INFLATION AND QUANTUM GEOMETRODYNAMICS IN SCALAR-TENSOR THEORIES

DISSERTATION

zur Erlangung des Doktorgrades
der Fakultät für Mathematik und Physik
der Albert-Ludwigs-Universität Freiburg

vorgelegt von

Matthijs van der Wild

2019

Dekan	Prof. Dr. Gregor Herten
Referent	Prof. Dr. Jochum van der Bij
Koreferent	Prof. Dr. Harald Ita
Tag der mündlichen Prüfung	23.10.2019

COLOPHON

The work described in this document was funded in part by the DFG, under grant number GRK 2044, and in part by Freiburg University, under the *Abschlussstipendium*.

This document was typeset using ConTEXt. The body text is set in the extended Baskervald font at 10 points. The mathematical text is set in the TEX Gyre Pagello Math font at 10 points. The diagrams were made using METAPOST.

The text on the next page quotes Terry Pratchett's *The Wyrd Sisters*.

This book was written using 100% recycled words.

ABSTRACT

The work described in this thesis is an investigation of scalar-tensor theories. Scalar-tensor theories generalise Einstein's theory of relativity by replacing the universal coupling constant of gravity, the Planck mass, with a dynamical scalar field. One motivation to consider scalar-tensor theories can be found within cosmology itself. Observations indicate that, during the very early universe, space underwent a period of accelerated expansion. This phase of expansion is called inflation. The most common method to implement inflation introduces a new scalar field to generate the required vacuum energy. It is not yet clear what the fundamental nature of this scalar field should be, and general relativity does not provide an answer to this question. Scalar-tensor theories, on the other hand, naturally explain the existence of such a field. During inflation it can be expected that tiny quantum gravitational corrections can be found in the imprint of the cosmic microwave background radiation. These corrections can be studied within the formalism of quantum geometrodynamics, which is a canonical, non-perturbative theory of quantum gravity. An important object in this formalism is the Wheeler-DeWitt equation, which governs the dynamics of the quantised degrees of freedom. Although the Wheeler-DeWitt equation describes a full theory of quantum gravity, it has a well-defined classical limit, and quantum gravitational corrections can be systematically derived from a semiclassical expansion.

The thesis starts with a review of the formalism of general relativity. It introduces a general class of scalar-tensor theories, and presents its properties. It reviews the formalism of inflation, and ends with the presentation of the formalism of inflation within the context of a general scalar-tensor theory.

Furthermore, it presents the consequences of a general scalar-tensor theory in the context of semiclassical quantum gravity. The basis for this investigation is the canonical approach to quantum gravity, called quantum geometrodynamics. This is a canonical and non-perturbative formulation of quantum gravity, in which the Hamiltonian of the theory of general relativity is quantised in an analogous way as in non-relativistic quantum mechanics. The formalism is generalised to a general class of scalar-tensor theories. It is demonstrated how quantum geometrodynamics can be used to estimate the form and magnitude of the first quantum gravitational corrections.

Then, the thesis considers the formalism of quantum geometrodynamics within an inflationary model. The corrected Schrödinger equation that arises in the semiclassical expansion of the Wheeler-DeWitt equation is related to inflationary observables. The form of these corrections is derived and the magnitude is estimated. It is found that the quantum gravitational corrections in the inflationary power spectrum are too small to be detected by current experiments, although they are in principle observable. The corrections are found to have a clear observational signature.

CONTENTS

PUBLICATIONS

This work is based in part on the following works:

- [1] C. F. Steinwachs and M. L. van der Wild,
  Quantum gravitational corrections from the Wheeler-DeWitt equation for scalar-tensor theories,
  Class. Quant. Grav. 35(13), 135010, 2018, eprint **arXiv:1712.08543** [gr-qc].
- [2] C. F. Steinwachs and M. L. van der Wild,
  Quantum gravitational corrections to the inflationary power spectrum for scalar-tensor theories,
  accepted for publication in *Class. Quant. Grav.*, eprint **arXiv:1904.12861** [gr-qc].

