189}$$

Let us remind ourselves with few more results. We have (with $\omega_k=|\vec{k}|$)

$$D_F(x,x^{'}) = \int \frac{d^3k}{(2\pi)^3}\frac{1}{2\omega_k}e^{-ik(x-x^{'})}. \tag{6.190}$$

Recall that $k(x-x^{'})=-k^0(x^0-x^{0'})+\vec{k}(\vec{x}-\vec{x}^{'})$. After Wick rotation in which $x^0\longrightarrow -ix_4$ and $k^0\longrightarrow -ik_4$ we obtain $k(x-x^{'})=k_4(x_4-x_4^{'})+\vec{k}(\vec{x}-\vec{x}^{'})$. The above integral becomes then [10]

$$\begin{aligned} D_F(x,x^{'}) &= \int \frac{d^3k}{(2\pi)^3}\frac{1}{2\omega_k}e^{-i\left(k_4(x_4-x_4^{'})-\vec{k}(\vec{x}-\vec{x}^{'})\right))} \\ &= \frac{1}{4\pi^2}\frac{1}{(x-x^{'})^2}. \end{aligned} \tag{6.191}$$

We consider now the case of parallel plates separated by a distance L. The plates are in the xy plane. We impose now different boundary conditions on the field by assuming that $\hat{\phi}$ is confined in the z direction between the two plates at $z=0$ and $z=L$. Thus the field must vanishes at these two plates, viz

$$\hat{\phi}|_{z=0}=\hat{\phi}|_{z=L}=0. \tag{6.192}$$

As a consequence the plane wave e^{ik_3z} will be replaced with the standing wave $\sin k_3z$ where the momentum in the z direction is quantized as

$$k_3=\frac{n\pi}{L}\ ,\ n\in Z^+. \tag{6.193}$$

Thus the frequency ω_k becomes

$$\omega_n=\sqrt{k_1^2+k_2^2+(\frac{n\pi}{L})^2}. \tag{6.194}$$

We will think of the propagator (6.191) as the electrostatic potential (in 4 dimensions) generated at point y from a unit charge at point x, viz

$$V\equiv D_F(x,x^{'}) = \frac{1}{4\pi^2}\frac{1}{(x-x^{'})^2}. \tag{6.195}$$

We will find the propagator between parallel plates starting from this potential using the method of images. It is obvious that this propagator must satisfy

$$D_F(x,x^{'})=0\ ,\ z=0,L\text{ and }z^{'}=0,L. \tag{6.196}$$

[10]Exercise: derive the second line of this equation.

Instead of the two plates at $x=0$ and $x=L$ we consider image charges (always with respect to the two plates) placed such that the two plates remain grounded. First we place an image charge -1 at $(x,y,-z)$ which makes the potential at the plate $z=0$ zero. The image of the charge at $(x,y,-z)$ with respect to the plane at $z=L$ is a charge $+1$ at $(x,y,z+2L)$. This last charge has an image with respect to $z=0$ equal -1 at $(x,y,-z-2L)$ which in turn has an image with respect to $z=L$ equal $+1$ at $(x,y,z+4L)$. This process is to be continued indefinitely. We have then added the following image charges

$$q=+1\ ,\ (x,y,z+2nL)\ ,\ n=0,1,2,... \tag{6.197}$$

$$q=-1\ ,\ (x,y,-z-2nL)\ ,\ n=0,1,2,... \tag{6.198}$$

The way we did this we are guaranteed that the total potential at $z=0$ is 0. The contribution of the added image charges to the plate $z=L$ is also zero but this plate is still not balanced properly precisely because of the original charge at (x,y,z).

The image charge of the original charge with respect to the plate at $z=L$ is a charge -1 at $(x,y,2L-z)$ which has an image with respect to $z=0$ equal $+1$ at $(x,y,-2L+z)$. This last image has an image with respect to $z=L$ equal -1 at $(x,y,4L-z)$. This process is to be continued indefinitely with added charges given by

$$q=+1\ ,\ (x,y,z+2nL)\ ,\ n=-1,-2,... \tag{6.199}$$

$$q=-1\ ,\ (x,y,-z-2nL)\ ,\ n=-1,-2,... \tag{6.200}$$

By the superposition principle the total potential is the sum of the individual potentials. We get immediately

$$V\equiv D_F(x,x^{'}) = \frac{1}{4\pi^2}\sum_{n=-\infty}^{+\infty}\bigg[\frac{1}{(x-x'-2nLe_3)^2}-\frac{1}{(x-x'-2(nL+z)e_3)^2}\bigg]. \tag{6.201}$$

This satisfies the boundary conditions (6.196). By the uniqueness theorem this solution must therefore be the desired propagator. At this point we can undo the Wick rotation and return to Minkowski spacetime.

6.4.3 Another Derivation Using The Energy-Momentum Tensor

The stress-energy-momentum tensor in flat space with minimal coupling $\xi=0$ and $m=0$ is given by

$$T_{\mu\nu}=\partial_\mu\phi\partial_\nu\phi-\frac{1}{2}\eta_{\mu\nu}\partial_\alpha\phi\partial^\alpha\phi. \tag{6.202}$$

The stress-energy-momentum tensor in flat space with conformal coupling $\xi = 1/6$ and $m = 0$ is given by [11]

$$T_{\mu\nu} = \frac{2}{3}\partial_\mu\phi\partial_\nu\phi + \frac{1}{6}\eta_{\mu\nu}\partial_\alpha\phi\partial^\alpha\phi + \frac{1}{3}\phi\partial_\mu\partial_\nu\phi. \tag{6.203}$$

This tensor is traceless, i.e $T_\mu{}^\mu = 0$ which reflects the fact that the theory is conformal. This tensor is known as the new improved stress-energy-momentum tensor.

In the quantum theory $T_{\mu\nu}$ becomes an operator $\hat{T}_{\mu\nu}$ and we are interested in the expectation value of $\hat{T}_{\mu\nu}$ in the vacuum state $<0|\hat{T}_{\mu\nu}|0>$. We are of course interested in the energy density which is equal to $<0|\hat{T}_{00}|0>$ in flat spacetime. We compute (using the Klein-Gordon equation $\partial_\mu\partial^\mu\hat{\phi} = 0$)

$$\begin{aligned}
<0|\hat{T}_{00}|0> &= \frac{2}{3}<0|\partial_0\hat{\phi}\partial_0\hat{\phi}|0> - \frac{1}{6}<0|\partial_\alpha\hat{\phi}\partial^\alpha\hat{\phi}|0> + \frac{1}{3}<0|\hat{\phi}\partial_\mu\partial_\nu\hat{\phi}|0> \\
&= \frac{5}{6}<0|\partial_0\hat{\phi}\partial_0\hat{\phi}|0> - \frac{1}{6}<0|\partial_i\hat{\phi}\partial_i\hat{\phi}|0> + \frac{1}{3}<0|\hat{\phi}\partial_0^2\hat{\phi}|0> \\
&= \frac{5}{6}<0|\partial_0\hat{\phi}\partial_0\hat{\phi}|0> - \frac{1}{6}<0|\partial_i\hat{\phi}\partial_i\hat{\phi}|0> + \frac{1}{3}<0|\hat{\phi}\partial_i^2\hat{\phi}|0> \\
&= \frac{5}{6}<0|\partial_0\hat{\phi}\partial_0\hat{\phi}|0> + \frac{1}{6}<0|\partial_i\hat{\phi}\partial_i\hat{\phi}|0>.
\end{aligned} \tag{6.204}$$

We regularize this object by putting the two fields at different points x and y as follows

$$\begin{aligned}
<0|\hat{T}_{00}|0> &= \frac{5}{6}<0|\partial_0\hat{\phi}(x)\partial_0\hat{\phi}(y)|0> + \frac{1}{6}<0|\partial_i\hat{\phi}(x)\partial_i\hat{\phi}(y)|0> \\
&= \left[\frac{5}{6}\partial_0^x\partial_0^y + \frac{1}{6}\partial_i^x\partial_i^y\right]<0|\hat{\phi}(x)\hat{\phi}(y)|0>.
\end{aligned} \tag{6.205}$$

Similarly we obtain with minimal coupling the result

$$<0|\hat{T}_{00}|0> = \left[\frac{1}{2}\partial_0^x\partial_0^y + \frac{1}{2}\partial_i^x\partial_i^y\right]<0|\hat{\phi}(x)\hat{\phi}(y)|0>. \tag{6.206}$$

In infinite space the scalar field operator has the expansion (with $w_k = |k|$, $[\bar{a}_k, \bar{a}_{k'}^+] = V\delta_{k,k'}$, etc)

$$\hat{\phi} = \int \frac{d^3k}{(2\pi)^3}\frac{1}{\sqrt{2\omega_k}}\bigg(\bar{a}_k e^{-i\omega_k t + i\vec{k}\vec{x}} + \bar{a}_k^+ e^{i\omega_k t - i\vec{k}\vec{x}}\bigg). \tag{6.207}$$

In the space between parallel plates the field can then be expanded as

$$\hat{\phi} = \sqrt{\frac{2}{L}}\sum_n \int \frac{d^2k}{(2\pi)^2}\frac{1}{\sqrt{2\omega_n}}\sin\frac{n\pi}{L}z\bigg(\bar{a}_{k,n}e^{-i\omega_n t + i\vec{k}\vec{x}} + \bar{a}_{k,n}^+ e^{i\omega_n t - i\vec{k}\vec{x}}\bigg). \tag{6.208}$$

[11]Exercise: derive this result.

The creation and annihilation operators satisfy the commutation relations $[\bar{a}_{k,n},\bar{a}^+_{p,m}]=\delta_{nm}(2\pi)^2\delta^2(k-p)$, etc.

We use the result

$$\begin{aligned} D_F(x-y) &= <0|T\hat{\phi}(x)\hat{\phi}(y)|0> \\ &= \frac{1}{4\pi^2}\sum_{n=-\infty}^{+\infty}\bigg[\frac{1}{(x-y-2nLe_3)^2}-\frac{1}{(x-y-2(nL+x^3)e_3)^2}\bigg]. \end{aligned} \tag{6.209}$$

We introduce (with $a=-nL,-(nL+x^3)$)

$$\mathcal{D}_a=(x-y+2ae_3)^2=-(x^0-y^0)^2+(x^1-y^1)^2+(x^2-y^2)^2+(x^3-y^3+2a)^2. \tag{6.210}$$

We then compute

$$\partial_0^x\partial_0^y\frac{1}{\mathcal{D}_a}=-\frac{2}{\mathcal{D}_a^2}-8(x^0-y^0)^2\frac{1}{\mathcal{D}_a^3}. \tag{6.211}$$

$$\partial_i^x\partial_i^y\frac{1}{\mathcal{D}_a}=\frac{2}{\mathcal{D}_a^2}-8(x^i-y^i)^2\frac{1}{\mathcal{D}_a^3}\ ,\ i=1,2. \tag{6.212}$$

$$\partial_3^x\partial_3^y\frac{1}{\mathcal{D}_{-nL}}=\frac{2}{\mathcal{D}_{-nL}^2}-8(x^3-y^3+2nL)^2\frac{1}{\mathcal{D}_{-nL}^3}. \tag{6.213}$$

$$\partial_3^x\partial_3^y\frac{1}{\mathcal{D}_{-(nL+x^3)}}=-\frac{2}{\mathcal{D}_{-(nL+x^3)}^2}+8(x^3+y^3+2nL)^2\frac{1}{\mathcal{D}_{-(nL+x^3)}^3}. \tag{6.214}$$

We can immediately compute

$$\begin{aligned} <0|\hat{T}_{00}|0>^L_{\xi=0} &= \frac{1}{4\pi^2}\sum_{n=-\infty}^{+\infty}\bigg[\frac{2}{\mathcal{D}_{-nL}^2}-4(x^3-y^3+2nL)^2\frac{1}{\mathcal{D}_{-nL}^3}-4(x^3+y^3+2nL)^2\frac{1}{\mathcal{D}_{-(nL+x^3)}^3}\bigg] \\ &\longrightarrow -\frac{1}{32\pi^2}\sum_{n=-\infty}^{+\infty}\frac{1}{(nL)^4}-\frac{1}{16\pi^2}\sum_{n=-\infty}^{+\infty}\frac{1}{(nL+x^3)^4}. \end{aligned} \tag{6.215}$$

This is still divergent. The divergence comes from the original charge corresponding to $n=0$ in the first two terms in the limit $x\longrightarrow y$. All other terms coming from image charges are finite.

The same quantity evaluated in infinite space is

$$<0|\hat{T}_{00}|0>^{\infty}_{\xi=0} = \int\frac{d^3k}{(2\pi)^3}\frac{\omega_k}{2}e^{-ik(x-y)}. \tag{6.216}$$

This is divergent and the divergence must be the same divergence as in the case of parallel plates in the limit $L \longrightarrow \infty$, viz

$$< 0|\hat{T}_{00}|0 >^{\infty}_{\xi=0} = -\frac{1}{32\pi^2}\frac{1}{(nL)^4}|_{n=0}. \tag{6.217}$$

Hence the normal ordered vacuum expectation value of the energy-momentum-tensor is given by

$$< 0|\hat{T}_{00}|0 >^{L}_{\xi=0} - < 0|\hat{T}_{00}|0 >^{\infty}_{\xi=0} = -\frac{1}{32\pi^2}\sum_{n\neq 0}\frac{1}{(nL)^4} - \frac{1}{16\pi^2}\sum_{n=-\infty}^{+\infty}\frac{1}{(nL+x^3)^4}. \tag{6.218}$$

This is still divergent at the boundaries $x^3 \longrightarrow 0, L$.

In the conformal case we compute in a similar way the vacuum expectation value of the energy-momentum-tensor

$$\begin{aligned} < 0|\hat{T}_{00}|0 >^{L}_{\xi=\frac{1}{6}} &= \frac{1}{12\pi^2}\sum_{n=-\infty}^{+\infty}\bigg[-\frac{2}{\mathcal{D}^2_{-nL}} + \frac{4}{\mathcal{D}^2_{-(nL+x^3)}} - 4(x^3-y^3+2nL)^2\frac{1}{\mathcal{D}^3_{-nL}} \\ &- 4(x^3+y^3+2nL)^2\frac{1}{\mathcal{D}^3_{-(nL+x^3)}}\bigg] \\ &\longrightarrow -\frac{1}{32\pi^2}\sum_{n=-\infty}^{+\infty}\frac{1}{(nL)^4}. \end{aligned} \tag{6.219}$$

The normal ordered expression is

$$\begin{aligned} < 0|\hat{T}_{00}|0 >^{L}_{\xi=\frac{1}{6}} - < 0|\hat{T}_{00}|0 >^{\infty}_{\xi=\frac{1}{6}} &= -\frac{1}{32\pi^2}\sum_{n\neq 0}\frac{1}{(nL)^4} \\ &= -\frac{1}{16\pi^2L^4}\sum_{n=1}^{\infty}\frac{1}{n^4} \\ &= -\frac{1}{16\pi^2L^4}\zeta(4). \end{aligned} \tag{6.220}$$

The zeta function is given by

$$\zeta(4) = \sum_{n=1}^{\infty}\frac{1}{n^4} = \frac{\pi^4}{90}. \tag{6.221}$$

Thus

$$< 0|\hat{T}_{00}|0 >^{L}_{\xi=\frac{1}{6}} - < 0|\hat{T}_{00}|0 >^{\infty}_{\xi=\frac{1}{6}} = -\frac{\pi^2}{1440L^4}. \tag{6.222}$$

This is precisely the vacuum energy density of the conformal scalar field. The electromagnetic field is also a conformal field with two degrees of freedom and thus the corresponding vacuum energy density is

$$\rho_{\rm em} = -\frac{\pi^2}{720L^4}. \tag{6.223}$$

This corresponds to the attractive Casimir force. The energy between the two plates (where A is the surface area of the plates) is

$$E_{\rm em} = -\frac{\pi^2}{720L^4}AL. \tag{6.224}$$

The force is defined by

$$\begin{aligned} F_{\rm em} &= -\frac{dE_{\rm em}}{dL} \\ &= -\frac{\pi^2}{240L^4}A. \end{aligned} \tag{6.225}$$

The Casimir force is the force per unit area given by

$$\frac{F_{\rm em}}{A} = -\frac{\pi^2}{240L^4}. \tag{6.226}$$

6.4.4 From Renormalizable Field Theory

We consider the Lagrangian density (recall the metric is taken to be of signature $-+++...+$ and we will consider mostly $1+2$ dimensions)

$$\mathcal{L} = -\frac{1}{2}\partial_\mu\phi\partial^\mu\phi - \frac{1}{2}m^2\phi^2 - \frac{1}{2}\lambda\phi^2\sigma. \tag{6.227}$$

The static background field σ for parallel plates separated by a distance $2L$ will be chosen to be given by

$$\sigma = \frac{1}{\Delta}\bigg(\theta(|z| - L + \frac{\Delta}{2}) - \theta(|z| - L - \frac{\Delta}{2})\bigg). \tag{6.228}$$

Δ is the width of the plates and thus we are naturally interested in the sharp limit $\Delta \longrightarrow 0$. Obviously we have the normalization

$$\begin{aligned} \Delta\int dz\sigma(z) &= 2\int_{-L+\Delta/2}^{0} dz\theta(z) - 2\int_{-L-\Delta/2}^{0} dz\theta(z) \\ &= 2.\Delta. \end{aligned} \tag{6.229}$$

We compute the Fourier transform

$$\begin{aligned} \tilde{\sigma}(q) &= \int dz e^{iqz}\sigma(z) \\ &= \frac{1}{\Delta}\int_{-L-\Delta/2}^{-L+\Delta/2} dz e^{iqz} + \frac{1}{\Delta}\int_{L-\Delta/2}^{L+\Delta/2} dz e^{iqz} \\ &= \frac{4}{q\Delta}\cos qL\sin\frac{q\Delta}{2}. \end{aligned} \tag{6.230}$$

In the limit $\Delta \longrightarrow 0$ we obtain

$$\tilde{\sigma}(q) = 2\cos qL \to \sigma(z) = \delta(z-L) + \delta(z+L). \quad (6.231)$$

The boundary condition limit $\phi(\pm L) = 0$ is obtained by letting $\lambda \longrightarrow \infty$. This is the Dirichlet limit.