1 INTRODUCTION

1.1 THE THEORETICAL DESCRIPTION OF GRAVITY

In recent years experiments, such as those performed with the *Planck* satellite, have given a new wealth of data that can be used to piece together the structure of the cosmos. This has resulted in a theoretical description in the form of the *Lambda cold dark matter* (ΛCDM) model of cosmology which, ultimately, is based on Einstein's theory of general relativity.

As a description of gravity, the theory of general relativity is highly successful. It accurately describes gravitational time dilatation and redshift, planetary motion in the solar system, and the bending of light as it moves through space. General relativity has withstood the test of time up to the present, as it provides an accurate description of recent observations. Perhaps the most prominent examples are the experimental observation of gravitational waves and the shadow of the supermassive black hole in the centre of the galaxy M87. With all these data in favour of Einstein's theory one cannot help but wonder whether there is a deeper, more fundamental theory of which general relativity is a low-energy limit.

These considerations are partially based on observations that cannot be explained by general relativity alone. The observation of galaxies at the edge of the observable universe found that the universe is to a large extent the same in every direction. Since there is nothing that indicates to distinguish the solar system from any other location in spacetime, it is therefore reasonable to postulate that the universe is the same at each point in space. Although an isoptropic and homogeneous universe is expected from the cosmic Big Bang, this seems to be in direct contradiction with the principle of causality, and the theory of general relativity offers no explanation.

An extension of general relativity introduces a vacuum energy which drives a phase of accelerated expansion during the early universe. This vacuum energy is provided by the potential of a dynamical scalar field, the existence of which has to be postulated. Over time, the scalar field reaches the minimum of its potential, which leads to the decay of the vacuum energy into particles. Tiny density fluctuations in the scalar field give rise to the inhomogeneities that eventually form the large-scale structure that is observed in the sky. This, in a nutshell, describes the formalism of cosmic inflation.

Although inflation is able to explain the formation of large-scale structure and the homogeneity and isotropy of the universe, it offers no motivation for the fundamental nature of the scalar field. This is somewhat unsatisfying, as the scalar field provides a dynamical origin for inflation. However, it is possible to find models in which the existence of this field is explained naturally. One particular class of models replaces the universal coupling constant of gravity, the Planck mass, with a scalar field that is dynamical. The resulting theory is called a scalar-tensor theory. This seemingly simple modification of Einstein's theory of gravity has profound implications. The dynamical nature of the scalar field results in a coupling of matter to gravity that is no longer universal: particles in different regions of spacetime would, in general, experience a different gravitating force. Therefore, scalar-tensor theories are not entirely compatible with equivalence principle in the Einsteinian sense.

Scalar-tensor theories open up the possibility to unify the standard model of particle physics with the standard model of cosmology in a single effective field theory that bridges the gap between the electroweak symmetry breaking scale and the Planck scale. One particularly attractive model identifies the Higgs field of particle physics with the cosmological inflaton which, from a field theoretical perspective, has the feature of being minimal, as it does not require the introduction of additional fundamental scalar fields.

It is natural to wonder what other consequences could arise from an alternative formulation of gravity. Since scalar-tensor theories form a connection between gravity and particle physics, they could conceivably give rise to an enhancement of quantum gravitational effects. These effects are typically suppressed by inverse powers of the Planck mass M_{P} in quantised formulations of general relativity. In order to estimate the magnitude of these quantum gravitational effects it is convenient to express the Planck mass in terms of the reduced Planck constant $\hbar$, the speed of light c and Newton's gravitational constant G_{N}:

$$M_{\mathrm{P}} \equiv \sqrt{\frac{\hbar c}{8\pi G_{\mathrm{N}}}} \approx 2.4 \times 10^{18}\,\mathrm{GeV}\,c^{-2}.$$

This is an enormous number and, as a result, quantum gravitational effects are beyond the reach of the current generation of detectors. However, these small effects may become sizeable—or at least conceptually observable—within the formalism of scalar-tensor theories, where the Planck mass is replaced with a dynamical field.

A conceptually simple formulation of quantum gravity is based on the canonical formulation of general relativity. The canonical variables in this formulation are the components of the spatial metric, which are described by a Hamiltonian. This Hamiltonian leads to constraint equations that, when appropriately quantised, results in a non-perturbative theory of quantum gravity called quantum geometrodynamics. The central object in quantum geometrodynamics is the Wheeler-DeWitt equation, which describes the full quantum dynamics of the canonical degrees of freedom. As a result, the full Wheeler-DeWitt equation is tremendously complicated, and can only be solved in special cases. Nevertheless, useful quantum gravitational information can still be extracted from the equation, as it allows a systematic semiclassical expansion to be performed in regions where the wave function is approximately classical. The effects of quantum gravity arise from such a procedure in the same way as relativistic corrections would arise from a perturbative expansion of the Klein-Gordon equation.

A natural scenario in which quantum gravitational corrections can be related to observations is the aforementioned inflationary epoch of the early universe. Tiny quantum fluctuations have left their imprint in the anisotropies of the cosmic background radiation. It may be expected that quantum gravitational fluctuations may be part of these quantum fluctuations. They are expected to play a role in the early universe, as the energy density of the universe can be estimated to be within a few orders of magnitude of the Planck scale.

1.2 CONTENTS OF THIS THESIS

The discussion above sets the stage for what follows in the rest of this work. The goal of this dissertation is twofold. First and foremost, the formalism of quantum geometrodynamics will be generalised, so that it includes a very general class of scalar-tensor theories. Second of all, this formalism is applied to inflation in order to investigate the impact of the non-minimal coupling of scalar-tensor theories to the magnitude of the quantum gravitational corrections to the inflationary power spectra.

This work is structured as follows. In **chapter 2** a summary of the conventional formulation of general relativity is presented. The dynamical nature of spacetime imposes the condition that the theory is invariant under certain classes of transformations, known as *coordinate transformations*, *diffeomorphisms* or *gauge transformations*. The result is that the set of fields that is used to describe the spacetime carries with it certain redundancies, and as a result not all the degrees of freedom of the gravitational field are dynamical. In this chapter a formalism, called the Arnowitt-Deser-Misner (ADM) formalism, will be presented that makes this redundancy evident, and facilitates the identification of the dynamical degrees of freedom, as well as the constraints that arise from the non-dynamical fields.

In **chapter 3** a modification of the theory of general relativity is explored. This modification, called a scalar-tensor theory, replaces the universal coupling constant of gravity to matter with a general, non-trivial, coupling between the gravitational metric tensor and matter fields. These theories have different formulations, which are related to non-linear field redefinitions, which are presented in this chapter. Furthermore, the chapter presents an investigation of the consequences of a non-trivial coupling of gravity to matter for a selected number of cosmologically relevant theories.

In **chapter 4** a review of the application of general relativity to cosmology is presented, which forms the basis of the description of the period of cosmic inflation in the early universe. This formalism is extended such that it includes a general class of scalar-tensor theories. The chapter will conclude with a derivation of the inflationary parameters. The results of this chapter will serve as a comparison with the results derived in the later chapters.

In **chapter 5** the general consequences are investigated that the diffeomorphism invariance of gravitational theories, such as the theory of general relativity or scalar-tensor theories, places on the canonical formulation of these theories. This formulation culminates in the Wheeler-DeWitt equation, which is the basis for a non-perturbative theory of quantum gravity called quantum geometrodynamics. The chapter continues with an outline of how classical physics and quantum mechanics can be obtained from a semiclassical expansion of the full WDW equation. The chapter concludes with an explicit example in the formalism of general relativity with a scalar field.