Before we continue let us give the Casimir force for parallel plates ($\sigma = \delta(z-a) + \delta(z+a)$) in the case of $1+1$ dimensions. This is given by

$$F(L,\lambda,m) = -\frac{\lambda^2}{\pi}\int_m^\infty \frac{t^2 dt}{\sqrt{t^2-m^2}} \frac{e^{-4Lt}}{4t^2 - 4\lambda t + \lambda^2(1-e^{-4Lt})}. \quad (6.232)$$

It vanishes quadratically in λ when $\lambda \longrightarrow 0$ as it should be since it is a force induced by the coupling of the scalar field ϕ to the background σ. In the boundary condition limit $\lambda \longrightarrow \infty$ we obtain

$$F(L,\infty,m) = -\frac{1}{\pi}\int_m^\infty \frac{t^2 dt}{\sqrt{t^2-m^2}} \frac{e^{-4Lt}}{1-e^{-4Lt}}. \quad (6.233)$$

This is independent of the material. Furthermore it reduces in the massless limit to the usual result, viz (with $a = 2L$)

$$F(L,\infty,0) = -\frac{\pi}{24a^2}. \quad (6.234)$$

The vacuum polarization energy of the field ϕ in the background σ is the Casimir energy. More precisely the Casimir energy is the vacuum energy in the presence of the boundary minus the vacuum energy without the boundary, viz

$$E[\sigma] = \frac{1}{2}\sum_n \omega_n[\sigma] - \frac{1}{2}\sum_n \omega_n[\sigma=0]. \quad (6.235)$$

The path integral is given by

$$Z = \int \mathcal{D}\phi e^{i\int d^Dx \mathcal{L}}. \quad (6.236)$$

The vacuum energy is given formally by

$$\begin{aligned} W[\sigma] &= \frac{1}{i}\ln Z \\ &= \frac{i}{2}Tr\ln\Big[\partial_\mu\partial^\mu - m^2 - \lambda\sigma\Big] + \text{constant}. \end{aligned} \quad (6.237)$$

Thus

$$W[\sigma] - W[\sigma=0] = \frac{i}{2}Tr\ln\Big[1 - \frac{1}{\partial^\mu\partial_\mu - m^2}\lambda\sigma\Big]. \quad (6.238)$$

The diagrammatic expansion of this term is given by the sum of all one-loop Feynman diagrams shown in figure 1 of reference [31]. The two-point function is obtained from W by differentiating with respect to an appropriate source twice, viz

$$G(x,y)=\frac{\delta^2 W[\sigma,J]}{\partial J(x)\partial J(y)}. \tag{6.239}$$

The two-point function is then what controls the Casimir energy. From the previous section we have for a massless theory the result

$$\begin{aligned} E[\sigma] &= \int d^3x < \hat{T}^{00} >_{\xi=0} \\ &= \frac{1}{2}\int d^3x(\partial_x^0\partial_y^0-\vec{\nabla}_x^2)D_{F\sigma}(x,y)|_{x=y} \\ &= \int\frac{d\omega}{2\pi}\omega^2\int d^3x D_{F\sigma}(\omega,\vec{x},\vec{x})+\text{constant}. \end{aligned} \tag{6.240}$$

In other words

$$E[\sigma]-E[\sigma=0] = \int\frac{d\omega}{2\pi}\omega^2\int d^3x\bigg[D_{F\sigma}(\omega,\vec{x},\vec{x})-D_{F0}(\omega,\vec{x},\vec{x})\bigg]. \tag{6.241}$$

As it turns the density of states created by the background is precisely [12]

$$\frac{dN}{d\omega} = \frac{\omega}{\pi}\bigg[D_{F\sigma}(\omega,\vec{x},\vec{x})-D_{F0}(\omega,\vec{x},\vec{x})\bigg]. \tag{6.242}$$

Using this last equation in the previous one gives precisely (6.235).

Alternatively we can rewrite the Casimir energy as

$$\begin{aligned} E[\sigma]-E[\sigma=0] &= \frac{1}{2}\int d^3x(\partial_x^0\partial_y^0-\vec{\nabla}_x^2)\bigg[D_{F\sigma}(x,y)-D_{F0}(x,y)\bigg]|_{x=y} \\ &= \frac{1}{2}\int d^3x(\partial_x^0\partial_y^0-\vec{\nabla}_x^2)\frac{1}{\partial_\mu\partial^\mu}\bigg[\frac{1}{\partial_\mu\partial^\mu}\lambda\sigma+\frac{1}{\partial_\mu\partial^\mu}\lambda\sigma\frac{1}{\partial_\mu\partial^\mu}\lambda\sigma+...\bigg]|_{x=y} \\ &= -\frac{1}{2}\int d^3x\bigg[\frac{1}{\partial_\mu\partial^\mu}\lambda\sigma+\frac{1}{\partial_\mu\partial^\mu}\lambda\sigma\frac{1}{\partial_\mu\partial^\mu}\lambda\sigma+...\bigg]|_{x=y}. \end{aligned} \tag{6.243}$$

This term is again given by the sum of all one-loop Feynman diagrams shown in figure 1 of reference [31]. We observe that

$$E[\sigma]-E[\sigma=0]=-i\lambda\frac{\partial}{\partial\lambda}\bigg[W[\sigma]-W[\sigma=0]\bigg]. \tag{6.244}$$

Both the one-point function (tadpole) and the two-point function (the self-energy) of the sigma field are superficially divergent for $D\leq 3$ and thus require renormalization. We introduce a counterterm given by

$$\mathcal{L}=c_1\sigma+c_2\sigma^2. \tag{6.245}$$

[12]Exercise: Construct an explicit argument.

The coefficients c_1 and c_2 are determined from the renormalization conditions

$$< \sigma >= 0. \tag{6.246}$$

$$< \sigma\sigma > |_{p^2=-\mu^2} = 0. \tag{6.247}$$

The $< \sigma >$ and $< \sigma\sigma >$ stand for proper vertices and not Green's functions of the field σ.

The total Casimir energy for a smooth background is finite. It can become divergent when the background becomes sharp ($\Delta \longrightarrow 0$) and strong ($\lambda \longrightarrow \infty$). The tadpole is always 0 by the renormalization condition. The two-point function of the sigma field diverges as we remove Δ and as a consequence the renormalized Casimir energy diverges in the Dirichlet limit. The three-point function also diverges (logarithmically) in the sharp limit whereas all higher orders in λ are finite.

Any further study of these issues and a detailed study of the competing perspective of Milton [22, 32, 33] is beyond the scope of these lectures.

6.4.5 Is Vacuum Energy Really Real?

The main point of [29] is that experimental confirmation of the Casimir effect does not really establish the reality of zero point fluctuations in quantum field theory. We leave the reader to go through the very sensible argumentation presented in that article.

Chapter 7

Horava-Lifshitz Gravity

7.1 The ADM Formulation

In this section we follow [1, 45].

We consider a fixed spacetime manifold $\mathcal{M}$ of dimension $D+1$. Let g_{ab} be the four-dimensional metric of the spacetime manifold $\mathcal{M}$. We consider a codimension-one foliation of the spacetime manifold $\mathcal{M}$ given by the spatial hypersurface (Cauchy surfaces) Σ_t of constant time t. Let n^a be the unit normal vector field to the hypersurfaces Σ_t. This induces a three-dimensional metric h_{ab} on each Σ_t given by the formula

$$h_{ab} = g_{ab} + n_a n_b. \tag{7.1}$$

The time flow in this spacetime will be given by a time flow vector field t^a which satisfies $t^a \nabla_a t = 1$. We decompose t^a into its normal and tangential parts with respect to the hypersurface Σ_t. The normal and tangential parts are given by the so-called lapse function N and shift vector N^a respectively defined by

$$N = -g_{ab} t^a n^b. \tag{7.2}$$

$$N^a = h^a{}_b t^b. \tag{7.3}$$

Let us make all this more explicit. Let $t = t(x^\mu)$ be a scalar function on the four-dimensional spacetime manifold $\mathcal{M}$ defined such that constant t gives a family of non-intersecting spacelike hypersurfaces Σ_t. Let y^i be the coordinates on the hypersurfaces Σ_t. We introduce a congruence of curves parameterized by t which connect the hypersurfaces Σ_t in such a way that points on each of the hypersurfaces intersected by the same curve are given the same spatial coordinates y^i. We have then

$$x^\mu \longrightarrow y^\mu = (t, y^i). \tag{7.4}$$

The tangent vectors to the hypersurface Σ_t are

$$e_i^\mu = \frac{\partial x^\mu}{\partial y^i}. \tag{7.5}$$

The tangent vectors to the congruence of curves is

$$t^\mu = \frac{\partial x^\mu}{\partial t}. \tag{7.6}$$

The vector t^μ satisfies trivially $t^\mu \nabla_\mu t = 1$, i.e. t^μ gives the direction of flow of time. The normal vector to the hypersurface Σ_t is defined by

$$n_\mu = -N \frac{\partial t}{\partial x^\mu}. \tag{7.7}$$

The normalization N is the lapse function. It is given precisely by (7.2), viz $N = -n_\mu t^\mu$. Clearly then N is the normal part of the vector t^μ with respect to the hypersurface Σ_t. Obviously we have $n_\mu e_i^\mu = 0$ and from the normalization $n_\mu n^\mu = -1$ we must also have

$$N^2 \frac{\partial t}{\partial x^\mu} \frac{\partial x_\mu}{\partial t} = -1 \ , \ N = (n^\mu \nabla_\mu t)^{-1}. \tag{7.8}$$

We can decompose t^μ as

$$t^\mu = N n^\mu + N^i e_i^\mu. \tag{7.9}$$

The three functions N^i define the components of the shift (spatial) vector. We compute immediately that

$$\begin{aligned} dx^\mu &= \frac{\partial x^\mu}{\partial t} dt + \frac{\partial x^\mu}{\partial y^i} dy^i \\ &= t^\mu dt + e_i^\mu dy^i \\ &= (N dt) n^\mu + (dy^i + N^i dt) e_i^\mu. \end{aligned} \tag{7.10}$$

Also

$$\begin{aligned} ds^2 &= g_{\mu\nu} dx^\mu dx^\nu \\ &= g_{\mu\nu} \bigg[N^2 dt^2 n^\mu n^\nu + (dy^i + N^i dt)(dy^j + N^j dt) e_i^\mu e_j^\nu \bigg] \\ &= -N^2 dt^2 + h_{ij} (dy^i + N^i dt)(dy^j + N^j dt). \end{aligned} \tag{7.11}$$

The three-dimensional metric h_{ij} is the induced metric on the hypersurface Σ_t. It is given explicitly by

$$h_{ij} = g_{\mu\nu} e_i^\mu e_j^\nu. \tag{7.12}$$

From the other hand in the coordinate system y^μ we have

$$\begin{aligned} ds^2 &= \gamma_{\mu\nu}dy^\mu dy^\nu \\ &= \gamma_{00}dt^2+2\gamma_{0j}dtdy^j+\gamma_{ij}dy^idy^j. \end{aligned} \tag{7.13}$$

By comparing (7.11) and (7.13) we obtain

$$\gamma_{\mu\nu}=\begin{pmatrix} \gamma_{00} & \gamma_{0j} \\ \gamma_{i0} & \gamma_{ij} \end{pmatrix}=\begin{pmatrix} -N^2+h_{ij}N^iN^j & h_{ij}N^i \\ h_{ij}N^j & h_{ij} \end{pmatrix}=\begin{pmatrix} -N^2+N^iN_i & N_j \\ N_i & h_{ij} \end{pmatrix}. \tag{7.14}$$

The condition $\gamma_{\mu\nu}\gamma^{\nu\lambda}=\delta_\mu^\lambda$ reads explicitly

$$\begin{aligned} &(-N^2+N^iN_i)\gamma^{00}+N_i\gamma^{i0}=1 \\ &(-N^2+N^iN_i)\gamma^{0j}+N_i\gamma^{ij}=0 \\ &N_j\gamma^{00}+h_{ij}\gamma^{i0}=0 \\ &N_j\gamma^{0k}+h_{ij}\gamma^{ik}=\delta_j^k. \end{aligned} \tag{7.15}$$

We define h^{ij} in the usual way, viz $h_{ij}h^{jk}=\delta_i^k$. We get immediately the solution

$$\gamma^{\mu\nu}=\begin{pmatrix} \gamma^{00} & \gamma^{0j} \\ \gamma^{i0} & \gamma^{ij} \end{pmatrix}=\begin{pmatrix} -\frac{1}{N^2} & \frac{1}{N^2}N^j \\ \frac{1}{N^2}N^i & h^{ij}-\frac{1}{N^2}N^iN^j \end{pmatrix}. \tag{7.16}$$

We also compute (we work in $1+2$ for simplicity)

$$\begin{aligned} \det\gamma &= \det\begin{pmatrix} -N^2+N^iN_i & N_1 & N_2 \\ N_1 & h_{11} & h_{12} \\ N_2 & h_{21} & h_{22} \end{pmatrix} \\ &= (-N^2+N^iN_i)\det h-N_1(N_1h_{22}-N_2h_{12})+N_2(N_1h_{21}-N_2h_{11}). \end{aligned} \tag{7.17}$$

By using $N_i=h_{ij}N^j$ we find

$$\det\gamma = -N^2\det h \tag{7.18}$$

We have then the result

$$\sqrt{-g}d^4x=\sqrt{-\gamma}d^4y=N\sqrt{h}d^4y. \tag{7.19}$$

We conclude that all information about the original four-dimensional metric $g_{\mu\nu}$ is contained in the lapse function N, the shift vector N^i and the three-dimensional metric h_{ij}.

The three-dimensional metric h_{ij} can also be understood in terms of projectors as follows. The projector normal to the hypersurface Σ_t is defined by

$$P^N_{\mu\nu}=-n_\mu n_\nu. \tag{7.20}$$

This satisfies $(P^N)^2 = P^N$ and $P^N n = n$ as it should. The normal component of any vector V^μ with respect to the hypersurface Σ_t is given by $V^\mu n_\mu$. The projector $P^N_{\mu\nu}$ can also be understood as the metric along the normal direction. Indeed we have

$$P^N_{\mu\nu} dx^\mu dx^\nu = -n_\mu n_\nu dx^\mu dx^\nu = -N^2 dt^2. \tag{7.21}$$

The tangent projector is then obviously given by

$$\begin{aligned} P^T_{\mu\nu} &= g_{\mu\nu} - P^N_{\mu\nu} \\ &= g_{\mu\nu} + n_\mu n_\nu. \end{aligned} \tag{7.22}$$

This should be understood as the metric along the tangent directions since

$$P^T_{\mu\nu} dx^\mu dx^\nu = ds^2 + N^2 dt^2 = h_{ij}(dy^i + N^i dt)(dy^j + N^j dt). \tag{7.23}$$

The three-dimensional metric is therefore given by

$$h_{\mu\nu} \equiv P^T_{\mu\nu} = g_{\mu\nu} + n_\mu n_\nu. \tag{7.24}$$

Indeed we have

$$h_{\mu\nu} \frac{\partial x^\mu}{\partial y^\alpha} \frac{\partial x^\nu}{\partial y^\beta} dy^\alpha dy^\beta = h_{ij}(dy^i + N^i dt)(dy^j + N^j dt). \tag{7.25}$$

Or equivalently

$$\begin{aligned} & h_{\mu\nu} e^\mu_i e^\nu_j = h_{ij} \Leftrightarrow g_{\mu\nu} e^\mu_i e^\nu_j = h_{ij} \\ & N_i = h_{\mu\nu} t^\mu e^\nu_i \Leftrightarrow N_i = h_{ij} N^j \\ & N_i N^i \equiv h_{\mu\nu} t^\mu t^\nu. \end{aligned} \tag{7.26}$$

We compute also

$$h^\mu_\nu t^\nu = g^{\mu\alpha} h_{\alpha\nu} t^\nu = N^i e^\mu_i \equiv N^\mu. \tag{7.27}$$

This should be compared with (7.3).

It is a theorem that the three-dimensional metric $h_{\mu\nu}$ will uniquely determine a covariant derivative operator on Σ_t. This will be denoted D_μ and defined in an obvious way by

$$D_\mu X_\nu = h^\alpha_\mu h^\beta_\nu \nabla_\alpha X_\beta. \tag{7.28}$$

In other words D_μ is the projection of the four-dimensional covariant derivative ∇_μ onto Σ_t.

A central object in the discussion of how the hypersurfaces Σ_t are embedded in the four-dimensional spacetime manifold $\mathcal{M}$ is the extrinsic curvature $K_{\mu\nu}$. This is given essentially by 1) comparing the normal vector n_μ at a point p and the parallel transport of the normal vector n_μ at a nearby point q along a geodesic connecting q to p on the hypersurface Σ_t and then 2) projecting the result onto the hypersurface Σ_t. The first part is clearly given by the covariant derivative whereas the projection is done through the three-dimensional metric tensor. Hence the extrinsic curvature must be defined by

$$\begin{aligned} K_{\mu\nu} &= -h_\mu^\alpha h_\nu^\beta \nabla_\alpha n_\beta \\ &= -h_\mu^\alpha \nabla_\alpha n_\nu. \end{aligned} \tag{7.29}$$

In the second line of the above equation we have used $n^\beta \nabla_\alpha n_\beta = 0$ and $\nabla_\alpha g_{\mu\nu} = 0$. We can check that $K_{\mu\nu}$ is symmetric and tangent, viz [1]

$$K_{\mu\nu} = K_{\nu\mu} \ , \ h_\mu^\alpha K_{\alpha\nu} = K_{\mu\nu}. \tag{7.30}$$

We recall the definition of the curvature tensor in four dimensions which is given by

$$(\nabla_\alpha \nabla_\beta - \nabla_\beta \nabla_\alpha)\omega_\mu = R_{\alpha\beta\mu}{}^{\nu}\omega_\nu. \tag{7.31}$$

By analogy the curvature tensor of Σ_t can be defined by

$$(D_\alpha D_\beta - D_\beta D_\alpha)\omega_\mu =^{(3)} R_{\alpha\beta\mu}{}^{\nu}\omega_\nu. \tag{7.32}$$

We compute

$$\begin{aligned} D_\alpha D_\beta \omega_\mu &= D_\alpha (h_\beta^\rho h_\mu^\nu \nabla_\rho \omega_\nu) \\ &= h_\alpha^\delta h_\beta^\theta h_\mu^\gamma \nabla_\delta (h_\theta^\rho h_\gamma^\nu \nabla_\rho \omega_\nu) \\ &= h_\alpha^\delta h_\beta^\rho h_\mu^\nu \nabla_\delta \nabla_\rho \omega_\nu - h_\mu^\nu K_{\alpha\beta} n^\rho \nabla_\rho \omega_\nu - h_\beta^\rho K_{\alpha\mu} n^\nu \nabla_\rho \omega_\nu. \end{aligned} \tag{7.33}$$

In the last line of the above equation we have used the result

$$h_\alpha^\delta h_\beta^\theta \nabla_\delta h_\theta^\rho = -K_{\alpha\beta} n^\rho. \tag{7.34}$$

We also compute

$$\begin{aligned} h_\beta^\rho n^\nu \nabla_\rho \omega_\nu &= h_\beta^\rho \nabla_\rho (n^\nu \omega_\nu) + K_\beta^\nu \omega_\nu \\ &= D_\rho (n^\nu \omega_\nu) + K_\beta^\nu \omega_\nu \\ &= K_\beta^\nu \omega_\nu. \end{aligned} \tag{7.35}$$

Thus

$$D_\alpha D_\beta \omega_\mu = h_\alpha^\delta h_\beta^\rho h_\mu^\nu \nabla_\delta \nabla_\rho \omega_\nu - h_\mu^\nu K_{\alpha\beta} n^\rho \nabla_\rho \omega_\nu - K_{\alpha\mu} K_\beta^\nu \omega_\nu. \tag{7.36}$$

Similar calculation gives

$$D_\beta D_\alpha \omega_\mu = h_\alpha^\delta h_\beta^\rho h_\mu^\nu \nabla_\rho \nabla_\delta \omega_\nu - h_\mu^\nu K_{\alpha\beta} n^\rho \nabla_\rho \omega_\nu - K_{\beta\mu} K_\alpha^\nu \omega_\nu. \tag{7.37}$$

[1]Exercise: Verify these results.