The first main result of the thesis is described in **chapter 6**, where the Arnowitt-Deser-Misner formalism is used to derive the Hamiltonian for a general class of scalar-tensor theories. The formalism of quantum geometrodynamics is then used in order to derive the WDW equation this general model. It is then shown in detail how a systematic semiclassical expansion of the timeless WDW equation for scalar-tensor theories gives rise to the classical formalism. The semiclassical expansion is then continued, which leads to a Schrödinger equation for the quantised degrees of freedom. The inclusion of higher order terms in the semiclassical expansion then leads to quantum gravitational corrections to the Schrödinger equation. The structure of these corrections is compared with its counterpart in the semiclassical expansion in general relativity with a minimally coupled matter field, and will briefly be discussed in the context of Higgs inflation.

The second main result of the thesis is described in **chapter 7**, where the formalism of quantum geometrodynamics is applied to a very general class of inflationary scalar-tensor theories. This results in the WDW equation for a general class of inflationary scalar-tensor models. A systematic semiclassical expansion of this equation reproduces the results obtained in **chapter 4**. The inclusion of terms of higher order in the semiclassical expansion results in the quantum gravitational corrections to the evolution equations of the cosmic fluctuations. The solutions to these equations are compared with their counterparts in general relativity with a minimally coupled scalar field.

1.3 CONVENTIONS

The majority of this work is expressed in terms of *natural units*, in which quantities are expressed in terms of physical constants. These constants are normalised to unity. In particular, the reduced Planck's constant and the speed of light are normalised to the value $\hbar = c = 1$. The Planck mass M_P is then equal to $M_\mathrm{P} = 1/\sqrt{8\pi G_\mathrm{N}}$, where G_N is Newton's constant.

The equations of a geometric theory of gravitation allows a certain freedom in the parametrisation of the gravitational metric $g_{\mu\nu}$. In this work the convention will be that in locally inertial coordinates the metric can be brought into the form

$$g_{\mu\nu} = \mathrm{diag}(-1, 1, 1, 1).$$

The degree in which a given metric depends on space is determined by the Christoffel symbol Γ, the components of which are defined as

$$\Gamma^\rho_{\mu\nu} = \tfrac{1}{2} g^{\rho\lambda}(\partial_\mu g_{\lambda\nu} + \partial_\nu g_{\mu\lambda} - \partial_\lambda g_{\mu\nu}).$$

The curvature of spacetime is described by the Riemann tensor $R^\rho{}_{\sigma\mu\nu}$, which in this work is defined in terms of the Christoffel symbol as

$$R^\rho{}_{\sigma\mu\nu} = \partial_\mu \Gamma^\rho_{\sigma\nu} - \partial_\nu \Gamma^\rho_{\sigma\mu} + \Gamma^\rho_{\mu\lambda}\Gamma^\lambda_{\sigma\nu} - \Gamma^\rho_{\nu\lambda}\Gamma^\lambda_{\sigma\mu}.$$

The Ricci tensor $R_{\mu\nu}$ is defined by the contraction of the first index of the Riemann tensor with the third.

The symmetric part $T_{(ab)}$ and antisymmetric part $T_{[ab]}$ of a tensor T_{ab} will respectively be denoted as

$$T_{(ab)} \equiv \tfrac{1}{2}(T_{ab} + T_{ba}), \qquad T_{[ab]} \equiv \tfrac{1}{2}(T_{ab} - T_{ba}).$$

DYNAMICS OF SPACETIME

2

The most accurate description of gravitation in physics is formulated as a dynamical theory of spacetime. Various such theories have been developed during the past decades, but the conceptually simplest is Einstein's theory of *general relativity*. The theory remains successful up to the present, where in recent years general relativity was subjected to tests such as the direct detection of gravitational waves or the direct imaging of the shadow of a black hole [3,4].

This chapter starts with a brief summary of the necessary formalism for the covariant formulation of general relativity. The remaining part of the chapter is dedicated to the foliation of spacetime in terms of hypersurfaces. Such a foliation is necessary for a Hamiltonian formulation of gravitational theories, of which general relativity is the most natural example. A particular application of spacetime foliation is the *Arnowitt-Deser-Misner* (ADM) formalism [5], which will later find its application in the canonical quantisation procedure. This formulation and its subsequent canonical quantisation are described in **chapter 5**.

There is no unique way to foliate spacetime, and indeed it is a feature of the diffeomorphism invariance of general relativity that the choice of foliation is often dependent on the computational applications. In addition, the hypersurfaces can be timelike, spacelike or lightlike in character, owing to the Lorentzian structure of space and time. Emphasis in this work is placed on the former two.

The chapter will, after its presentation of the formulation of the gravitational action in terms of the ADM formalism, conclude with a brief discussion of the restrictions the symmetries of the universe place on the foliation of spacetime.

For more extended overviews of the formalism presented here the reader is invited to study the many excellent available reference works. Examples of such are found in the references [6–15].

2.1 COVARIANT FORMULATION OF GRAVITY

General relativity is described by the *Einstein-Hilbert action*

$$S_{\mathrm{EH}}[g] = \kappa \int_{\mathcal{M}} (R - 2\Lambda)\, \sqrt{-g}\, \mathrm{d}^4 x, \tag{2.1}$$

up to boundary terms that will be derived at the end of this chapter. Here, R is the Ricci tensor that is calculated from the metric tensor g, κ is half the square of the Planck mass $M_{\mathrm{P}} = (8\pi G_{\mathrm{N}})^{-1/2}$ and Λ is the cosmological constant. The integration covers the entire spacetime manifold $\mathcal{M}$.

Any theory that describes nature should in some way incorporate matter, which is described by its action S_{m}. The energy and momentum content of matter is then described by the energy-momentum tensor $T_{\mu\nu}$, which is defined as

$$T_{\mu\nu} \equiv \frac{2}{\sqrt{-g}} \frac{\delta S_{\mathrm{m}}}{\delta g^{\mu\nu}}.$$

The Einstein-Hilbert action leads to the *Einstein field equations of motion*, which determine the dynamics of the gravitational metric. These dynamics are fixed by the condition that they ensure that the action is invariant under small perturbations. This can be formulated as the condition $\delta S = 0$. The variation of the Einstein-Hilbert action then leads to, up to boundary terms, the field equations for the metric tensor:

$$R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \tfrac{1}{2}\kappa^{-1}T_{\mu\nu}. \tag{2.2}$$

It follows from the Bianchi identity and the metric-compatibility condition that $T_{\mu\nu}$ is covariantly conserved:

$$\nabla^\mu T_{\mu\nu} = \tfrac{1}{2}\kappa^{-1}\nabla^\mu \left(R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu}\right) = 0.$$

A consistent derivation of the Einstein field equations requires a careful analysis of the boundary terms. This will be done in **section 2.6**, for completeness.

2.2 FOLIATION OF SPACETIME

The Hamiltonian formulation of general relativity is the necessary framework in which inflationary cosmology and canonical quantum gravity can be described. Since the Hamiltonian can be regarded as the generator of time evolution, the Hamiltonian framework places a special significance on the coordinate of time. In contrast, general relativity unites space and time in the single object of spacetime, which, in addition, is itself dynamical. Any Hamiltonian formulation of general relativity would therefore have to formulate its covariance such that one can distinguish a time parameter along which evolution can be defined. This is typically accomplished within the ADM formalism. This section gives a systematic expansion of spacetime in terms of spatial hypersurfaces, which are connected by the integral curves of a vector that takes the role of the generator of time.

Consider a smooth manifold $\mathcal{M}$ of dimension $d+1$. The manifold $\mathcal{M}$ is equipped with a metric g, which in local coordinates can be expressed as $g_{\mu\nu}$. $\mathcal{M}$ is taken to be Lorentzian, by which it is meant that $\mathcal{M}$ can be brought in the local canonical form

$$g_{\mu\nu} = \mathrm{diag}(-1, 1, 1, \ldots, 1, 1).$$

In order to guarantee that a meaningful foliation of spacetime exists it is necessary that $\mathcal{M}$ is not entirely arbitrary, but subject to certain restrictions, which will be made precise below.