Hence we obtain the first Gauss-Codacci relation given by

$$ {}^{(3)}R_{\alpha\beta\mu}{}^{\nu}\omega_\nu = h^\delta_\alpha h^\rho_\beta h^\theta_\mu R_{\delta\rho\theta}{}^{\kappa}\omega_\kappa - K_{\alpha\mu}K^\nu_\beta\omega_\nu + K_{\beta\mu}K^\nu_\alpha\omega_\nu. \tag{7.38} $$

In other words

$$ {}^{(3)}R_{\alpha\beta\mu}{}^{\nu} = h^\delta_\alpha h^\rho_\beta h^\theta_\mu R_{\delta\rho\theta}{}^{\kappa}h^\nu_\kappa - K_{\alpha\mu}K^\nu_\beta + K_{\beta\mu}K^\nu_\alpha. \tag{7.39} $$

The first term represents the intrinsic part of the three-dimensional curvature obtained by simply projecting out the four-dimensional curvature onto the hypersurface Σ_t whereas the second term represents the extrinsic part of the three-dimensional curvature which arises from the embedding of Σ_t into the spacetime manifold.

The second Gauss-Codacci relation is given by

$$ D_\mu K^\mu_\nu - D_\nu K^\mu_\mu = -h^\alpha_\nu R_{\alpha\kappa}n^\kappa. \tag{7.40} $$

The proof goes as follows. We use $h^\nu_\mu h^\lambda_\nu = h^\lambda_\mu$ and $K^\lambda_\mu = g^{\lambda\nu}K_{\mu\nu} = -h^\alpha_\mu\nabla_\alpha n^\lambda$ to find

$$ \begin{aligned} D_\mu K^\mu_\nu - D_\nu K^\mu_\mu &= h^\rho_\mu h^\mu_\sigma h^\lambda_\nu \nabla_\rho K^\sigma_\lambda - h^\rho_\nu h^\mu_\sigma h^\lambda_\mu \nabla_\rho K^\sigma_\lambda \\ &= h^\rho_\sigma h^\lambda_\nu \nabla_\rho K^\sigma_\lambda - h^\rho_\sigma h^\lambda_\nu \nabla_\lambda K^\sigma_\rho \\ &= -h^\rho_\sigma h^\lambda_\nu \nabla_\rho(h^\alpha_\lambda\nabla_\alpha n^\sigma) + h^\rho_\sigma h^\lambda_\nu \nabla_\lambda(h^\alpha_\rho\nabla_\alpha n^\sigma) \\ &= -h^\rho_\sigma h^\lambda_\nu\bigg(\nabla_\rho h^\alpha_\lambda.\nabla_\alpha n^\sigma + h^\alpha_\lambda\nabla_\rho\nabla_\alpha n^\sigma - \nabla_\lambda h^\alpha_\rho.\nabla_\alpha n^\sigma - h^\alpha_\rho\nabla_\lambda\nabla_\alpha n^\sigma\bigg). \end{aligned} \tag{7.41} $$

The first and third terms are zero. Explicitly we have (using $\nabla_\alpha g_{\mu\nu} = 0$ and $n^\mu h^\nu_\mu = 0$)

$$ \begin{aligned} -h^\rho_\sigma h^\lambda_\nu\bigg(\nabla_\rho h^\alpha_\lambda.\nabla_\alpha n^\sigma - \nabla_\lambda h^\alpha_\rho.\nabla_\alpha n^\sigma\bigg) &= h^\lambda_\nu K_{\sigma\lambda}n^\alpha\nabla_\alpha n^\sigma - h^\rho_\sigma K_{\nu\rho}n^\alpha\nabla_\alpha n^\sigma \\ &= 0. \end{aligned} \tag{7.42} $$

We have then

$$ \begin{aligned} D_\mu K^\mu_\nu - D_\nu K^\mu_\mu &= -h^\rho_\sigma h^\lambda_\nu\big(h^\alpha_\lambda\nabla_\rho\nabla_\alpha n^\sigma - h^\alpha_\rho\nabla_\lambda\nabla_\alpha n^\sigma\big) \\ &= -h^\rho_\sigma h^\alpha_\nu\big(\nabla_\rho\nabla_\alpha - \nabla_\alpha\nabla_\rho\big)n^\sigma \\ &= -h^{\rho\sigma}h^\alpha_\nu R_{\rho\alpha\sigma\kappa}n^\kappa \\ &= h^{\rho\sigma}h^\alpha_\nu R_{\rho\alpha\kappa\sigma}n^\kappa \\ &= g^{\rho\sigma}h^\alpha_\nu R_{\rho\alpha\kappa\sigma}n^\kappa \\ &= h^\alpha_\nu R_{\rho\alpha\kappa}{}^{\rho}n^\kappa \\ &= -h^\alpha_\nu R_{\alpha\rho\kappa}{}^{\rho}n^\kappa \\ &= -h^\alpha_\nu R_{\alpha\kappa}n^\kappa. \end{aligned} \tag{7.43} $$

The goal now is to compute in terms of three-dimensional quantities the scalar curvature R. We start from

$$R = -Rg_{\mu\nu}n^\mu n^\nu = -2(R_{\mu\nu} - G_{\mu\nu})n^\mu n^\nu = -2R_{\mu\nu}n^\mu n^\nu + R_{\mu\nu\alpha\beta}h^{\mu\alpha}h^{\nu\beta}. \tag{7.44}$$

We compute

$$\begin{aligned} R_{\mu\nu\alpha\beta}h^{\mu\alpha}h^{\nu\beta} &= h_{\beta\rho}R_{\mu\nu\alpha}{}^{\rho}h^{\mu\alpha}h^{\nu\beta} \\ &= g^{\beta\eta}g^{\kappa\sigma}\big(h^\mu_\kappa h^\nu_\eta h^\alpha_\sigma R_{\mu\nu\alpha}{}^{\rho}h^\theta_\rho\big)h^\beta_\theta \\ &= g^{\beta\eta}g^{\kappa\sigma}\big({}^{(3)}R_{\kappa\eta\sigma}{}^{\theta} + K_{\kappa\sigma}K^\theta_\eta - K_{\eta\sigma}K^\theta_\kappa\big)h^\beta_\theta \\ &= g^{\kappa\sigma}\big({}^{(3)}R_{\kappa\eta\sigma}{}^{\theta} + K_{\kappa\sigma}K^\theta_\eta - K_{\eta\sigma}K^\theta_\kappa\big) \\ &= {}^{(3)}R + K^2 - K_{\mu\nu}K^{\mu\nu}. \end{aligned} \tag{7.45}$$

Next we compute

$$\begin{aligned} R_{\mu\nu}n^\mu n^\nu &= R_{\mu\alpha\nu}{}^{\alpha}n^\mu n^\nu \\ &= -g^{\alpha\rho}n^\nu R_{\nu\rho\alpha\mu}n^\mu \\ &= n^\nu\nabla_\mu\nabla_\mu n^\mu - n^\nu\nabla_\nu\nabla_\mu n^\mu \\ &= \nabla_\mu(n^\nu\nabla_\nu n^\mu) - \nabla_\nu(n^\nu\nabla_\mu n^\mu) - \nabla_\mu n^\nu.\nabla_\nu n^\mu + \nabla_\nu n^\nu.\nabla_\mu n^\mu. \end{aligned} \tag{7.46}$$

The rate of change of the normal vector along the normal direction is expressed by the quantity

$$a^\mu = n^\nu\nabla_\nu n^\mu. \tag{7.47}$$

We have

$$K = K^\mu_\mu = -h^\alpha_\mu\nabla_\alpha n^\mu = -g^\alpha_\mu\nabla_\alpha n^\mu = -\nabla_\mu n^\mu. \tag{7.48}$$

By using now $K_{\mu\nu} = -h^\alpha_\mu\nabla_\alpha n_\nu = -h^\alpha_\nu\nabla_\alpha n_\mu$ and $h^{\nu\beta}K_{\mu\nu} = K^\beta_\mu$ we can show that

$$\begin{aligned} K_{\mu\nu}K^{\mu\nu} &= -K_{\mu\nu}h^{\nu\beta}\nabla_\beta n^\mu \\ &= -K^\beta_\mu\nabla_\beta n^\mu \\ &= h^\rho_\mu\nabla_\rho n^\beta\nabla_\beta n^\mu \\ &= \nabla_\mu n^\beta.\nabla_\beta n^\mu. \end{aligned} \tag{7.49}$$

We obtain then the result

$$R_{\mu\nu}n^\mu n^\nu = \nabla_\mu(Kn^\mu + a^\mu) - K_{\mu\nu}K^{\mu\nu} + K^2. \tag{7.50}$$

The end result is given by

$$R = \mathcal{L}_{\rm ADM} - 2\nabla_\mu (K n^\mu + a^\mu). \tag{7.51}$$

The so-called ADM (Arnowitt, Deser and Misner) Lagrangian is given by

$$\mathcal{L}_{\rm ADM} = ^{(3)}R - K^2 + K_{\mu\nu}K^{\mu\nu}. \tag{7.52}$$

In other words

$$\sqrt{-g}\mathcal{L}_{\rm ADM} = \sqrt{h} N ({}^{(3)}R - K^2 + K_{\mu\nu}K^{\mu\nu}). \tag{7.53}$$

The extrinsic curvature $K_{\mu\nu}$ is the covariant analogue of the time derivative of the metric as we will now show. First we recall the definition of the Lie derivative of a tensor T along a vector V. For a function we have obviously $\mathcal{L}_V f = V(f) = V^\mu \partial_\mu f$ whereas for a vector the Lie derivative is defined by $\mathcal{L}_V U^\mu = [V, U]^\mu$. This is essentially the commutator which is the reason why the commutator is called sometimes the Lie bracket. The Lie derivative of an arbitrary tensor is given by

$$\mathcal{L}_V T^{\mu_1 ... \mu_k}_{\nu_1 ... \nu_l} = V^\sigma \nabla_\sigma T^{\mu_1 ... \mu_k}_{\nu_1 ... \nu_l} - \nabla_\lambda V^{\mu_1} T^{\lambda \mu_2 ... \mu_k}_{\nu_1 ... \nu_l} - ... + \nabla_{\nu_1} V^\lambda T^{\mu_1 ... \mu_k}_{\lambda \nu_2 ... \nu_l} + ... \tag{7.54}$$

A very important example is the Lie derivative of the metric given by

$$\mathcal{L}_V g_{\mu\nu} = \nabla_\mu V_\nu + \nabla_\nu V_\mu. \tag{7.55}$$

Let us now go back to the extrinsic curvature $K_{\mu\nu}$. We have (using $n^c h_{cb} = 0$, $t^c = N n^c + N^c$)

$$\begin{aligned}
K_{ab} &= -h_a^\alpha \nabla_\alpha n_b \\
&= -\frac{1}{2} h_a^\alpha h_b^\beta (\nabla_\alpha n_\beta + \nabla_\beta n_\alpha) \\
&= -\frac{1}{2} h_a^\alpha h_b^\beta \mathcal{L}_n h_{\alpha\beta} \\
&= -\frac{1}{2} h_a^\alpha h_b^\beta \big(n^c \nabla_c h_{\alpha\beta} + \nabla_\alpha n^c . h_{c\beta} + \nabla_\beta n^c . h_{\alpha c} \big) \\
&= -\frac{1}{2N} h_a^\alpha h_b^\beta \bigg(N n^c \nabla_c h_{\alpha\beta} + \nabla_\alpha (N n^c). h_{c\beta} + \nabla_\beta (N n^c). h_{\alpha c} \bigg) \\
&= -\frac{1}{2N} h_a^\alpha h_b^\beta \big(\mathcal{L}_t h_{\alpha\beta} - \mathcal{L}_N h_{\alpha\beta} \big).
\end{aligned} \tag{7.56}$$

However we have (using $N^c = h_d^c t^d$)

$$\begin{aligned}
h_a^\alpha h_b^\beta \mathcal{L}_N h_{\alpha\beta} &= h_a^\alpha h_b^\beta \bigg(N^c \nabla_c h_{\alpha\beta} + \nabla_\alpha N^c . h_{c\beta} + \nabla_\beta N^c . h_{c\alpha} \bigg) \\
&= t^d D_d h_{ab} + D_a N_b + D_b N_a \\
&= D_a N_b + D_b N_a.
\end{aligned} \tag{7.57}$$

The time derivative of the metric is defined by

$$\dot{h}_{ab} = h_a^\alpha h_b^\beta \mathcal{L}_t h_{\alpha\beta}. \tag{7.58}$$

Hence

$$K_{ab} = -\frac{1}{2N} \big(\dot{h}_{ab} - D_a N_b - D_b N_a \big). \tag{7.59}$$

7.2 Introducing Horava-Lifshitz Gravity

In this section we follow [46–48] but also [49–52].

We will consider a fixed spacetime manifold $\mathcal{M}$ of dimension $D+1$ with an extra structure given by a codimension-one foliation $\mathcal{F}$. Each leaf of the foliation is a spatial hypersurface Σ_t of constant time t with local coordinates given by x^i. Obviously general diffeomorphisms, including Lorentz transformations, do not respect the foliation $\mathcal{F}$. Instead we have invariance under the foliation preserving diffeomorphism group $\mathrm{Diff}_{\mathcal{F}}(\mathcal{M})$ consisting of space-independent time reparametrizations and time-dependent spatial diffeomorphisms given by

$$t \longrightarrow t^{'}(t)\ ,\ \vec{x} \longrightarrow \vec{x}^{'}(t,\vec{x}). \tag{7.60}$$

The infinitesimal generators are clearly given by

$$\delta t = f(t)\ ,\ \delta x^i = \xi^i(t,\vec{x}). \tag{7.61}$$

The time-dependent spatial diffeomorphisms allow us arbitrary changes of the spatial coordinates x^i on each constant time hypersurfaces Σ_t. The fact that time reparametrization is space-independent means that the foliation of the spacetime manifold $\mathcal{M}$ by the constant time hypersurfaces Σ_t is not a choice of coordinate, as in general relativity, but it is a physical property of spacetime itself.

This property of spacetime is implemented explicitly by positing that spacetime is anisotropic in the sense that time and space do not scale in the same way, viz

$$x^i \longrightarrow bx^i\ ,\ t \longrightarrow b^z t. \tag{7.62}$$

The exponent z is called the dynamical critical exponent and it measures the degree of anisotropy postulated to exist between space and time. This exponent is a dynamical quantity in the theory which is not determined by the gauge transformations corresponding to the foliation preserving diffeomorphisms. The above scaling rules (7.62) are not invariant under foliation preserving diffeomorphisms and they should only be understood as the scaling properties of the theory at the UV free field fixed point.

7.2.1 Lifshitz Scalar Field Theory

We start by explaining the above point a little further in terms of so-called Lifshitz field theory. Lifshitz scalar field theory describes a tricritical triple point at which three different phases (disorder, uniform (homogeneous) and non-uniform (spatially modulated)) meet. A Lifshitz scalar field is given by the action

$$S = \frac{1}{2}\int dt \int d^D x\big(\dot{\Phi}^2 - \frac{1}{4}(\Delta\Phi)^2\big). \tag{7.63}$$

This action defines a Gaussian (free) RG fixed point with anisotropic scaling rules (7.62) with $z = 2$. The two terms in the above action must have the same mass dimension and as a consequence we obtain $[t] = [x]^2$. By choosing $\hbar = 1$ the mass dimension of x is P^{-1} where P is some typical momentum and hence the mass dimension of t is P^{-2}. We have then

$$[x] = P^{-1} \ , \ [t] = P^{-2}. \tag{7.64}$$

The mass dimension of the scalar field is therefore given by

$$[\Phi] = P^{\frac{D-2}{2}}. \tag{7.65}$$

The values $z = 2$ and $(D-2)/2$ should be compared with the relativistic values $z = 1$ and $(D-1)/2$. The lower critical dimension of the Lifshitz scalar at which the two-point function becomes logarithmically divergent is $2+1$ instead of the usual $1+1$ of the relativistic scalar field.

We can add at the UV free fixed point a relevant perturbation given by

$$W = -\frac{c^2}{2}\int dt \int d^D x \partial_i \Phi \partial_i \Phi. \tag{7.66}$$

By using the various mass dimensions at the UV free fixed point the coupling constant c has mass dimension P. The theory will flow in the infrared to the value $z = 1$ since this perturbation dominates the second term of (7.63) at low energies. In other words at large distances Lorentz symmetry emerges accidentally.

This crucial result is also equivalent to the statement that the ground state wave function of the system (7.63) is given essentially by the above relevant perturbation. This can be shown as follows. The Hamiltonian derived from (7.63) is trivially given by

$$H = \frac{1}{2}\int d^D x \big(P^2 + \frac{1}{4}(\Delta\Phi)^2\big). \tag{7.67}$$

The term $(\Delta\Phi)^2$ appears therefore as the potential. The momentum P can be realized as

$$P = -i\frac{\delta}{\delta\Phi}. \tag{7.68}$$

The Hamiltonian can then be rewritten as

$$H = \frac{1}{2}\int d^D x Q^+ Q \ , \ Q = iP - \frac{1}{2}\Delta\Phi. \tag{7.69}$$

The ground state wave function is a functional of the scalar field Φ which satisfies $H\Psi_0[\Phi] = 0$ or equivalently

$$Q\Psi_0 = 0 \Rightarrow \big(\frac{\delta}{\delta\Phi} - \frac{1}{2}\Delta\Phi\big)\Psi_0[\Phi] = 0. \tag{7.70}$$

A simple solution is given by

$$\Psi_0[\Phi]=\exp\big(-\frac{1}{4}\int d^Dx\partial_i\Phi\partial_i\Phi\big). \tag{7.71}$$

The theory given by the action (7.63) satisfy the so-called detailed balance condition in the sense that the potential part can be derived from a variational principle given precisely by the action (7.66), viz

$$\frac{\delta W}{\delta\Phi}=c^2\Delta\Phi. \tag{7.72}$$

7.2.2 Foliation Preserving Diffeomorphisms and Kinetic Action

We will assume for simplicity that the global topology of spacetime is given by

$$\mathcal{M}=\mathbf{R}\times\Sigma. \tag{7.73}$$

Σ is a compact $D-$dimensional space with trivial tangent bundle. This is equivalent to the statement that all global topological effects will be ignored and all total derivative and boundary terms are dropped in the action.

The Riemannian structure on the foliation $\mathcal{F}$ is given by the three dimensional metric g_{ij}, the shift vector N_i and the lapse function N as in the ADM decomposition of general relativity. The lapse function can be either projectable or non-projectable depending on whether or not it depends on time only and thus it is constant on the spatial leafs or it depends on spacetime. As it turns out projectable Horava-Lifshitz gravity contains an extra degree of freedom known as the scalar graviton.