A *hypersurface* Σ of $\mathcal{M}$ is a submanifold that can, in principle, be timelike, lightlike or spacelike. Such a submanifold can be formed as a level surface of some scalar field T, such that

$$T(x^\mu) = \text{constant}.$$

The existence of such a scalar field requires that spacetime has a certain structure: it must be able to support Cauchy surfaces.

A *Cauchy surface* is a spacelike hypersurface Σ in $\mathcal{M}$ such that each timelike or lightlike curve intersects Σ only once. A manifold that admits a Cauchy surface is called *globally hyperbolic*. The nature of the spacetime foliation considered in this work limits the class of acceptable spacetimes to globally hyperbolic spacetimes. This is not a strong restriction; most spacetimes of both astrophysical and cosmological interest are of this nature. The topology of a spacetime that is globally hyperbolic is the product $\mathbb{R} \times \Sigma$, where $\mathbb{R}$ corresponds with a timelike coordinate, and Σ is a general manifold which, in the context of the canonical formalism that is introduced in **chapter 5**, is typically spacelike. For globally hyperbolic spacetimes there exists a smooth scalar field T, which has a nowhere vanishing gradient. A *foliation* is then defined by the level surfaces Σ_t of T, which define d-dimensional submanifolds of $\mathcal{M}$. Since the gradient of T is non-zero everywhere, level surfaces are non-intersecting:

$$\Sigma_t \cap \Sigma_{t'} = \varnothing \text{ for } t \neq t'.$$

Each hypersurface is called a *leaf* of the foliation. In the following the character of the leaves—whether they are timelike or spacelike—is kept general, but it is assumed that their union composes the whole of $\mathcal{M}$:

$$\mathcal{M} = \bigcup_{t\in\mathbb{R}} \Sigma_t.$$

2.3 INTRINSIC HYPERSURFACE GEOMETRY

For a given foliation on the spacetime manifold $\mathcal{M}$ one can study the intrinsic geometry of tensor fields on the leaves Σ_t.

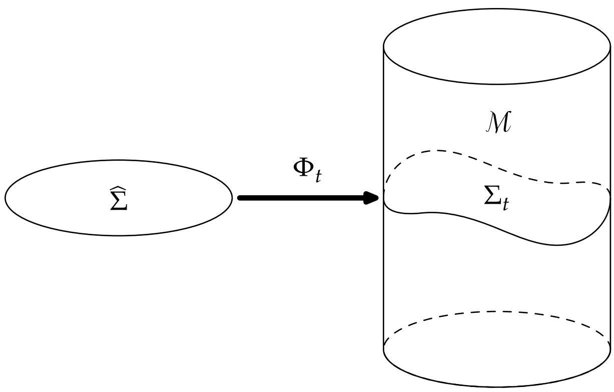

Figure 2.1 Hypersurfaces Σ_t in $\mathcal{M}$ as an embedding of $\hat{\Sigma}$.

The leaves can be described as the *embedding* of a manifold $\Phi_t : \hat{\Sigma} \to \Sigma_t$ in $\mathcal{M}$, as depicted in **figure 2.1**. This $\hat{\Sigma}$ has dimension d, and can thus intrinsically be described by coordinates x^a $(a = 1, \ldots, d)$. The embedding map Φ_t induces the Jacobi matrix $e^\mu{}_a$ as

$$e^\mu{}_a \equiv \frac{\partial X^\mu}{\partial x^a},$$

and can be used to push forward tangent vectors from Σ_t to $\mathcal{M}$. For example:

$$\partial_a = e^\mu{}_a \partial_\mu.$$

The $e^\mu{}_a$ form a linearly independent set of tangent vectors in Σ_t. Vectors n_μ normal to Σ_t are then implicitly defined through the relation

$$e^\mu{}_a n_\mu = 0.$$

The tangent vectors $e^\mu{}_a$ can be used to construct a metric γ_{ab} on $\widehat{\Sigma}$. The $e^\mu{}_a$ define a pullback from $\mathcal{M}$ to $\widehat{\Sigma}$. A metric on $\widehat{\Sigma}$ is naturally given by the pullback of the restriction of $g_{\mu\nu}$ to Σ_t:

$$\gamma_{ab} = e^\mu{}_a e^\nu{}_b g_{\mu\nu}.$$

This is a metric on $\widehat{\Sigma}$, as it is a scalar under coordinate transformations of the X^μ in $\mathcal{M}$, but a rank 2 tensor under coordinate transformations of the x^a in $\widehat{\Sigma}$.

One can go on and define vector and tensor fields on $\widehat{\Sigma}$. A covariant derivative D can be defined for these fields which is compatible with the induced metric. This covariant derivative then satisfies the Ricci identity on $\widehat{\Sigma}$. For example, for a vector v^a one obtains

$$[\mathrm{D}_c, \mathrm{D}_d]\, v^a = {}^{(\mathrm{s})}R^a{}_{bcd} v^b, \tag{2.3}$$

where ${}^{(\mathrm{s})}R^a{}_{bcd}$ is the Riemann tensor constructed from γ_{ab}.

2.4 EXTRINSIC HYPERSURFACE GEOMETRY

Spacetime foliations can be described purely in terms of hypersurfaces in the ambient manifold $\mathcal{M}$, where Σ_t is not considered to be an intrinsic manifold.

Given the scalar field T a covector T_μ normal to the hypersurfaces Σ_t can be constructed through

$$T_\mu = \nabla_\mu T.$$

The norm of T_μ is parametrised by a scalar function N:

$$g^{\mu\nu} T_\mu T_\nu = N^{-2}.$$

The function N is called the *lapse function.* It plays an important role in the canonical formulation of general relativity. Of importance for this section is the unit normal covector n_μ, which is defined as

$$n_\mu \equiv \varepsilon N \nabla_\mu T, \tag{2.4}$$

where $\varepsilon = 1$ for timelike hypersurfaces and $\varepsilon = -1$ for spacelike hypersurfaces. The sign is chosen such that the normal vector n^μ points in the direction of increasing T. The norm of n_μ depends on the character of the hypersurfaces, that is, whether they are spacelike, timelike or lightlike:

$$n^\mu n_\mu = \varepsilon. \tag{2.5}$$

From the point of view of the foliation of the manifold, n_μ distinguishes a preferred direction in $\mathcal{M}$. It is then natural to define the projectors parallel and perpendicular to n_μ:

$$\mathrm{P}_{\parallel}{}^{\mu}{}_{\nu} = \varepsilon n^\mu n_\nu,$$
$$\mathrm{P}_{\perp}{}^{\mu}{}_{\nu} = \delta^\mu{}_\nu - \varepsilon n^\mu n_\nu.$$

It is straightforward to see that

$$\mathrm{P}_{\parallel}{}^{\mu}{}_{\nu} + \mathrm{P}_{\perp}{}^{\mu}{}_{\nu} = \delta^\mu_\nu,$$
$$\mathrm{P}_{\parallel}{}^{\mu}{}_{\rho}\,\mathrm{P}_{\parallel}{}^{\rho}{}_{\nu} = \mathrm{P}_{\parallel}{}^{\mu}{}_{\nu},$$
$$\mathrm{P}_{\perp}{}^{\mu}{}_{\rho}\,\mathrm{P}_{\perp}{}^{\rho}{}_{\nu} = \mathrm{P}_{\perp}{}^{\mu}{}_{\nu}.$$

Tensor fields in the kernel of $\mathrm{P}_{\parallel}{}^{\mu}{}_{\nu}$ are called ***tangential*** or, if the hypersurfaces are spacelike, ***spatial tensors***. Since their contraction with n_μ vanishes, they can be pulled back to $\hat{\Sigma}$ with no loss of information.