We want here to demonstrate some of the above results. We first write down the metric in the ADM decomposition as

$$\begin{aligned} ds^2 &= -N^2c^2dt^2+g_{ij}(dx^i+N^idt)(dx^j+N^jdt)\\ &= (-N^2+g_{ij}N^iN^j/c^2)(dx^0)^2+(g_{ij}N^j/c)dx^idx^0+(g_{ij}N^i/c)dx^jdx^0+g_{ij}dx^idx^j. \end{aligned} \tag{7.74}$$

Now we consider the general diffeomorphism transformation

$$\begin{aligned} x^{'0} &= x^0+cf(t,x^i)+O(\frac{1}{c})\\ x^{'i} &= x^i+\xi^i(t,x^j)+O(\frac{1}{c}). \end{aligned} \tag{7.75}$$

This is an expansion in powers of $1/c$. For simplicity we will also assume that the generators f and ξ^i are small. We compute immediately

$$\begin{aligned} g_{ij} &= \frac{\partial x^{'\mu}}{\partial x^i}\frac{\partial x^{'\nu}}{\partial x^j}g^{'}_{\mu\nu}\\ &= g_i^kg_j^lg^{'}_{kl}+g_i^k\frac{\partial\xi^l}{\partial x^j}g^{'}_{kl}+g_j^l\frac{\partial\xi^k}{\partial x^i}g^{'}_{kl}+c\frac{\partial f}{\partial x^j}g_i^kg^{'}_{k0}+c\frac{\partial f}{\partial x^i}g_j^kg^{'}_{k0}... \end{aligned} \tag{7.76}$$

In the limit $c \longrightarrow \infty$ the last two terms diverge and thus one must choose the generator of time reparametrization f such that $f = f(t)$. In this case the above diffeomorphism (7.75) becomes precisely a foliation preserving diffeomorphism. We obtain in this case

$$g_{ij}^{'} = g_{ij} - g_i^k \frac{\partial \xi^l}{\partial x^j} g_{kl} - g_j^l \frac{\partial \xi^k}{\partial x^i} g_{kl}. \tag{7.77}$$

Equivalently the gauge transformation of the three dimensional metric corresponding to a foliation preserving diffeomorphism is

$$\begin{aligned} \delta g_{ij} &= g_{ij}^{'}(x^{'}) - g_{ij}(x) \\ &= g_{ij}^{'}(x) - g_{ij}(x) + f \frac{\partial g_{ij}}{\partial t} + \xi^k \frac{\partial g_{ij}}{\partial x^k} \\ &= -\frac{\partial \xi^l}{\partial x^j} g_{il} - \frac{\partial \xi^k}{\partial x^i} g_{kj} + f \frac{\partial g_{ij}}{\partial t} + \xi^k \frac{\partial g_{ij}}{\partial x^k}. \end{aligned} \tag{7.78}$$

Similarly we compute the gauge transformation of the shift vector corresponding to a foliation preserving diffeomorphism as follows. We have

$$\begin{aligned} g_{i0} &= \frac{\partial x^{'\mu}}{\partial x^i} \frac{\partial x^{'\nu}}{\partial x^0} g_{\mu\nu}^{'} \\ &= g_{i0}^{'} + \frac{\partial f}{\partial t} g_{i0}^{'} + \frac{1}{c} \frac{\partial \xi^k}{\partial t} g_{ik}^{'} + \frac{\partial \xi^k}{\partial x^i} g_{k0}^{'}. \end{aligned} \tag{7.79}$$

Equivalently we have

$$g_{ij}^{'} N^{'j} = g_{ij} N^j - \frac{\partial f}{\partial t} N_i - \frac{\partial \xi^k}{\partial t} g_{ik} - \frac{\partial \xi^k}{\partial x^i} N_k. \tag{7.80}$$

We rewrite this as

$$g_{ij}(N^{'j} - N^j) = -\frac{\partial f}{\partial t} N_i - \frac{\partial \xi^k}{\partial t} g_{ik} + \frac{\partial \xi^l}{\partial x^j} g_{il} N^j. \tag{7.81}$$

We have then

$$\begin{aligned} \delta N_i &= g_{ij}^{'}(x^{'}) N^{'j}(x^{'}) - g_{ij}(x) N^j(x) \\ &= -\frac{\partial f}{\partial t} N_i + f \frac{\partial N_i}{\partial t} - \frac{\partial \xi^k}{\partial t} g_{ik} + \xi^k \frac{\partial N_i}{\partial x^k} - \frac{\partial \xi^k}{\partial x^i} N_k. \end{aligned} \tag{7.82}$$

A similar calculation for the lapse function goes as follows. We have

$$\begin{aligned} g_{00} &= \frac{\partial x^{'\mu}}{\partial x^0} \frac{\partial x^{'\nu}}{\partial x^0} g_{\mu\nu}^{'} \\ &= g_{00}^{'} + 2 \frac{\partial f}{\partial t} g_{00}^{'} + \frac{2}{c} \frac{\partial \xi^i}{\partial t} g_{i0}^{'}. \end{aligned} \tag{7.83}$$

Explicitly we find from this equation after some calculation (recalling that $g_{ij} N^j = N_i$)

$$N^{'}(x) - N(x) = -\frac{\partial f}{\partial t} N. \tag{7.84}$$

Thus

$$\begin{aligned}\delta N &= N^{'}(x^{'})-N(x)\\ &= f\frac{\partial N}{\partial t}+\xi^k\frac{\partial N}{\partial x^k}-\frac{\partial f}{\partial t}N.\end{aligned} \tag{7.85}$$

We can use the above gauge transformations to make the choice

$$N=1\ ,\ N_i=0. \tag{7.86}$$

These are called the Gaussian coordinates.

Now we want to write an action principle for this theory. It will be given by the difference of a kinetic term and a potential term as follows

$$S=S_K-S_V. \tag{7.87}$$

The kinetic term is formed from the most general scalar term compatible with foliation preserving diffeomorphisms which must be quadratic in the time derivative of the three dimensional metric in order to maintain unitarity. It must be of the canonical form $\int dtd^Dx\dot{\Phi}^2$. Explicitly we may write

$$S_K=\frac{1}{2\kappa^2}\int dtd^DxN\sqrt{g}\frac{\partial g_{ij}}{\partial t}G^{ijkl}\frac{\partial g_{kl}}{\partial t}. \tag{7.88}$$

The time derivative of the three dimensional metric in the above action (7.88) must in fact be replaced by K_{ij} while the metric G^{ijkl} on the space of metrics can be determined from the requirement of invariance under foliation preserving diffeomorphisms as we will show in the following.

We know from our study of the ADM decomposition of general relativity that the covariant time derivative of the three dimensional metric is given by the extrinsic curvature, viz

$$K_{ij} = -\frac{1}{2N}\big(\dot{g}_{ij}-\nabla_iN_j-\nabla_jN_i\big)\ ,\ \dot{g}_{ij}=g_i^ag_j^b\mathcal{L}_tg_{ab}. \tag{7.89}$$

In this section we have decided to denote the three dimensional covariant derivative by ∇_i in the same way that we have decided to denote the three dimensional metric by g_{ij}. We may choose the local coordinates such that the vector field t^a has components $(c,0,...,0)$ and as a consequence the diffeomorphism corresponding to time evolution is precisely given by $(x^0,x^1,...,x^D)\longrightarrow(x^0+\delta x^0,x^1,...,x^D)$ and hence $\dot{g}_{ij}=\partial g_{ij}/\partial t$.

From the ADM decomposition (7.53) we see that the combination $K_{ij}K^{ij}-K^2$ where $K=g^{ij}K_{ij}$ is the only combination which is invariant under four dimensional diffeomorphisms. Under the three dimensional (foliation preserving) diffeomorphisms it is obvious that both terms $K_{ij}K^{ij}$ and K^2 are, by construction, separately invariant. We are led therefore to consider the kinetic action

$$S_K=\frac{1}{2\kappa^2}\int dtd^DxN\sqrt{g}(K_{ij}K^{ij}-\lambda K^2). \tag{7.90}$$

Let us determine the mass dimension of the different objects. Let us set $\hbar = 1$. From the Heisenberg uncertainty principle we know that the mass dimension of x is precisely P^{-1} where P is some typical momentum. In order to reflect the properties of the fixed point we will set a scale Z of dimension $[Z] = [x]^z/[t]$ to be dimensionless, i.e. $[c] = P^{z-1}$. This choice is consistent with the scaling rules (7.62). The mass dimension of t is therefore given by P^{-z}. The volume element is hence of mass dimension

$$[dtd^Dx] = P^{-z-D}. \tag{7.91}$$

Now from the line element (7.74) we see that dx^i and N^idt have the same mass dimension and hence the mass dimension of N^i is P^{z-1}. The mass dimension of the line element ds^2 must be the same as the mass dimension of dx^2, i.e. $[ds] = P^{-1}$ and as a consequence $[g_{ij}] = P^0$. Similarly we can conclude that the mass dimension of N is P^0. In summary we have

$$[g_{ij}] = [N] = P^0 \ , \ [N^i] = P^{z-1}. \tag{7.92}$$

From the above results we conclude that the mass dimension of the extrinsic curvature is given by

$$[K_{ij}] = P^z. \tag{7.93}$$

We can now derive the mass dimension of the coupling constant κ. We have

$$[S_K] \equiv P^0 = \frac{1}{[\kappa]^2} P^{-z-D} P^{2z} \Longrightarrow [\kappa] = P^{\frac{z-D}{2}}. \tag{7.94}$$

Thus in $D = 3$ spatial dimensions we must have $z = 3$ in order for κ to be dimensionless and hence the theory power-counting renormalizable.

The second coupling constant λ is also dimensionless. It only appears because the two terms $K_{ij}K^{ij}$ and K^2 are separately invariant under the three dimensional (foliation preserving) diffeomorphisms.

The kinetic action (7.90) can be rewritten in a trivial way as

$$S_K = \frac{1}{2\kappa^2} \int dtd^DxN\sqrt{g}K_{ij}G^{ijkl}K_{kl}. \tag{7.95}$$

The metric on the space of metrics G^{ijkl} is a generalized version of the so-called Wheeler-DeWitt metric given explicitly by

$$G^{ijkl} = \frac{1}{2}(g^{ik}g^{jl} + g^{il}g^{jk}) - \lambda g^{ij}g^{kl}. \tag{7.96}$$

This is the only form consistent with three dimensional (foliation preserving) diffeomorphisms. Full spacetime diffeomorphism invariance corresponding to general relativity fixes the value of λ as $\lambda = 1$. The inverse of G is defined by

$$G_{ijmn}G^{mnkl} = \frac{1}{2}(g_i^kg_j^l + g_i^lg_j^k). \tag{7.97}$$

We find explicitly

$$G_{ijkl}=\frac{1}{2}(g_{ik}g_{jl}-g_{il}g_{jk})-\frac{\lambda}{D\lambda-1}g_{ij}g_{kl}. \tag{7.98}$$

We will always assume for $D=3$ that $\lambda\neq 1/3$ for obvious reasons. The precise role of λ is still not very clear and we will try to study it more carefully in the following.

7.2.3 Potential Action and Detail Balance

The total action of Horava-Lifshitz gravity is a difference between the kinetic action constructed above and a potential action, viz

$$S=S_K-S_V. \tag{7.99}$$

The potential term, in the spirit of effective field theory, must contain all terms consistent with the foliation preserving diffeomorphisms which are of mass dimension less or equal than the kinetic action. These potential terms will contain in general spatial derivatives but not time derivatives which are already taken into account in the kinetic action. These potential terms must be obviously scalars under foliation preserving diffeomorphisms.

The mass dimension of the kinetic term is $[K_{ij}K_{ij}]=P^{2z}=P^6$. Thus the potential action must contain all covariant scalars which are of mass dimensions less or equal than 6. These terms are built from g_{ij} and N and their spatial derivatives. Because g_{ij} and N are both dimensionless the scalar term of mass dimension n must contain n spatial derivatives since $[x_i]=P^{-1}$. For projectable Horava-Lifshitz gravity the lapse function does not depend on space and hence all terms can only depend on the metric g_{ij} and its spatial derivatives. Obviously terms with odd number of spatial derivatives are not covariant. There remains terms with mass dimensions 0, 2, 4 and 6.

The term of mass dimension 0 is precisely the cosmological constant while the term of mass dimension 2 is the Ricci scalar, viz.

$$\begin{aligned}&\text{mass dimension}=0\ ,\ R^0\\&\text{mass dimension}=2\ ,\ R.\end{aligned} \tag{7.100}$$

The terms of mass dimensions 4 and 6 are given by the lists

$$\begin{aligned}&\text{mass dimension}=4\ ,\ R^2,R_{ij}R^{ij}\\&\text{mass dimension}=6\ ,\ R^3,RR_i^jR_j^i,R_j^iR_k^jR_i^k,R\nabla^2R,\nabla_iR_{jk}\nabla^iR^{jk}.\end{aligned} \tag{7.101}$$

The operators of mass dimensions 0, 2 and 4 are relevant (super renormalizable) while the operators of dimension 6 are marginal (renormalizable). The quadratic terms modify the propagator and add interactions while cubic terms in the curvature provide only interaction terms.

The term $\nabla_i R_{jk}\nabla^j R^{ik}$ is not included in the list because it is given by a linear combination of the above terms up to a total derivative. The potential action of projectable Horava-Lifshitz gravity is then given by

$$S_V = \int dtd^Dx\sqrt{g}NV[g_{ij}]. \tag{7.102}$$

$$V[g_{ij}] = g_0 + g_1 R + g_2 R^2 + g_3 R_{ij}R^{ij} + g_4 R^3 + g_5 R R_{ij}R^{ij} + g_6 R_i^j R_j^k R_k^i + g_7 R\nabla^2 R + g_8 \nabla_i R_{jk}\nabla^i R^{jk}. \tag{7.103}$$

The lowest order potential coincides with general relativity. In general relativity the projectability condition can always be chosen at least locally as a gauge choice which is not the case for Horava-Lifshitz gravity.

A remark now on non-projectable Horava-Lifshitz gravity is in order. In this case the lapse function depend on time and space which matches the spacetime-dependence of the lapse function in general relativity. Furthermore it can be shown that $a_i = \partial_i \ln N$ transforms as a vector under the diffeomorphism group $\mathrm{Diff}_{\mathcal{F}}(\mathcal{M})$ and as a consequence more terms such as $a_i a^i$, $\nabla_i a^i$ must be included in the potential action. The lowest order potential in this case is found to be given by

$$V[g_{ij}] = g_0 + g_1 R + \alpha a_i a^i + \beta\nabla_i a^i. \tag{7.104}$$

It is very hard to see whether or not the RG flow of the coupling constants α and β goes to zero in the infrared in order to recover general relativity. In [53] it was shown that the non-vanishing of α and β in the IR leads to the existence of a scalar mode.

Alternatively we can rewrite the total action as follows. The first part is the Hilbert-Einstein action given by

$$S_{EH} = \frac{1}{2\kappa^2}\int dt \int d^Dx N\sqrt{g}\bigg[K_{ij}K^{ij} - K^2 - 2\kappa^2 g_1 R - 2\kappa^2 g_0\bigg]. \tag{7.105}$$

Recall that $[t] = P^{-3}$ and $[x] = P^{-1}$. We scale time as $t^{'} = \zeta^2 t$ where ζ is of mass dimension P. It is clear that $[t^{'}] = P^{-1} = [x]$ and thus in the new system of coordinates $(t^{'}, x^i)$ we can choose as usual $c = 1$. We have then

$$S_{EH} = \frac{1}{2(\kappa\zeta)^2}\int dt^{'} \int d^Dx N\sqrt{g}\bigg[K_{ij}K^{ij} - K^2 - 2(\kappa\zeta)^2 g_1 R - 2(\kappa\zeta)^2 g_0\bigg]. \tag{7.106}$$

The coupling constant g_1 is of mass dimension P^4. Thus we may choose g_1 or equivalently ζ such that

$$-2(\kappa\zeta)^2 g_1 = 1. \tag{7.107}$$

We can now make the identification

$$\frac{1}{2(\kappa\zeta)^2} = \frac{1}{2}M^2_{\text{Planck}} = \frac{1}{16\pi G_{\text{Newton}}}\ ,\ (\kappa\zeta)^2 g_0 = \Lambda. \tag{7.108}$$

Thus the Hilbert-Einstein action is given by

$$S_{EH}=\frac{1}{2}M_{\rm Planck}^2\int dt^{'}\int d^DxN\sqrt{g}\bigg[K_{ij}K^{ij}-K^2+R-2\Lambda\bigg]. \tag{7.109}$$

To obtain the Horava-Lifshitz action we need to add 8 Lorentz-violating terms given by (with $\xi=1-\lambda$ and $g_2=\hat{g}_2\zeta^2$, $g_3=\hat{g}_3\zeta^2$ since g_2 and g_3 are of mass dimensions P^2)

$$\begin{aligned}S_{LV} &= \frac{1}{2\kappa^2}\int dt\int d^DxN\sqrt{g}\xi K^2+\int dt\int d^DxN\sqrt{g}\bigg[-\hat{g}_2\zeta^2R^2-\hat{g}_3\zeta^2R_{ij}R^{ij}-g_4R^3-g_5RR_{ij}R^{ij}\\ &- g_6R_i^jR_j^kR_k^i-g_7R\nabla^2R-g_8\nabla_iR_{jk}\nabla^iR^{jk}\bigg].\end{aligned} \tag{7.110}$$

Equivalently

$$\begin{aligned}S_{LV} &= \frac{1}{2(\kappa\xi)^2}\int dt^{'}\int d^DxN\sqrt{g}\bigg[\xi K^2-2\hat{g}_2(\kappa\zeta)^2R^2-2\hat{g}_3(\kappa\zeta)^2R_{ij}R^{ij}-2g_4\kappa^2R^3-2g_5\kappa^2RR_{ij}R^{ij}\\ &- 2g_6\kappa^2R_i^jR_j^kR_k^i-2g_7\kappa^2R\nabla^2R-2g_8\kappa^2\nabla_iR_{jk}\nabla^iR^{jk}\bigg].\end{aligned} \tag{7.111}$$

We may set $\kappa=1$ for simplicity. These Lorentz-violating terms lead to a scalar mode for the graviton with mass of order $O(\xi)$. Furthermore these terms are not small since they become comparable to the Einstein-Hilbert action for momenta of the order $M_i=M_{\rm Pl}/{g_i}^{0.5}$, $i=2,3$ and $M_i=M_{\rm Pl}/{g_i}^{0.25}$, $i=4,5,6,7$. The Planck scale $M_{\rm Pl}$ is independent of the various Lorentz-violating scales M_i which can be driven arbitrarily high by fine tuning of the dimensionless coupling constants g_i.

We will now impose the condition of detailed balance on the potential action. Thus we require that the potential action is of the special form

$$S_V=\frac{\kappa^2}{8}\int dtd^Dx\sqrt{g}NE^{ij}G_{ijkl}E^{kl}. \tag{7.112}$$

The tensor E is derived from some Euclidean $D-$dimensional action W as follows

$$\sqrt{g}E^{ij}=\frac{\delta W}{\delta g_{ij}}. \tag{7.113}$$

It is clearly that with the detailed balance condition the potential is a perfect square. As it turns out detailed balance lead to a cosmological constant of the wrong sign and parity violation. However it remains true that renormalization with detailed balance condition of the $(D+1)-$dimensional theory is equivalent to the renormalization of the $D-$dimensional action W together with the renormalization of the relative couplings between kinetic and scalar terms which is clearly much simpler than renormalization of a generic theory in $(D+1)-$dimensions.

For theories which are spatially isotropic we can choose the action W to be precisely the Hilbert-Einstein action in D dimensions. This is a relativistic theory with Euclidean signature given by the action

$$W=\frac{1}{\kappa_W^2}\int d^Dx\sqrt{g}(R-2\Lambda_W). \tag{7.114}$$

A standard calculation gives

$$\delta W=\frac{1}{\kappa_W^2}\int d^Dx\sqrt{g}\delta g^{ij}\big(R_{ij}-\frac{1}{2}g_{ij}R+g_{ij}\Lambda_W\big). \tag{7.115}$$

Equivalently

$$\frac{\delta W}{\delta g^{ij}}=\frac{1}{\kappa_W^2}\sqrt{g}\big(R_{ij}-\frac{1}{2}g_{ij}R+g_{ij}\Lambda_W\big). \tag{7.116}$$

Thus

$$E_{ij}=\frac{1}{\kappa_W^2}\big(R_{ij}-\frac{1}{2}g_{ij}R+g_{ij}\Lambda_W\big). \tag{7.117}$$

The potential action becomes therefore

$$S_V=\frac{\kappa^2}{8\kappa_W^4}\int dtd^Dx\sqrt{g}N\big(R^{ij}-\frac{1}{2}g^{ij}R+g^{ij}\Lambda_W\big)G_{ijkl}\big(R^{kl}-\frac{1}{2}g^{kl}R+g^{kl}\Lambda_W\big). \tag{7.118}$$

For very short distances (UV) the curvature is clearly the dominant term in W and thus the potential action S_V is dominated by terms quadratic in the curvature. In this case the mass dimension of the potential action $P^{4-z-D}[\kappa]^2/[\kappa_W]^4$ must be equal to the mass dimension of the kinetic action $P^{z-D}/[\kappa]^2$. This leads to the results

$$[\kappa]^2=P^{z-D}\ ,\ \frac{[\kappa]^2}{[\kappa_W]^2}=P^{z-2}. \tag{7.119}$$

We have then anisotropic scaling with $z=2$ and power counting renormalizability in $1+2$ dimensions. In a spacetime with $1+3$ dimensions we have $[\kappa]^2=P^{z-3}$ and $[\kappa_W]^2=P^{-1}$. The fact that the coupling constant κ_W is dimensionfull means the above theory in $1+3$ dimensions can only work as an effective field theory valid which is up to energies set by the energy scale $1/[\kappa_W]^2$.

At large distances (IR) the dominant term in W is the cosmological constant Λ_W and thus the potential action is dominated by linear and quadratic terms in Λ_W. This is essentially equivalent to the Einstein-Hilbert gravity theory given by the combination $R-2\Lambda$ and thus effectively the anisotropic scaling becomes the usual value $z=1$. In other words in $1+3$ dimensions, the above Horava-Lifshitz gravity has a $z=2$ fixed point in the UV and flows to a $z=1$ fixed point in the IR.

However we really need to construct a Horava-Lifshitz gravity with a $z=3$ fixed point in the UV and flows to a $z=1$ fixed point in the IR. As explained before the $z=3$ anisotropic scaling in $1+3$ dimensions is exactly what is needed for power counting renormalizability. The theory must satisfy detailed balance and thus one must look for a tensor E_{ij} which is such that it gives a $z=3$ scaling. It is easy to convince ourselves that E_{ij} must be third order in spatial derivatives so that the dominant term in the potential action S_V contains six spatial derivatives and hence will balance the two time derivatives in the kinetic action. With such an E_{ij} we will have

$$[\kappa]^2 = P^{z-D} \ , \ \frac{[\kappa]^2}{[\kappa_W]^2} = P^{z-3}. \tag{7.120}$$

There is a unique candidate for E_{ij} which is known as the Cotton tensor. This is a tensor which is third order in spatial derivatives given explicitly by

$$C^{ij} = \epsilon^{ikl}\nabla_k(R_l^j - \frac{1}{4}Rg_l^j). \tag{7.121}$$

We now state some results concerning the Cotton tensor without any proof. This is a symmetric tensor $C^{ij} = C^{ji}$, traceless $g_{ij}C^{ij} = 0$, conserved $\nabla_i C^{ij} = 0$ which transforms under Weyl transformations of the metric $g_{ij} \longrightarrow \exp(2\Omega)g_{ij}$ as $C^{ij} \longrightarrow \exp(-5\Omega)C^{ij}$, i.e. it is conformal with weight $-5/2$.

In dimensions $D > 3$ conformal flatness of a Riemannian metric is equivalent to the vanishing of the Weyl tensor defined by

$$C_{ijkl} = R_{ijkl} - \frac{1}{D-2}\big(g_{ik}R_{jl} - g_{il}R_{jk} - g_{jk}R_{il} + g_{jl}R_{ik}\big) + \frac{1}{(D-1)(D-2)}(g_{ik}g_{jl} - g_{il}g_{jk})R. \tag{7.122}$$

We can verify that the Weyl tensor is the completely traceless part of the Riemann tensor. In $D = 3$ the Weyl tensor vanishes identically and conformal flatness becomes equivalent to the vanishing of the Cotton tensor.

The Cotton tensor can be derived from an action principle given precisely by the Chern-Simon gravitational action defined by

$$W = \frac{1}{w^2}\int_\Sigma \omega_3(\Gamma). \tag{7.123}$$

$$\begin{aligned} \omega_3(\Gamma) &= \mathrm{Tr}\Big(\Gamma\wedge d\Gamma + \frac{2}{3}\Gamma\wedge\Gamma\wedge\Gamma\Big) \\ &= \epsilon^{ijk}\Big(\Gamma^m_{il}\partial_j\Gamma^l_{km} + \frac{2}{3}\Gamma^n_{il}\Gamma^l_{jm}\Gamma^m_{kn}\Big)d^3x. \end{aligned} \tag{7.124}$$

Chapter 8

Note on References

The personal choice of references, used in these notes, includes: 1) Wald (general relativity and differential geometry), 2) Hartle (elementary exposition of cosmology and observational cosmology), 3) Carroll (black holes and advanced cosmology), 4) Mukhanov (inflationary cosmology: maybe the best book on cosmology especially for a theoretical physicist), 5) Birrell and Davies (QFT on curved backgrounds: one of the best QFT books I have ever seen). For a successful treatment of the problem of quantizing gravity we think that Horava-Lifshitz gravity is the only serious candidate which adhere to the tradition of QFT. The references on this topic are the original papers by Horava. These are the primary references followed here but more references can be found in the listing at the end of these lecture notes. However, we stress that the list of references included in these lectures only reflect the choice, preference and prejudice of the author and is not intended to be complete, exhaustive and thorough in any sense whatsoever.

Appendix A

Differential Geometry Primer

A.1 Manifolds

A.1.1 Maps, Open Set and Charts

Definition 1: A map ϕ between two sets M and N, viz $\phi : M \longrightarrow N$ is a rule which takes every element of M to exactly one element of N, i.e it takes M into N.

This is a generalization of the notion of a function. The set M is the domain of M while the subset of N that M gets mapped into the image of ϕ. We have the following properties:

- An injective (one-to-one) map is a map in which every element of N has at most one element of M mapped into it. Example: $f = e^x$ is injective.
- A surjective (onto) map is a map in which every element of N has at least one element of M mapped into it. Example: $f = x^3 - x$ is surjective.
- A bijective (and therefore invertible) map is a map which is both injective and surjective.
- A map from R^m to R^n is a collection of n functions ϕ^i of m variables x^i given by

$$\phi^i(x^1, ..., x^m) = y^i \ , \ i = 1, ..., n. \tag{A.1}$$

- The map $\phi : R^m \longrightarrow R^n$ is a C^p map if every component ϕ^i is at least a C^p function, i.e. if the pth derivative exists and is continuous. A C^∞ map is called a smooth map.
- A diffeomorphism is a bijective map $\phi : M \longrightarrow N$ which is smooth and with an inverse $\phi^{-1} : N \longrightarrow M$ which is also smooth. The two sets M and N are said to be diffeomorphic which means essentially that they are identical.

Definition 2: An open ball centered around a point $y \in R^n$ is the set of all points $x \in R^n$ such that $|x - y| < r$ for some $r \in R$ where $|x - y|^2 = \sum_{i=1}^n (x_i - y_i)^2$. This is clearly the inside of a sphere S^{n-1} in R^n of radius r centered around the point y.

Definition 3: An open set $V \subset R^n$ is a set in which every point $y \in V$ is the center of an open ball which is inside V. Clearly an open set is a union of open balls. Also it is obvious that an open set is the inside of a $(n-1)-$dimensional surface in R^n.

Definition 4: A chart (coordinate system) is a subset U of a set M together with a one-to-one map $\phi : U \longrightarrow R^n$ such that the image $V = \phi(U)$ is an open set in R^n. We say that U is an open set in M. The map $\phi : U \longrightarrow \phi(U)$ is clearly invertible. See figure $1.a$.

Definition 5: A C^∞ atals is a collection of charts $\{(U_\alpha, \phi_\alpha)\}$ which must satisfy the 2 conditions:

- The union is M, viz $\cup_\alpha U_\alpha = M$.
- If two charts U_α and U_β intersects then we can consider the maps $\phi_\alpha \circ \phi_\beta^{-1}$ and $\phi_\beta \circ \phi_\alpha^{-1}$ defined as

$$\phi_\alpha \circ \phi_\beta^{-1} : \phi_\beta(U_\alpha \cap U_\beta) \longrightarrow \phi_\alpha(U_\alpha \cap U_\beta) \ , \ \phi_\beta \circ \phi_\alpha^{-1} : \phi_\alpha(U_\alpha \cap U_\beta) \longrightarrow \phi_\beta(U_\alpha \cap U_\beta). \tag{A.2}$$

  Clearly $\phi_\alpha(U_\alpha \cap U_\beta) \subset R^n$ and $\phi_\beta(U_\alpha \cap U_\beta) \subset R^n$. See figure $1.b$. These two maps are required to be C^∞, i.e. smooth.

It is clear that definition 4 provides a precise formulation of the notion that a manifold "will locally look like R^n" whereas definition 5 provides a precise formulation of the statement that a manifold "will be constructed from pieces of R^n (in fact the open sets U_α) which are sewn together smoothly".

A.1.2 Manifold: Definition and Examples

Definition 6: A C^∞ $n-$dimensional manifold M is a set M together with a maximal atlas, i.e. an atlas which contains every chart that is compatible with the conditions of definition 5. This requirement means in particular that two identical manifolds defined by two different atlases will not be counted as different manifolds.

Example 1: The Euclidean spaces R^n, the spheres S^n and the tori T^n are manifolds.

Example 2: Riemann surfaces are two-dimensional manifolds. A Riemann surface of genus g is a kind of two-dimensional torus with a g holes. The two-dimensional torus has genus $g = 1$ whereas the sphere is a two-dimensional torus with genus $g = 0$.

Example 3: Every compact orientable two-dimensional surface without boundary is a Riemann surface and thus is a manifold.

Example 4: The group of rotations in R^n (which is denoted by $SO(n)$) is a manifold. Any Lie group is a manifold.

Example 5: The product of two manifolds M and $M^{'}$ of dimensions n and $n^{'}$ respectively is a manifold $M \times M^{'}$ of dimension $n + n^{'}$.

Example 6: We display on figure 2 few spaces which are not manifolds. The spaces displayed on figure 3 are manifolds but they are either "not differentiable" (the cone) or "with boundary" (the line segment).

Example 7: Let us consider the circle S^1. Let us try to cover the circle with a single chart (S^1, θ) where $\theta : S^1 \longrightarrow R$. The image $\theta(S^1)$ is not open in R if we include both $\theta = 0$ and $\theta = 2\pi$ since clearly $\theta(0) = \theta(2\pi)$ (the map is not bijective). If we do not include both points then the chart does not cover the whole space. The solution is to use (at least) two charts as shown on figure 4.

Example 8: We consider a sphere S^2 in R^3 defined by the equation $x^2 + y^2 + z^2 = 1$. First let us recall the stereographic projection from the north pole onto the plane $z = -1$. For any point P on the sphere (excluding the north pole) there is a unique line through the north pole $N = (0, 0, 1)$ and $P = (x, y, z)$ which intersects the $z = -1$ plane at the point $p^{'} = (X, Y)$. From the cross sections shown on figure 5 we have immediately

$$X = \frac{2x}{1-z} \; , \; Y = \frac{2y}{1-z}. \tag{A.3}$$

The first chart will be therefore given by the subset $U_1 = S^2 - \{N\}$ and the map

$$\phi_1(x, y, z) = (X, Y) = (\frac{2x}{1-z}, \frac{2y}{1-z}). \tag{A.4}$$

The stereographic projection from the south pole onto the plane $z = 1$. Again for any point P on the sphere (excluding the south pole) there is a unique line through the south pole $N^{'} = (0, 0, -1)$ and $P = (x, y, z)$ which intersects the $z = 1$ plane at the point $p^{'} = (X^{'}, Y^{'})$. Now we have

$$X^{'} = \frac{2x}{1+z} \; , \; Y^{'} = \frac{2y}{1+z}. \tag{A.5}$$

The second chart will be therefore given by the subset $U_2 = S^2 - \{N^{'}\}$ and the map

$$\phi_2(x, y, z) = (X^{'}, Y^{'}) = (\frac{2x}{1+z}, \frac{2y}{1+z}). \tag{A.6}$$

The two charts (U_1, ϕ_1) and (U_2, ϕ_2) cover the whole sphere. They overlap in the region $-1 < z < +1$. In this overlap region we have the map

$$(X^{'}, Y^{'}) = \phi_2 \circ \phi_1^{-1}(X, Y). \tag{A.7}$$

We compute first the inverse map ϕ_1^{-1} as

$$x=\frac{4}{4+X^2+Y^2}X\ ,\ y=\frac{4}{4+X^2+Y^2}Y\ ,\ z=-\frac{4-X^2-Y^2}{4+X^2+Y^2}. \tag{A.8}$$

Next by substituting in the formulas of $X^{'}$ and $Y^{'}$ we obtain

$$X^{'}=\frac{4X}{X^2+Y^2}\ ,\ Y^{'}=\frac{4Y}{X^2+Y^2}. \tag{A.9}$$

This is simply a change of coordinates.

A.1.3 Vectors and Directional Derivative

In special relativity Minkowski spacetime is also a vector space. In general relativity spacetime is a curved manifold and is not necessarily a vector space. For example the sphere is not a vector space because we do not know how to add two points on the sphere to get another point on the sphere. The sphere which is naturally embedded in R^3 admits at each point P a tangent plane. The notion of a "tangent vector space" can be constructed for any manifold which is embedded in R^n. As it turns out manifolds are generally defined in intrinsic terms and not as surfaces embedded in R^n (although they can: Whitney's embedding theorem) and as such the notion of a "tangent vector space" should also be defined in intrinsic terms,i.e. with reference only to the manifold in question.

Directional Derivative: There is a one-to-one correspondence between vectors and directional derivatives in R^n. Indeed the vector $v=(v^1,...,v^n)$ in R^n defines the directional derivative $\sum_\mu v^\mu\partial_\mu$ which acts on functions on R^n. These derivatives are clearly linear and satisfy the Leibniz rule. We will therefore define tangent vectors on a general manifold as directional derivatives which satisfy linearity and the Leibniz rule. Remark that the directional derivative $\sum_\mu v^\mu\partial_\mu$ is a map from the set of all smooth functions into R.

Definition 7: Let now $\mathcal{F}$ be the set of all smooth functions f on a manifold M, viz $f:M\longrightarrow R$. We define a tangent vector v at the point $p\in M$ as the map $v:\mathcal{F}\longrightarrow R$ which is required to satisfy linearity and the Leibniz rule, viz

$$v(af+bg)=av(f)+bv(g)\ ,\ v(fg)=f(p)v(g)+g(p)v(f)\ ,\ a,b\in R\ ,\ f,g\in\mathcal{F}. \tag{A.10}$$

We have the following results:

- For a constant function ($h(p)=c$) we have from linearity $v(c^2)=cv(c)$ whereas the Leibniz rule gives $v(c^2)=2cv(c)$ and thus $v(c)=0$.

- The set V_p of all tangents vectors v at p form a vector space since $(v_1+v_2)(f)=v_1(f)+v_2(f)$ and $(av)(f)=av(f)$ where $a\in R$.

- The dimension of V_p is precisely the dimension n of the manifold M. The proof goes as follows. Let $\phi : O \subset M \longrightarrow U \subset R^n$ be a chart which includes the point p. Clearly for any $f \in \mathcal{F}$ the map $f \circ \phi^{-1} : U \longrightarrow R$ is smooth since both f and ϕ are smooth maps. We define the maps $X_\mu : \mathcal{F} \longrightarrow R$, $\mu = 1, ..., n$ by

$$X_\mu(f) = \frac{\partial}{\partial x^\mu}(f \circ \phi^{-1})|_{\phi(p)}. \quad (A.11)$$

Given a smooth function $F : R^n \longrightarrow R$ and a point $a = (a^1, ..., a^n) \in R^n$ then there exists smooth functions H_μ such that for any $x = (x^1, ..., x^n) \in R^n$ we have the result

$$F(x) = F(a) + \sum_{\mu=1}^{n}(x^\mu - a^\mu)H_\mu(x) \ , \ H_\mu(a) = \frac{\partial F}{\partial x^\mu}|_{x=a}. \quad (A.12)$$

We choose $F = f \circ \phi^{-1}$, $x \in U$ and $a = \phi(p) \in U$ we have

$$f \circ \phi^{-1}(x) = f \circ \phi^{-1}(a) + \sum_{\mu=1}^{n}(x^\mu - a^\mu)H_\mu(x). \quad (A.13)$$

Clearly $\phi^{-1}(x) = q \in O$ and thus

$$f(q) = f(p) + \sum_{\mu=1}^{n}(x^\mu - a^\mu)H_\mu(\phi(q)). \quad (A.14)$$

We think of each coordinate x^μ as a smooth function from U into R, viz $x^\mu : U \longrightarrow R$. Thus the map $x^\mu \circ \phi : O \longrightarrow R$ is such that $x^\mu(\phi(q)) = x^\mu$ and $x^\mu(\phi(p)) = a^\mu$. In other words

$$f(q) = f(p) + \sum_{\mu=1}^{n}(x^\mu \circ \phi(q) - x^\mu \circ \phi(p))H_\mu(\phi(q)). \quad (A.15)$$

Let now v be an arbitrary tangent vector in V_p. We have immediately

$$\begin{aligned} v(f) &= v(f(p)) + \sum_{\mu=1}^{n} v(x^\mu \circ \phi - x^\mu \circ \phi(p))H_\mu \circ \phi(q)|_{q=p} + \sum_{\mu=1}^{n}(x^\mu \circ \phi(q) - x^\mu \circ \phi(p))|_{q=p} v(H_\mu \circ \phi) \\ &= \sum_{\mu=1}^{n} v(x^\mu \circ \phi)H_\mu \circ \phi(p). \end{aligned} \quad (A.16)$$

But

$$H_\mu \circ \phi(p) = H_\mu(a) = \frac{\partial}{\partial x^\mu}(f \circ \phi^{-1})|_{x=a} = X_\mu(f). \quad (A.17)$$

Thus

$$v(f) = \sum_{\mu=1}^{n} v(x^\mu \circ \phi)X_\mu(f) \Rightarrow v = \sum_{\mu=1}^{n} v^\mu X_\mu \ , \ v^\mu = v(x^\mu \circ \phi). \quad (A.18)$$

This shows explicitly that the X_μ satisfy linearity and the Leibniz rule and thus they are indeed tangent vectors to the manifold M at p. The fact that an arbitrary tangent vector v can be expressed as a linear combination of the n vectors X_μ shows that the vectors X_μ are linearly independent, span the vector space V_p and that the dimension of V_p is exactly n.

Coordinate Basis: The basis $\{X_\mu\}$ is called a coordinate basis. We may pretend that

$$X_\mu \equiv \frac{\partial}{\partial x^\mu}. \tag{A.19}$$

Indeed if we work in a different chart $\phi^{'}$ we will have

$$X^{'}_\mu(f) = \frac{\partial}{\partial x^{'\mu}}(f\circ \phi^{'-1})|_{x^{'}=\phi^{'}(p)}. \tag{A.20}$$

We compute

$$\begin{aligned} X_\mu(f) &= \frac{\partial}{\partial x^\mu}(f\circ\phi^{-1})|_{x=\phi(p)} \\ &= \frac{\partial}{\partial x^\mu} f\circ\phi^{'-1}(\phi^{'}\circ\phi^{-1})|_{x=\phi(p)} \\ &= \sum_{\nu=1}^{n}\frac{\partial x^{'\nu}}{\partial x^\mu}\frac{\partial}{\partial x^{'\nu}}(f\circ\phi^{'-1}(x^{'}))|_{x^{'}=\phi^{'}(p)} \\ &= \sum_{\nu=1}^{n}\frac{\partial x^{'\nu}}{\partial x^\mu}X^{'}_\nu(f). \end{aligned} \tag{A.21}$$

The tangent vector v can be rewritten as

$$v = \sum_{\mu=1}^{n} v^\mu X_\mu = \sum_{\mu=1}^{n} v^{'\mu}X^{'}_\mu. \tag{A.22}$$

We conclude immediately that

$$v^{'\nu} = \sum_{\nu=1}^{n}\frac{\partial x^{'\nu}}{\partial x^\mu}v^\mu. \tag{A.23}$$

This is the vector transformation law under the coordinate transformation $x^\mu \longrightarrow x^{'\mu}$.

Vectors as Directional Derivatives: A smooth curve on a manifold M is a smooth map from R into M, viz $\gamma : R \longrightarrow M$. A tangent vector at a point p can be thought of as a directional derivative operator along a curve which goes through p. Indeed a tangent vector T at $p = \gamma(t) \in M$ can be defined by

$$T(f) = \frac{d}{dt}(f\circ\gamma(t))|_p. \tag{A.24}$$

The function f is $\in \mathcal{F}$ and thus $f\circ\gamma : R \longrightarrow R$. Given a chart ϕ the point p will be given by $p=\phi^{-1}(x)$ where $x=(x^1,...,x^n)\in R^n$. Hence

$$\gamma(t)=\phi^{-1}(x). \tag{A.25}$$

In other words the map γ is mapped into a curve $x(t)$ in R^n. We have immediately

$$T(f)=\frac{d}{dt}(f\circ\phi^{-1}(x))|_p=\sum_{\mu=1}^n\frac{\partial}{\partial x^\mu}(f\circ\phi^{-1}(x))\frac{dx^\mu}{dt}|_p=\sum_{\mu=1}^n X_\mu(f)\frac{dx^\mu}{dt}|_p. \tag{A.26}$$

The components T^μ of the vector T are therefore given by

$$T^\mu=\frac{dx^\mu}{dt}|_p. \tag{A.27}$$

A.1.4 Dual Vectors and Tensors

Definition 8: Let V_p be the tangent vector space at a point p of a manifold M. Let V_p^* be the space of all linear maps ω^* from V_p into R, viz $\omega^* : V_p \longrightarrow R$. The space V_p^* is the so-called dual vector space to V_p where addition and multiplication by scalars are defined in an obvious way. The elements of V_p^* are called dual vectors.

The dual vector space V_p^* is also called the cotangent dual vector space at p (also the vector space of one-forms at p). The elements of V_p^* are then called cotangent dual vectors. Another nomenclature is to refer to the elements of V_p^* as covariant vectors whereas the elements of V_p are referred to as contravariant vectors.

Dual Basis: Let X_μ, $\mu=1,...,n$ be a basis of V_p. The basis elements of V_p^* are given by vectors $X^{\mu*}$, $\mu=1,...,n$ which are defined by

$$X^{\mu*}(X_\nu)=\delta^\mu_\nu. \tag{A.28}$$

The Kronecker delta is defined in the usual way. The proof that $\{X^{\mu*}\}$ is a basis is straightforward. The basis $\{X^{\mu*}\}$ of V_p^* is called the dual basis to the basis $\{X_\mu\}$ of V_p. The basis elements X_μ may be thought of as the partial derivative operators $\partial/\partial x^\mu$ since they transform under a change of coordinate systems (corresponding to a change of charts $\phi\longrightarrow\phi^{'}$) as

$$X_\mu \;=\; \sum_{\nu=1}^n\frac{\partial x^{'\nu}}{\partial x^\mu}X^{'}_\nu. \tag{A.29}$$

We immediately deduce that we must have the transformation law

$$X^{\mu*} \;=\; \sum_{\nu=1}^n\frac{\partial x^\mu}{\partial x^{'\nu}}X^{\nu*^{'}}. \tag{A.30}$$

Indeed we have in the transformed basis

$$X^{\mu*^{'}}(X^{'}_\nu)=\delta^\mu_\nu. \tag{A.31}$$

From this result we can think of the basis elements $X^{\mu *}$ as the gradients dx^{μ}, viz

$$X^{\mu *'} \equiv dx^{\mu}. \tag{A.32}$$

Let $v = \sum_{\mu} v^{\mu} X_{\mu}$ be an arbitrary tangent vector in V_p, then the action of the dual basis elements $X^{\mu *}$ on v is given by

$$X^{\mu *}(v) = v^{\mu}. \tag{A.33}$$

The action of a general element $\omega^* = \sum_{\mu} \omega_{\mu} X^{\mu *}$ of V_p^* on v is given by

$$\omega^*(v) = \sum_{\mu} \omega_{\mu} v^{\mu}. \tag{A.34}$$

Recall the transformation law

$$v^{'\nu} = \sum_{\nu=1}^{n} \frac{\partial x^{'\nu}}{\partial x^{\mu}} v^{\mu}. \tag{A.35}$$

Again we conclude the transformation law

$$\omega_{\nu}^{'} = \sum_{\nu=1}^{n} \frac{\partial x^{\mu}}{\partial x^{'\nu}} \omega_{\mu}. \tag{A.36}$$

Indeed we confirm that

$$\omega^*(v) = \sum_{\mu} \omega_{\mu}^{'} v^{'\mu}. \tag{A.37}$$

Double Dual Vector Space: Let now V_p^{**} be the space of all linear maps v^{**} from V_p^* into R, viz $v^{**} : V_p^* \longrightarrow R$. The vector space V_p^{**} is naturally isomorphic (an isomorphism is one-to-one and onto map) to the vector space V_p since to each vector $v \in V_p$ we can associate the vector $v^{**} \in V_p^{**}$ by the rule

$$v^{**}(\omega^*) = \omega^*(v) \ , \ \omega^* \in V_p^*. \tag{A.38}$$

If we choose $\omega^* = X^{\mu *}$ and $v = X_{\nu}$ we get $v^{**}(X^{\mu *}) = \delta_{\nu}^{\mu}$. We should think of v^{**} in this case as $v = X_{\nu}$.

Definition 9**:** A tensor T of type (k, l) over the tangent vector space V_p is a multilinear map form $(V_p^* \times V_p^* \times ... \times V_p^*) \times (V_p \times V_p \times ... \times V_p)$ (with k cotangent dual vector space V_p^* and l tangent vector space V_p) into R, viz

$$T : V_p^* \times V_p^* \times ... \times V_p^* \times V_p \times V_p \times ... \times V_p \longrightarrow R. \tag{A.39}$$

The vectors $v \in V_p$ are therefore tensors of type $(1, 0)$ whereas the cotangent dual vectors $v \in V_p^*$ are tensors of type $(0, 1)$. The space $\mathcal{T}(k, l)$ of all tensors of type (k, l) is a vector space (obviously) of dimension $n^k.n^l$ since $\dim V_p = \dim V_p^* = n$.

Contraction: The contraction of a tensor T with respect to its ith cotangent dual vector and jth tangent vector positions is a map $C : \mathcal{T}(k,l) \longrightarrow \mathcal{T}(k-1,l-1)$ defined by

$$CT = \sum_{\mu=1}^{n} T(..., X^{\mu *}, ...; ..., X_{\mu}, ...). \tag{A.40}$$

The basis vector $X^{\mu *}$ of the cotangent dual vector space V_p^* is inserted into the ith position whereas the basis vector X_μ of the tangent vector space V_p is inserted into the jth position.

A tensor of type $(1,1)$ can be viewed as a linear map from V_p into V_p since for a fixed $v \in V_p$ the map $T(.,v)$ is an element of V_p^{**} which is the same as V_p, i.e. $T(.,v)$ is a map from V_p into V_p. From this result it is obvious that the contraction of a tensor of the type $(1,1)$ is essentially the trace and as such it must be independent of the basis $\{X_\mu\}$ and its dual $\{X^{\mu *}\}$. Contraction is therefore a well defined operation on tensors.

Outer Product: Let T be a tensor of type (k,l) and "components" $T(X^{1*}, ..., X^{k*}; Y_1, ..., Y_l)$ and $T^{'}$ be a tensor of type $(k^{'}, l^{'})$ and components $T^{'}(X^{k+1*}, ..., X^{k+k^{'}*}; Y_{l+1}, ..., Y_{l+l^{'}})$. The outer product of these two tensors which we denote $T \otimes T^{'}$ is a tensor of type $(k+k^{'}, l+l^{'})$ defined by the "components" $T(X^{1*}, ..., X^{k*}; Y_1, ..., Y_l)T^{'}(X^{k+1*}, ..., X^{k+k^{'}*}; Y_{l+1}, ..., Y_{l+l^{'}})$.

Simple Tensors: Simple tensors are tensors obtained by taking the outer product of cotangent dual vectors and tangent vectors. The $n^k.n^l$ simple tensors $X_{\mu_1} \otimes ... \otimes X_{\mu_k} \otimes X^{\nu_1 *} \otimes ... \otimes X^{\nu_l *}$ form a basis of the vector space $\mathcal{T}(k,l)$. In other words any tensor T of type (k,l) can be expanded as

$$T = \sum_{\mu_i} \sum_{\nu_i} T^{\mu_1 ... \mu_k}{}_{\nu_1 ... \nu_l} X_{\mu_1} \otimes ... \otimes X_{\mu_k} \otimes X^{\nu_1 *} \otimes ... \otimes X^{\nu_l *}. \tag{A.41}$$

By using $X^{\mu *}(X_\nu) = \delta^{\mu\nu}$ and $X_\mu(X^{\nu *}) = \delta^{\mu\nu}$ we calculate

$$T^{\mu_1 ... \mu_k}{}_{\nu_1 ... \nu_l} = T(X^{\mu_1 *} \otimes ... \otimes X^{\mu_k *} \otimes X_{\nu_1} \otimes ... \otimes X_{\nu_l}). \tag{A.42}$$

These are the components of the tensor T in the basis $\{X_\mu\}$. The contraction of the tensor T is now explicitly given by

$$(CT)^{\mu_1 ... \mu_{k-1}}{}_{\nu_1 ... \nu_{l-1}} = \sum_{\mu=1}^{n} T^{\mu_1 ... \mu ... \mu_{k-1}}{}_{\nu_1 ... \mu ... \nu_{l-1}} \tag{A.43}$$

The outer product of two tensors can also be given now explicitly in the basis $\{X_\mu\}$ in a quite obvious way.

We conclude by writing down the transformation law of a tensor under a change of coordinate systems. The transformation law of $X_{\mu_1} \otimes ... \otimes X_{\mu_k} \otimes X^{\nu_1 *} \otimes ... \otimes X^{\nu_l *}$ is obviously given by

$$X_{\mu_1}\otimes ...\otimes X_{\mu_k}\otimes X^{\nu_1 *}\otimes ...\otimes X^{\nu_l *}=\sum_{\mu_i^{'}}\sum_{\nu_i^{'}}\frac{\partial x^{'\mu_1^{'}}}{\partial x^{\mu_1}}...\frac{\partial x^{'\mu_k^{'}}}{\partial x^{\mu_k}}\frac{\partial x^{\nu_1}}{\partial x^{'\nu_1^{'}}}...\frac{\partial x^{\nu_l}}{\partial x^{'\nu_l^{'}}}X_{\mu_1^{'}}\otimes ...\otimes X_{\mu_k^{'}}\otimes X^{\nu_1^{'} *}\otimes ...\otimes X^{\nu_l^{'} *}. \quad (A.4$$

Thus we must have

$$T=\sum_{\mu_i}\sum_{\nu_i}T^{'\mu_1...\mu_k}{}_{\nu_1...\nu_l}X_{\mu_1^{'}}\otimes ...\otimes X_{\mu_k^{'}}\otimes X^{\nu_1^{'} *}\otimes ...\otimes X^{\nu_l^{'} *}. \quad (A.45)$$

The transformed components $T^{'\mu_1...\mu_k}{}_{\nu_1...\nu_l}$ are defined by

$$T^{'\mu_1^{'}...\mu_k^{'}}{}_{\nu_1^{'}...\nu_l^{'}}=\sum_{\mu_i}\sum_{\nu_i}\frac{\partial x^{'\mu_1^{'}}}{\partial x^{\mu_1}}...\frac{\partial x^{'\mu_k^{'}}}{\partial x^{\mu_k}}\frac{\partial x^{\nu_1}}{\partial x^{'\nu_1^{'}}}...\frac{\partial x^{\nu_l}}{\partial x^{'\nu_l^{'}}}T^{\mu_1...\mu_k}{}_{\nu_1...\nu_l}. \quad (A.46)$$

A.1.5 Metric Tensor

A metric g is a tensor of type $(0,2)$, i.e. a linear map from $V_p\times V_p$ into R with the following properties:

- The map $g: V_p\times V_p\longrightarrow R$ is symmetric in the sense that $g(v_1,v_2)=g(v_2,v_1)$ for any $v_1,v_2\in V_p$.
- The map g is nondegenerate in the sense that if $g(v,v_1)=0$ for all $v\in V_p$ then one must have $v_1=0$.
- In a coordinate basis where the components of the metric are denoted by $g_{\mu\nu}$ we can expand the metric as

$$g=\sum_{\mu,\nu}g_{\mu\nu}dx^\mu\otimes dx^\nu. \quad (A.47)$$

  This can also be rewritten symbolically as

$$ds^2=\sum_{\mu,\nu}g_{\mu\nu}dx^\mu dx^\nu. \quad (A.48)$$

- The map g provides an inner product on the tangent space V_p which is not necessarily positive definite. Indeed given two vectors v and w of V_p, their inner product is given by

$$g(v,w)=\sum_{\mu,\nu}g_{\mu\nu}v^\mu w^\nu. \quad (A.49)$$

  By choosing $v=w=\delta x=x_f-x_i$ we see that $g(\delta x,\delta x)$ is an infinitesimal squared distance between the points f and i. Hence the use of the name "metric" for the tensor g. In fact $g(\delta x,\delta x)$ is the generalization of the interval (also called line element) of special relativity $ds^2=\eta_{\mu\nu}dx^\mu dx^\nu$ and the components $g_{\mu\nu}$ are the generalization of $\eta_{\mu\nu}$.

- There exists a (non-unique) orthonormal basis $\{X_\mu\}$ of V_p in which

$$g(X_\mu, X_\nu) = 0 \ , \ \text{if } \mu \neq \nu \text{ and } g(X_\mu, X_\nu) = \pm 1 \ , \ \text{if } \mu = \nu. \tag{A.50}$$

  The number of plus and minus signs is called the signature of the metric and is independent of choice of basis. In fact the number of plus signs and the number of minus signs are separately independent of choice of basis.

  A manifold with a metric which is positive definite is called Euclidean or Riemannian whereas a manifold with a metric which is indefinite is called Lorentzian or Pseudo-Riemannian. Spacetime in special and general relativity is a Lorentzian manifold.

- The map $g(.,v)$ can be thought of as an element of V_p^*. Thus the metric can be thought of as a map from V_p into V_p^* given by $v \longrightarrow g(.,v)$. Because of the nondegeneracy of g, the map $v \longrightarrow g(.,v)$ is one-to-one and onto and as a consequence it is invertible. The metric provides thus an isomorphism between V_p and V_p^*.

- The nondegeneracy of g can also be expressed by the statement that the determinant $g = \det(g_{\mu\nu}) \neq 0$. The components of the inverse metric will be denoted by $g^{\mu\nu} = g^{\nu\mu}$ and thus

$$g^{\mu\rho} g_{\rho\nu} = \delta^\mu_\nu \ , \ g_{\mu\rho} g^{\rho\nu} = \delta^\nu_\mu. \tag{A.51}$$

  The metric $g_{\mu\nu}$ and its inverse $g^{\mu\nu}$ can be used to raise and lower indices on tensors as in special relativity.

A.2 Curvature

A.2.1 Covariant Derivative

Definition 10**:** A covariant derivative operator ∇ on a manifold M is a map which takes a differentiable tensor of type (k,l) to a differentiable tensor of type $(k,l+1)$ which satisfies the following properties:

- Linearity:

$$\nabla(\alpha T + \beta S) = \alpha \nabla T + \beta \nabla S \ , \ \alpha, \beta \in R \ , \ T, S \in \mathcal{T}(k,l). \tag{A.52}$$

- Leibniz rule:

$$\nabla(T \otimes S) = \nabla T \otimes S + T \otimes \nabla S \ , \ T \in \mathcal{T}(k,l) \ , \ S \in \mathcal{T}(k^{'},l^{'}). \tag{A.53}$$

- Commutativity with contraction: In the so-called index notation a tensor $T \in \mathcal{T}(k,l)$ will be denoted by $T^{a_1 \ldots a_k}{}_{b_1 \ldots b_l}$ while the tensor $\nabla T \in \mathcal{T}(k,l+1)$ will be denoted by

$\nabla_c T^{a_1...a_k}{}_{b_1...b_l}$. The almost obvious requirement of commutativity with contraction means that for all $T \in \mathcal{T}(k,l)$ we must have

$$\nabla_d(T^{a_1...c...a_k}{}_{b_1...c...b_l}) = \nabla_d T^{a_1...c...a_k}{}_{b_1...c...b_l}. \tag{A.54}$$

- The covariant derivative acting on scalars must be consistent with tangent vectors being directional derivatives. Indeed for all $f \in \mathcal{F}$ and $t^a \in V_p$ we must have

$$t^a \nabla_a f = t(f). \tag{A.55}$$

- Torsion free: For all $f \in \mathcal{F}$ we have

$$\nabla_a \nabla_b f = \nabla_b \nabla_a f. \tag{A.56}$$

Ordinary Derivative: Let $\{\partial/\partial x^\mu\}$ and $\{dx^\mu\}$ be the coordinate bases of the tangent vector space and the cotangent vector space respectively in some coordinate system ψ. An ordinary derivative operator ∂ can be defined in the region covered by the coordinate system ψ as follows. If $T^{\mu_1...\mu_k}{}_{\nu_1...\nu_l}$ are the components of the tensor $T^{a_1...a_k}{}_{b_1...b_l}$ in the coordinate system ψ, then $\partial_\sigma T^{\mu_1...\mu_k}{}_{\nu_1...\nu_l}$ are the components of the tensor $\partial_c T^{a_1...a_k}{}_{b_1...b_l}$in the coordinate system ψ. The ordinary derivative operator ∂ satisfies all the above five requirements as a consequence of the properties of partial derivatives. However it is quite clear that the ordinary derivative operator ∂ is coordinate dependent.

Action of Covariant Derivative on Tensors: Let ∇ and $\tilde{\nabla}$ be two covariant derivative operators. By condition 4 of definition 10 their action on scalar functions must coincide, viz

$$t^a \nabla_a f = t^a \tilde{\nabla}_a f = t(f). \tag{A.57}$$

We compute now the difference $\tilde{\nabla}_a(f\omega_b) - \nabla_a(f\omega_b)$ where ω is some cotangent dual vector. We have

$$\begin{aligned}\tilde{\nabla}_a(f\omega_b) - \nabla_a(f\omega_b) &= \tilde{\nabla}_a f.\omega_b + f\tilde{\nabla}_a\omega_b - \nabla_a f.\omega_b - f\nabla_a\omega_b \\ &= f(\tilde{\nabla}_a\omega_b - \nabla_a\omega_b).\end{aligned} \tag{A.58}$$

The difference $\tilde{\nabla}_a\omega_b - \nabla_a\omega_b$ depends only on the value of ω_b at the point p although both $\tilde{\nabla}_a\omega_b$ and $\nabla_a\omega_b$ depend on how ω_b changes as we go away from the point p since they are derivatives. The proof goes as follows. Let $\omega_b^{'}$ be the value of the cotangent dual vector ω_b at a nearby point $p^{'}$, i.e. $\omega_b^{'} - \omega_b$ is zero at p. Thus by equation (A.12) there must exist smooth functions $f_{(\alpha)}$ which vanish at the point p and cotangent dual vectors $\mu_b^{(\alpha)}$ such that

$$\omega_b^{'} - \omega_b = \sum_\alpha f_{(\alpha)}\mu_b^{(\alpha)}. \tag{A.59}$$

We compute immediately

$$\tilde{\nabla}(\omega_b^{'}-\omega_b)-\nabla(\omega_b^{'}-\omega_b)=\sum_\alpha f_{(\alpha)}(\tilde{\nabla}_a\mu_b^{(\alpha)}-\nabla_a\mu_b^{(\alpha)}). \tag{A.60}$$

This is 0 since by assumption $f_{(\alpha)}$ vanishes at p. Hence we get the desired result

$$\tilde{\nabla}_a\omega_b^{'}-\nabla_a\omega_b^{'}=\tilde{\nabla}_a\omega_b-\nabla_a\omega_b. \tag{A.61}$$

In other words $\tilde{\nabla}_a\omega_b-\nabla_a\omega_b$ depends only on the value of ω_b at the point p. Putting this differently we say that the operator $\tilde{\nabla}_a-\nabla_a$ is a map which takes cotangent dual vectors at a point p into tensors of type $(0,2)$ at p (not tensor fields defined in a neighborhood of p) which is clearly a linear map by condition 1 of definition 10. We write

$$\nabla_a\omega_b=\tilde{\nabla}_a\omega_b-C^c\ {}_{ab}\omega_c. \tag{A.62}$$

The tensor $C^c\ {}_{ab}$ stands for the map $\tilde{\nabla}_a-\nabla_a$ and it is clearly a tensor of type $(1,2)$. By setting $\omega_a=\nabla_af=\tilde{\nabla}_af$ we get

$$\nabla_a\nabla_bf=\tilde{\nabla}_a\tilde{\nabla}_b-C^c\ {}_{ab}\nabla_cf. \tag{A.63}$$

By employing now condition 5 of definition 10 we get immediately

$$C^c\ {}_{ab}=C^c\ {}_{ba}. \tag{A.64}$$

Let us consider now the difference $\tilde{\nabla}_a(\omega_bt^b)-\nabla_a(\omega_bt^b)$ where t^b is a tangent vector. Since ω_bt^b is a function we have

$$\tilde{\nabla}_a(\omega_bt^b)-\nabla_a(\omega_bt^b)=0. \tag{A.65}$$

From the other hand we compute

$$\tilde{\nabla}_a(\omega_bt^b)-\nabla_a(\omega_bt^b)=\omega_b(\tilde{\nabla}_at^b-\nabla_at^b+C^b\ {}_{ac}t^c). \tag{A.66}$$

Hence we must have

$$\nabla_at^b=\tilde{\nabla}_at^b+C^b\ {}_{ac}t^c. \tag{A.67}$$

For a general tensor $T^{b_1...b_k}\ {}_{c_1...c_l}$ of type (k,l) the action of the covariant derivative operator will be given by the expression

$$\nabla_aT^{b_1...b_k}\ {}_{c_1...c_l}=\tilde{\nabla}_aT^{b_1...b_k}\ {}_{c_1...c_l}+\sum_iC^{b_i}\ {}_{ad}T^{b_1...d...b_k}\ {}_{c_1...c_l}-\sum_jC^d\ {}_{ac_j}T^{b_1...b_k}\ {}_{c_1...d...c_l}. \tag{A.68}$$

The most important case corresponds to the choice $\tilde{\nabla}_a=\partial_a$. In this case $C^c\ {}_{ab}$ is denoted $\Gamma^c\ {}_{ab}$ and is called Christoffel symbol. This is a tensor associated with the covariant derivative operator ∇_a and the coordinate system ψ in which the ordinary partial derivative ∂_a is defined. By passing to a different coordinate system $\psi^{'}$ the ordinary partial derivative changes from ∂_a to $\partial_a^{'}$ and hence the Christoffel symbol changes from $\Gamma^c\ {}_{ab}$ to $\Gamma^{'c}\ {}_{ab}$. The components of $\Gamma^c\ {}_{ab}$ in the coordinate system ψ will not be related to the components of $\Gamma^{'c}\ {}_{ab}$ in the coordinate system $\psi^{'}$ by the tensor transformation law since both the coordinate system and the tensor have changed.

A.2.2 Parallel Transport

Definition 11: Let C be a curve with a tangent vector t^a. Let v^a be some tangent vector defined at each point on the curve. The vector v^a is parallelly transported along the curve C if and only if

$$t^a \nabla_a v^b|_{\rm curve} = 0. \tag{A.69}$$

We have the following consequences and remarks:

- We know that

$$\nabla_a v^b = \partial_a v^b + \Gamma^b{}_{ac} v^c. \tag{A.70}$$

  Thus

$$t^a(\partial_a v^b + \Gamma^b{}_{ac} v^c) = 0. \tag{A.71}$$

  Let t be the parameter along the curve C. The components of the vector t^a in a coordinate basis are given by

$$t^\mu = \frac{dx^\mu}{dt}. \tag{A.72}$$

  In other words

$$\frac{dv^\nu}{dt} + \Gamma^\nu{}_{\mu\lambda} t^\mu v^\lambda = 0. \tag{A.73}$$

  From the properties of ordinary differential equations we know that this last equation has a unique solution. In other words we can map tangent vector spaces V_p and V_q at points p and q of the manifold if we are given a curve C connecting p and q and a derivative operator. The corresponding mathematical structure is called connection. In some usage the derivative operator itself is called a connection.

- By demanding that the inner product of two vectors v^a and w^a is invariant under parallel transport we obtain the condition

$$t^a \nabla_a (g_{bc} v^b w^c) = 0 \Rightarrow t^a \nabla_a g_{bc}.v^b w^c + g_{bc} w^c.t^a \nabla_a v^b + g_{bc} v^b.t^a \nabla_a w^c = 0. \tag{A.74}$$

  By using the fact that v^a and w^a are parallelly transported along the curve C we obtain the condition

$$t^a \nabla_a g_{bc}.v^b w^c = 0. \tag{A.75}$$

  This condition holds for all curves and all vectors and thus we get

$$\nabla_a g_{bc} = 0. \tag{A.76}$$

  Thus given a metric g_{ab} on a manifold M the most natural covariant derivative operator is the one under which the metric is covariantly constant.

- It is a theorem that given a metric g_{ab} on a manifold M, there exists a unique covariant derivative operator ∇_a which satisfies $\nabla_a g_{bc}=0$. The proof goes as follows. We know that $\nabla_a g_{bc}$ is given by

$$\nabla_a g_{bc}=\tilde{\nabla}_a g_{bc}-C^d{}_{ab}g_{dc}-C^d{}_{ac}g_{bd}. \tag{A.77}$$

By imposing $\nabla_a g_{bc}=0$ we get

$$\tilde{\nabla}_a g_{bc}=C^d{}_{ab}g_{dc}+C^d{}_{ac}g_{bd}. \tag{A.78}$$

Equivalently

$$\tilde{\nabla}_b g_{ac}=C^d{}_{ab}g_{dc}+C^d{}_{bc}g_{ad}. \tag{A.79}$$

$$\tilde{\nabla}_c g_{ab}=C^d{}_{ac}g_{db}+C^d{}_{bc}g_{ad}. \tag{A.80}$$

Immediately we conclude that

$$\tilde{\nabla}_a g_{bc}+\tilde{\nabla}_b g_{ac}-\tilde{\nabla}_c g_{ab}=2C^d{}_{ab}g_{dc}. \tag{A.81}$$

In other words

$$C^d{}_{ab}=\frac{1}{2}g^{dc}(\tilde{\nabla}_a g_{bc}+\tilde{\nabla}_b g_{ac}-\tilde{\nabla}_c g_{ab}). \tag{A.82}$$

This choice of $C^d{}_{ab}$ which solves $\nabla_a g_{bc}=0$ is unique. In other words the corresponding covariant derivative operator is unique.

- Generally a tensor $T^{b_1...b_k}{}_{c_1...c_l}$ is parallelly transported along the curve C if and only if

$$t^a\nabla_a T^{b_1...b_k}{}_{c_1...c_l}|_{\rm curve}=0. \tag{A.83}$$

A.2.3 The Riemann Curvature

Riemann Curvature Tensor: The so-called Riemann curvature tensor can be defined in terms of the failure of successive operations of differentiation to commute. Let us start with an arbitrary tangent dual vector ω_a and an arbitrary function f. We want to calculate $(\nabla_a\nabla_b-\nabla_b\nabla_a)\omega_c$. First we have

$$\nabla_a\nabla_b(f\omega_c)=\nabla_a\nabla_b f.\omega_c+\nabla_b f\nabla_a\omega_c+\nabla_a f\nabla_b\omega_c+f\nabla_a\nabla_b\omega_c. \tag{A.84}$$

Similarly

$$\nabla_b\nabla_a(f\omega_c)=\nabla_b\nabla_a f.\omega_c+\nabla_a f\nabla_b\omega_c+\nabla_b f\nabla_a\omega_c+f\nabla_b\nabla_a\omega_c. \tag{A.85}$$

Thus

$$(\nabla_a\nabla_b-\nabla_b\nabla_a)(f\omega_c)=f(\nabla_a\nabla_b-\nabla_b\nabla_a)\omega_c. \tag{A.86}$$

We can follow the same set of arguments which led from (A.58) to (A.62) to conclude that the tensor $(\nabla_a\nabla_b-\nabla_b\nabla_a)\omega_c$ depends only on the value of ω_c at the point p. In other words $\nabla_a\nabla_b-\nabla_b\nabla_a$ is a linear map which takes tangent dual vectors into tensors of type $(0,3)$. Equivalently we can say that the action of $\nabla_a\nabla_b-\nabla_b\nabla_a$ on tangent dual vectors is equivalent to the action of a tensor of type $(1,3)$. Thus we can write

$$(\nabla_a\nabla_b-\nabla_b\nabla_a)\omega_c=R_{abc}{}^{d}\omega_d. \tag{A.87}$$

The tensor $R_{abc}{}^{d}$ is precisely the Riemann curvature tensor.

Action on Tangent Vectors: Let now t^a be an arbitrary tangent vector. The scalar product $t^a\omega_a$ is a function on the manifold and thus

$$(\nabla_a\nabla_b-\nabla_b\nabla_a)(t^c\omega_c)=0. \tag{A.88}$$

But

$$(\nabla_a\nabla_b-\nabla_b\nabla_a)(t^c\omega_c)=(\nabla_a\nabla_b-\nabla_b\nabla_a)t^c.\omega_c+t^c.(\nabla_a\nabla_b-\nabla_b\nabla_a)\omega_c. \tag{A.89}$$

In other words

$$(\nabla_a\nabla_b-\nabla_b\nabla_a)t^d=-R_{abc}{}^{d}t^c \tag{A.90}$$

Generalization of this result and the previous one to higher tensors is given by

$$(\nabla_a\nabla_b-\nabla_b\nabla_a)T^{d_1\ldots d_k}{}_{c_1\ldots c_l}=-\sum_{i=1}^{k}R_{abe}{}^{d_i}T^{d_1\ldots e\ldots d_k}{}_{c_1\ldots c_l}+\sum_{i=1}^{l}R_{abc_i}{}^{e}T^{d_1\ldots d_k}{}_{c_1\ldots e\ldots c_l}. \tag{A.91}$$

Properties of the Curvature Tensor: We state without proof the following properties of the curvature tensor [1]:

- Anti-symmetry in the first two indices:

$$R_{abc}{}^{d}=-R_{bac}{}^{d}. \tag{A.92}$$

- Anti-symmetrization of the first three indices yields 0:

$$R_{[abc]}{}^{d}=0\ ,\ R_{[abc]}{}^{d}=\frac{1}{3}(R_{abc}{}^{d}+R_{cab}{}^{d}+R_{bca}{}^{d}). \tag{A.93}$$

- Anti-symmetry in the last two indices:

$$R_{abcd}=-R_{abdc}\ ,\ R_{abcd}=R_{abc}{}^{e}g_{ed}. \tag{A.94}$$

[1] Exercise: Verify these properties explicitly.

- Symmetry if the pair consisting of the first two indices is exchanged with the pair consisting of the last two indices:

$$R_{abcd}=R_{cdab}. \tag{A.95}$$

- Bianchi identity:

$$\nabla_{[a}R_{bc]d}{}^{e}=0\ ,\ \nabla_{[a}R_{bc]d}{}^{e}=\frac{1}{3}(\nabla_aR_{bcd}{}^{e}+\nabla_cR_{abd}{}^{e}+\nabla_bR_{cad}{}^{e}). \tag{A.96}$$

Ricci and Einstein Tensors: The Ricci tensor is defined by

$$R_{ac}=R_{abc}{}^{b}. \tag{A.97}$$

It is not difficult to show that $R_{ac}=R_{ca}$. This is the trace part of the Riemann curvature tensor. The so-called scalar curvature is defined by

$$R=R_a{}^{a}. \tag{A.98}$$

By contracting the Bianchi identity and using $\nabla_ag_{bc}=0$ we get

$$g_e{}^{c}(\nabla_aR_{bcd}{}^{e}+\nabla_cR_{abd}{}^{e}+\nabla_bR_{cad}{}^{e})=0\Rightarrow\nabla_aR_{bd}+\nabla_eR_{abd}{}^{e}-\nabla_bR_{ad}=0. \tag{A.99}$$

By contracting now the two indices b and d we get

$$g^{bd}(\nabla_aR_{bd}+\nabla_eR_{abd}{}^{e}-\nabla_bR_{ad})=0\Rightarrow\nabla_aR-2\nabla_bR_a{}^{b}=0. \tag{A.100}$$

This can be put in the form

$$\nabla^aG_{ab}=0. \tag{A.101}$$

The tensor G_{ab} is called Einstein tensor and is given by

$$G_{ab}=R_{ab}-\frac{1}{2}g_{ab}R. \tag{A.102}$$

Geometrical Meaning of the Curvature: The parallel transport of a vector from point p to point q is actually path-dependent. This path-dependence is directly measured by the curvature tensor as we will now show.

We consider a tangent vector v^a and a tangent dual vector ω_a at a point p of a manifold M. We also consider a curve C consisting of a small closed loop on a two-dimensional surface S parameterized by two real numbers s and t with the point p at the origin, viz $(t,s)|_p=(0,0)$. The first leg of this closed loop extends from p to the point $(\Delta t,0)$, the second leg extends from $(\Delta t,0)$ to $(\Delta t,\Delta s)$, the third leg extends from $(\Delta t,\Delta s)$ to $(0,\Delta s)$ and the last leg from $(0,\Delta s)$ to the point p. We parallel transport the vector v^a but not the tangent dual vector ω_a around this loop.

We form the scalar product $\omega_a v^a$ and compute how it changes under the above parallel transport. Along the first stretch between $p=(0,0)$ and $(\Delta t,0)$ we have the change

$$\delta_1 = \Delta t \frac{\partial}{\partial t}(v^a\omega_a)|_{(\Delta t/2,0)}. \tag{A.103}$$

This is obviously accurate upto correction of the order Δt^3. Let T^a be the tangent vector to the line segment connecting $p=(0,0)$ and $(\Delta t,0)$. It is clear that T^a is also the tangent vector to all the curves of constant s. The above change can then be rewritten as

$$\delta_1 = \Delta t T^b\nabla_b(v^a\omega_a)|_{(\Delta t/2,0)}. \tag{A.104}$$

Since v^a is parallelly transported we have $T^b\nabla_b v^a = 0$. We have then

$$\delta_1 = \Delta t v^a T^b\nabla_b\omega_a|_{(\Delta t/2,0)}. \tag{A.105}$$

The variation δ_3 corresponding to the third line segment between $(\Delta t,\Delta s)$ and $(0,\Delta s)$ must be given by

$$\delta_3 = -\Delta t v^a T^b\nabla_b\omega_a|_{(\Delta t/2,\Delta s)}. \tag{A.106}$$

We have then

$$\delta_1+\delta_3 = \Delta t\bigg[v^a T^b\nabla_b\omega_a|_{(\Delta t/2,0)} - v^a T^b\nabla_b\omega_a|_{(\Delta t/2,\Delta s)}\bigg]. \tag{A.107}$$

This is clearly 0 when $\Delta s\longrightarrow 0$ and as a consequence parallel transport is path-independent at first order. The vector v^a at $(\Delta t/2,\Delta s)$ can be thought of as the parallel transport of the vector v^a at $(\Delta t/2,0)$ along the curve connecting these two points, i.e. the line segment connecting $(\Delta t/2,0)$ and $(\Delta t/2,\Delta s)$. By the previous remark parallel transport is path-independent at first order which means that v^a at $(\Delta t/2,\Delta s)$ is equal to v^a at $(\Delta t/2,0)$ upto corrections of the order of Δs^2, Δt^2 and $\Delta s\Delta t$. Thus

$$\delta_1+\delta_3 = \Delta t v^a\bigg[T^b\nabla_b\omega_a|_{(\Delta t/2,0)} - T^b\nabla_b\omega_a|_{(\Delta t/2,\Delta s)}\bigg]. \tag{A.108}$$

Similarly $T^b\nabla_b\omega_a$ at $(\Delta t/2,\Delta s)$ is the parallel transport of $T^b\nabla_b\omega_a$ at $(\Delta t/2,0)$ and hence upto first order we must have

$$T^b\nabla_b\omega_a|_{(\Delta t/2,0)} - T^b\nabla_b\omega_a|_{(\Delta t/2,\Delta s)} = -\Delta s S^c\nabla_c(T^b\nabla_b\omega_a). \tag{A.109}$$

The vector S^a is the tangent vector to the line segment connecting $(\Delta t/2,0)$ and $(\Delta t/2,\Delta s)$ which is the same as the tangent vector to all the curves of constant t. Hence

$$\delta_1+\delta_3 = -\Delta t\Delta s v^a S^c\nabla_c(T^b\nabla_b\omega_a). \tag{A.110}$$

The final result is therefore

$$\begin{aligned}\delta(v^a\omega_a) &= \delta_1+\delta_3+\delta_2+\delta_4\\&= \Delta t\Delta s v^a[T^c\nabla_c(S^b\nabla_b\omega_a)-S^c\nabla_c(T^b\nabla_b\omega_a)]\\&= \Delta t\Delta s v^a[(T^c\nabla_c S^b-S^c\nabla_c T^b)\nabla_b\omega_a+T^cS^b(\nabla_c\nabla_b-\nabla_b\nabla_c)\omega_a]\\&= \Delta t\Delta s v^a T^cS^bR_{cba}{}^d\omega_d.\end{aligned} \tag{A.111}$$

In the third line we have used the fact that S^a and T^a commute. Indeed the commutator of the vectors T^a and S^a is given by the vector $[T,S]^a$ where $[T,S]^a=T^c\nabla_cS^a-S^c\nabla_cT^a$. This must vanish since T^a and S^a are tangent vectors to linearly independent curves. Since ω_a is not parallelly transported we have $\delta(v^a\omega_a)=\delta v^a.\omega_a$ and thus one can finally conclude that

$$\delta v^d = \Delta t\Delta s v^aT^cS^bR_{cba}{}^d. \tag{A.112}$$

The Riemann curvature tensor measures therefore the path-dependence of parallelly transported vectors.

Components of the Curvature Tensor: We know that

$$(\nabla_a\nabla_b-\nabla_b\nabla_a)\omega_c=R_{abc}{}^d\omega_d. \tag{A.113}$$

We know also

$$\nabla_a\omega_b=\partial_a\omega_b-\Gamma^c{}_{ab}\omega_c. \tag{A.114}$$

We compute then

$$\begin{aligned}\nabla_a\nabla_b\omega_c &= \nabla_a(\partial_b\omega_c-\Gamma^d{}_{bc}\omega_d)\\&= \partial_a(\partial_b\omega_c-\Gamma^d{}_{bc}\omega_d)-\Gamma^e{}_{ab}(\partial_e\omega_c-\Gamma^d{}_{ec}\omega_d)-\Gamma^e{}_{ac}(\partial_b\omega_e-\Gamma^d{}_{be}\omega_d)\\&= \partial_a\partial_b\omega_c-\partial_a\Gamma^d{}_{bc}.\omega_d-\Gamma^d{}_{bc}\partial_a\omega_d-\Gamma^e{}_{ab}\partial_e\omega_c+\Gamma^e{}_{ab}\Gamma^d{}_{ec}\omega_d-\Gamma^e{}_{ac}\partial_b\omega_e+\Gamma^e{}_{ac}\Gamma^d{}_{be}\omega_d.\end{aligned} \tag{A.115}$$

And

$$(\nabla_a\nabla_b-\nabla_b\nabla_a)\omega_c = \bigg(\partial_b\Gamma^d{}_{ac}-\partial_a\Gamma^d{}_{bc}+\Gamma^e{}_{ac}\Gamma^d{}_{be}-\Gamma^a{}_{bc}\Gamma^d{}_{ae}\bigg)\omega_d. \tag{A.116}$$

We get then the components

$$R_{abc}{}^d = \partial_b\Gamma^d{}_{ac}-\partial_a\Gamma^d{}_{bc}+\Gamma^e{}_{ac}\Gamma^d{}_{be}-\Gamma^e{}_{bc}\Gamma^d{}_{ae}. \tag{A.117}$$

A.2.4 Geodesics

Parallel Transport of a Curve along Itself: Geodesics are the straightest possible lines on a curved manifold. Let us recall that a tangent vector v^a is parallelly transported along a curve C with a tangent vector T^a if and only if $T^a\nabla_a v^b = 0$. A geodesics is a curve whose tangent vector T^a is parallelly transported along itself, viz

$$T^a\nabla_a T^b = 0. \tag{A.118}$$

This reads in a coordinate basis as

$$\frac{dT^\nu}{dt} + \Gamma^\nu\ {}_{\mu\lambda}T^\mu T^\lambda = 0. \tag{A.119}$$

In a given chart ϕ the curve C is mapped into a curve $x(t)$ in R^n. The components T^μ are given in terms of $x^\mu(t)$ by

$$T^\mu = \frac{dx^\mu}{dt}. \tag{A.120}$$

Hence

$$\frac{d^2x^\nu}{dt^2} + \Gamma^\nu\ {}_{\mu\lambda}\frac{dx^\mu}{dt}\frac{dx^\lambda}{dt} = 0. \tag{A.121}$$

This is a set of n coupled second order ordinary differential equations with n unknown $x^\mu(t)$. Given appropriate initial conditions $x^\mu(t_0)$ and $dx^\mu/dt|_{t=t_0}$ we know that there must exist a unique solution. Conversely given a tangent vector T_p at a point p of a manifold M there exists a unique geodesics which goes through p and is tangent to T_p.

Length of a Curve: The length l of a smooth curve C with tangent T^a on a manifold M with Riemannian metric g_{ab} is obviously given by

$$l = \int dt\sqrt{g_{ab}T^aT^b}. \tag{A.122}$$

The length is parametrization independent. Indeed we can show that [2]

$$l = \int dt\sqrt{g_{ab}T^aT^b} = \int ds\sqrt{g_{ab}S^aS^b}\ ,\ S^a = T^a\frac{dt}{ds}. \tag{A.123}$$

In a Lorentzian manifold, the length of a spacelike curve is also given by this expression. For a timelike curve for which $g_{ab}T^aT^b < 0$ the length is replaced with the proper time τ which is given by $c\tau = \int dt\sqrt{-g_{ab}T^aT^b}$. For a lightlike (or null) curve for which $g_{ab}T^aT^b = 0$ the length is always 0. Geodesics in a Lorentzian manifold can not change from timelike to spacelike or null and vice versa since the norm is conserved in a parallel transport. The length of a curve which changes from spacelike to timelike or vice versa is not defined.

[2]Exercise: Verify this equation explicitly.

Geodesics extremize the length as we will now show. We consider the length of a curve C connecting two points $p = C(t_0)$ and $q = C(t_1)$. In a coordinate basis the length is given explicitly by

$$l = \int_{t_0}^{t_1} dt \sqrt{g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}}. \tag{A.124}$$

The variation in l under an arbitrary smooth deformation of the curve C which keeps the two points p and q fixed is given by

$$\begin{aligned}
\delta l &= \frac{1}{2}\int_{t_0}^{t_1} dt\big(g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}\big)^{-\frac{1}{2}}\bigg(\frac{1}{2}\delta g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt} + g_{\mu\nu}\frac{dx^\mu}{dt}\frac{d\delta x^\nu}{dt}\bigg)\\
&= \frac{1}{2}\int_{t_0}^{t_1} dt\big(g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}\big)^{-\frac{1}{2}}\bigg(\frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\sigma}\delta x^\sigma\frac{dx^\mu}{dt}\frac{dx^\nu}{dt} + g_{\mu\nu}\frac{dx^\mu}{dt}\frac{d\delta x^\nu}{dt}\bigg)\\
&= \frac{1}{2}\int_{t_0}^{t_1} dt\big(g_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt}\big)^{-\frac{1}{2}}\bigg(\frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\sigma}\delta x^\sigma\frac{dx^\mu}{dt}\frac{dx^\nu}{dt} - \frac{d}{dt}(g_{\mu\nu}\frac{dx^\mu}{dt})\delta x^\nu + \frac{d}{dt}(g_{\mu\nu}\frac{dx^\mu}{dt}\delta x^\nu)\bigg).
\end{aligned} \tag{A.125}$$

We can assume without any loss of generality that the parametrization of the curve C satisfies $g_{\mu\nu}(dx^\mu/dt)(dx^\nu/dt) = 1$. In other words choose dt^2 to be precisely the line element (interval) and thus $T^\mu = dx^\mu/dt$ is the 4−velocity. The last term in the above equation becomes obviously a total derivative which vanishes by the fact that the considered deformation keeps the two end points p and q fixed. We get then

$$\begin{aligned}
\delta l &= \frac{1}{2}\int_{t_0}^{t_1} dt\delta x^\sigma\bigg(\frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\sigma}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt} - \frac{d}{dt}(g_{\mu\sigma}\frac{dx^\mu}{dt})\bigg)\\
&= \frac{1}{2}\int_{t_0}^{t_1} dt\delta x^\sigma\bigg(\frac{1}{2}\frac{\partial g_{\mu\nu}}{\partial x^\sigma}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt} - \frac{\partial g_{\mu\sigma}}{\partial x^\nu}\frac{dx^\nu}{dt}\frac{dx^\mu}{dt} - g_{\mu\sigma}\frac{d^2x^\mu}{dt^2}\bigg)\\
&= \frac{1}{2}\int_{t_0}^{t_1} dt\delta x^\sigma\bigg(\frac{1}{2}(\frac{\partial g_{\mu\nu}}{\partial x^\sigma} - \frac{\partial g_{\mu\sigma}}{\partial x^\nu} - \frac{\partial g_{\nu\sigma}}{\partial x^\mu})\frac{dx^\mu}{dt}\frac{dx^\nu}{dt} - g_{\mu\sigma}\frac{d^2x^\mu}{dt^2}\bigg)\\
&= \frac{1}{2}\int_{t_0}^{t_1} dt\delta x_\rho\bigg(\frac{1}{2}g^{\rho\sigma}(\frac{\partial g_{\mu\nu}}{\partial x^\sigma} - \frac{\partial g_{\mu\sigma}}{\partial x^\nu} - \frac{\partial g_{\nu\sigma}}{\partial x^\mu})\frac{dx^\mu}{dt}\frac{dx^\nu}{dt} - \frac{d^2x^\rho}{dt^2}\bigg)\\
&= \frac{1}{2}\int_{t_0}^{t_1} dt\delta x_\rho\bigg(-\Gamma^\rho{}_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt} - \frac{d^2x^\rho}{dt^2}\bigg).
\end{aligned} \tag{A.126}$$

The curve C extremizes the length between the two points p and a if and only if $\delta l = 0$. This leads immediately to the equation

$$\Gamma^\rho{}_{\mu\nu}\frac{dx^\mu}{dt}\frac{dx^\nu}{dt} + \frac{d^2x^\rho}{dt^2} = 0. \tag{A.127}$$

In other words the curve C must be a geodesic. Since the length between any two points on a Riemannian manifold (and between any two points which can be connected by a spacelike curve on a Lorentzian manifold) can be arbitrarily long we conclude that the shortest curve connecting the two points must be a geodesic as it is an extremum of length. Hence the shortest curve is the straightest possible curve. The converse is not true. A geodesic connecting two points is not necessarily the shortest path.

The proper time between any two points which can be connected by a timelike curve on a Lorentzian manifold can be arbitrarily small and thus the curve with greatest proper time (if it exists) must be a timelike geodesic as it is an extremum of proper time. However, a timelike geodesic connecting two points is not necessarily the path with maximum proper time.

Lagrangian: It is not difficult to convince ourselves that the geodesic equation can also be derived as the Euler-Lagrange equation of motion corresponding to the Lagrangian

$$L = \frac{1}{2} g_{\mu\nu} \frac{dx^\mu}{dt} \frac{dx^\nu}{dt}. \tag{A.128}$$

In fact given the metric tensor $g_{\mu\nu}$ we can write explicitly the above Lagrangian and from the corresponding Euler-Lagrange equation of motion we can read off directly the Christoffel symbols $\Gamma^\rho{}_{\mu\nu}$.

Bibliography

[1] R. M. Wald, "General Relativity," Chicago, Usa: Univ. Pr. (1984) 491p.

[2] V. Mukhanov, "Physical foundations of cosmology," Cambridge, UK: Univ. Pr. (2005) 421 p.

[3] S. M. Carroll, "Spacetime and geometry: An introduction to general relativity," San Francisco, USA: Addison-Wesley (2004) 513 p.

[4] J. Hartle, "Gravity: An introduction to Einstein's general relativity," Pearson, 2014.

[5] N. D. Birrell and P. C. W. Davies, "Quantum Fields In Curved Space," Cambridge, Uk: Univ. Pr. (1982) 340p.

[6] S. Weinberg, "Cosmology," Oxford, UK: Oxford Univ. Pr. (2008) 593 p.

[7] A. R. Liddle, "An Introduction to cosmological inflation," astro-ph/9901124.

[8] D. Baumann, "TASI Lectures on Inflation," arXiv:0907.5424 [hep-th].

[9] S. Weinberg, "The Cosmological Constant Problem," Rev. Mod. Phys. **61**, 1 (1989).

[10] S. M. Carroll, "The Cosmological constant," Living Rev. Rel. **4**, 1 (2001) [astro-ph/0004075].

[11] S. M. Carroll, "Why is the universe accelerating?," eConf C **0307282**, TTH09 (2003) [AIP Conf. Proc. **743**, 16 (2005)] [astro-ph/0310342].

[12] S. Dodelson, "Modern cosmology," Amsterdam, Netherlands: Academic Pr. (2003) 440 p

[13] T. Jacobson, "Introduction to quantum fields in curved space-time and the Hawking effect," gr-qc/0308048.

[14] L. H. Ford, "Quantum field theory in curved space-time," In *Campos do Jordao 1997, Particles and fields* 345-388 [gr-qc/9707062].

[15] V. Mukhanov and S. Winitzki, "Introduction to quantum effects in gravity," Cambridge, UK: Cambridge Univ. Pr. (2007) 273 p

[16] J. Martin, "Inflationary cosmological perturbations of quantum-mechanical origin," Lect. Notes Phys. **669**, 199 (2005) [hep-th/0406011].

[17] E. Bertschinger, "Cosmological dynamics: Course 1," astro-ph/9503125.

[18] J. M. Maldacena, "Non-Gaussian features of primordial fluctuations in single field inflationary models," JHEP **0305**, 013 (2003) [astro-ph/0210603].

[19] C. Itzykson and J. B. Zuber, "Quantum Field Theory," New York, Usa: Mcgraw-hill (1980) 705 P.(International Series In Pure and Applied Physics)

[20] George Siopsis, "Quantum Field Theory I," siopsis@tennessee.edu.

[21] Anthony Challinor, "Part-III Cosmology Course:The Perturbed Universe," a.d.challinor@ast.cam.ac.uk.

[22] K.A. Milton, "The Casimir Effect: Physical Manifestations of Zero-Point Energy ," World Scientific (2001).

[23] Angus Prain, "Vacuum Energy in Expanding Spacetime and Superoscillation Induced Resonance," Master thesis (2008).

[24] H. -H. Zhang, K. -X. Feng, S. -W. Qiu, A. Zhao and X. -S. Li, "On analytic formulas of Feynman propagators in position space," Chin. Phys. C **34**, 1576 (2010) [arXiv:0811.1261 [math-ph]].

[25] L. H. Ford, "Gravitational Particle Creation and Inflation," Phys. Rev. D **35**, 2955 (1987).

[26] A. Melchiorri *et al.* [Boomerang Collaboration], "A measurement of omega from the North American test flight of BOOMERANG," Astrophys. J. **536**, L63 (2000) [astro-ph/9911445].

[27] W. H. Kinney, A. Melchiorri and A. Riotto, "New constraints on inflation from the cosmic microwave background," Phys. Rev. D **63**, 023505 (2001) [astro-ph/0007375].

[28] A. Riotto, "Inflation and the Theory of Cosmological Perturbations," ICTP lectures (2002).

[29] R. L. Jaffe, "The Casimir effect and the quantum vacuum," Phys. Rev. D **72**, 021301 (2005) [hep-th/0503158].

[30] A. Kempf, "Mode generating mechanism in inflation with cutoff," Phys. Rev. D **63**, 083514 (2001) [astro-ph/0009209].

[31] N. Graham, R. L. Jaffe, V. Khemani, M. Quandt, O. Schroeder and H. Weigel, "The Dirichlet Casimir problem," Nucl. Phys. B **677**, 379 (2004) [hep-th/0309130].

[32] K. A. Milton, "Calculating casimir energies in renormalizable quantum field theory," Phys. Rev. D **68**, 065020 (2003) [hep-th/0210081].

[33] K. A. Milton, "Local and Global Casimir Energies: Divergences, Renormalization, and the Coupling to Gravity," Lect.Notes.Phys.834:39-95,2011 [arXiv:1005.0031].

[34] T. S. Bunch and P. C. W. Davies, "Quantum Field Theory in de Sitter Space: Renormalization by Point Splitting," Proc. Roy. Soc. Lond. A **360**, 117 (1978).

[35] E. Mottola, "Particle Creation in de Sitter Space," Phys. Rev. D **31**, 754 (1985).

[36] B. Allen, "Vacuum States in de Sitter Space," Phys. Rev. D **32**, 3136 (1985).

[37] D. Anninos, "De Sitter Musings," Int. J. Mod. Phys. A **27**, 1230013 (2012) [arXiv:1205.3855 [hep-th]].

[38] J. Cortez, D. Martin-de Blas, G. A. Mena Marugan and J. Velhinho, "Massless scalar field in de Sitter spacetime: unitary quantum time evolution," Class. Quantum Grav. **30**, 075015 (2013) [arXiv:1301.4920 [gr-qc]].

[39] G. Perez-Nadal, A. Roura and E. Verdaguer, "Backreaction from weakly and strongly non-conformal fields in de Sitter spacetime," PoS QG **-PH** (2007) 034.

[40] S. A. Fulling, "Aspects Of Quantum Field Theory In Curved Space-time," London Math. Soc. Student Texts **17**, 1 (1989).

[41] Julien Lesgourgues, "Inflationary Cosmology," Lecture notes of a course presented in the framework of the 3ieme cycle de physique de Suisse romande (2006).

[42] T. Eguchi, P. B. Gilkey and A. J. Hanson, "Gravitation, Gauge Theories and Differential Geometry," Phys. Rept. **66**, 213 (1980).

[43] G.'t Hooft, "Introduction to General Relativity".

[44] G.'t Hooft, "Introduction to the Theory of Black Holes".

[45] C. Mueller, "The Hamiltonian Formulation of General Relativity."

[46] P. Horava, "Membranes at Quantum Criticality," JHEP **0903**, 020 (2009) [arXiv:0812.4287 [hep-th]].

[47] P. Horava, "Quantum Gravity at a Lifshitz Point," Phys. Rev. D **79**, 084008 (2009) [arXiv:0901.3775 [hep-th]].

[48] P. Horava, "Spectral Dimension of the Universe in Quantum Gravity at a Lifshitz Point," Phys. Rev. Lett. **102**, 161301 (2009) [arXiv:0902.3657 [hep-th]].

[49] S. Mukohyama, "Horava-Lifshitz Cosmology: A Review," Class. Quant. Grav. **27**, 223101 (2010) [arXiv:1007.5199 [hep-th]].

[50] M. Visser, “Status of Horava gravity: A personal perspective,” J. Phys. Conf. Ser. **314**, 012002 (2011) [arXiv:1103.5587 [hep-th]].

[51] S. Weinfurtner, T. P. Sotiriou and M. Visser, “Projectable Horava-Lifshitz gravity in a nutshell,” J. Phys. Conf. Ser. **222**, 012054 (2010) [arXiv:1002.0308 [gr-qc]].

[52] Luis Piresa, ”Hoava-Lifshitz Gravity: Hamiltonian Formulation and Connections with CDT,” Master Thesis, Utrecht University.

[53] J. Bellorin and A. Restuccia, “Consistency of the Hamiltonian formulation of the lowest-order effective action of the complete Horava theory,” Phys. Rev. D **84**, 104037 (2011) [arXiv:1106.5766 [hep-th]].

[54] D. Orlando and S. Reffert, “On the Renormalizability of Horava-Lifshitz-type Gravities,” Class. Quant. Grav. **26**, 155021 (2009) [arXiv:0905.0301 [hep-th]].

[55] D. Orlando and S. Reffert, “On the Perturbative Expansion around a Lifshitz Point,” Phys. Lett. B **683**, 62 (2010) [arXiv:0908.4429 [hep-th]].

[56] G. Giribet, D. L. Nacir and F. D. Mazzitelli, “Counterterms in semiclassical Horava-Lifshitz gravity,” JHEP **1009**, 009 (2010) [arXiv:1006.2870 [hep-th]].