A projected covariant derivative D_μ along Σ_t can be defined through the projection of each index of the covariant derivative ∇_μ on $\mathcal{M}$ to Σ_t. For example, the covariant for a tangential vector v_μ can be written as

$$\mathrm{D}_\mu v_\nu = \mathrm{P}_{\perp}{}^{\alpha}{}_{\mu}\,\mathrm{P}_{\perp}{}^{\beta}{}_{\nu}\,\nabla_\alpha v_\beta. \tag{2.6}$$

For tangential tensors this derivative coincides with the intrinsic covariant derivative, and the pullback in this case amounts to a coordinate transformation from Σ_t to $\hat{\Sigma}$:

$$\mathrm{D}_a v_b = e^\mu{}_a e^\nu{}_b\,\mathrm{D}_\mu v_\nu.$$

In general, each index of a tensor in $\mathcal{M}$ can be decomposed in terms of parts parallel and orthogonal to the normal vector. This results in a polynomial in n_μ. For the decomposition of the metric one obtains

$$\begin{aligned} g_{\mu\nu} &= \delta^\alpha_\mu \delta^\beta_\nu g_{\alpha\beta} \\ &= \left(\mathrm{P}_{\perp}{}^{\alpha}{}_{\mu} + \mathrm{P}_{\parallel}{}^{\alpha}{}_{\mu}\right)\left(\mathrm{P}_{\perp}{}^{\beta}{}_{\nu} + \mathrm{P}_{\parallel}{}^{\beta}{}_{\nu}\right) g_{\mu\nu} \\ &= \mathrm{P}_{\perp}{}^{\alpha}{}_{\mu}\,\mathrm{P}_{\perp}{}^{\beta}{}_{\nu} g_{\alpha\beta} + \varepsilon n_\mu n_\nu. \end{aligned} \tag{2.7}$$

A tangential tensor will be denoted by a tilde, for example:

$$\mathrm{P}_{\perp}{}^{\alpha}{}_{\mu}\,\mathrm{P}_{\perp}{}^{\beta}{}_{\nu} g_{\alpha\beta} = \tilde{g}_{\mu\nu}.$$

The object $\tilde{g}_{\mu\nu}$ is known as the ***first fundamental form***. In order to keep notational clarity it will be convenient to define it as a separate symbol:

The first fundamental form is related to the induced metric via $\gamma_{ab} = e^\mu{}_a e^\nu{}_b \gamma_{\mu\nu}$.

$$\gamma_{\mu\nu} \equiv \tilde{g}_{\mu\nu}.$$

As $\gamma_{\mu\nu}$ corresponds to the induced metric on Σ_t, as far as tangential tensors are concerned, indices can be raised or lowered with $\gamma_{\mu\nu}$ just as well as with $g_{\mu\nu}$.

For n_μ one can define the acceleration $\tilde{a}_\mu$:

$$\tilde{a}_\mu \equiv n^\nu \nabla_\nu n_\mu. \tag{2.8}$$

It can be seen that $\tilde{a}_\mu$ is a tangential tensor, as its contraction with n_μ vanishes by (2.5). One can also define the *second fundamental form* $\tilde{K}_{\mu\nu}$ via the relation

$$\tilde{K}_{\mu\nu} \equiv \mathrm{P}_\perp{}^\alpha{}_\mu \mathrm{P}_\perp{}^\beta{}_\nu \nabla_\alpha n_\beta = \nabla_\mu n_\nu - \varepsilon n_\mu \tilde{a}_\nu. \tag{2.9}$$

It follows from (2.4) that $\tilde{K}_{\mu\nu}$ is symmetric. The geometric meaning of the second fundamental form can be inferred from figure 2.2. The normal vector n^μ describes a vector field in $\mathcal{M}$. Therefore, if the normal vector n^μ at point p is parallel transported to point q, the resulting vector $\tilde{n}^\mu$ will in general not coincide with the normal vector n^μ at q. The difference between these two vectors, when projected onto Σ_t, gives a measure for the curvature of Σ_t. This curvature is not intrinsic to Σ_t itself, but is an effect of its embedding in $\mathcal{M}$. For this reason $\tilde{K}_{\mu\nu}$ is frequently called the *extrinsic curvature.*

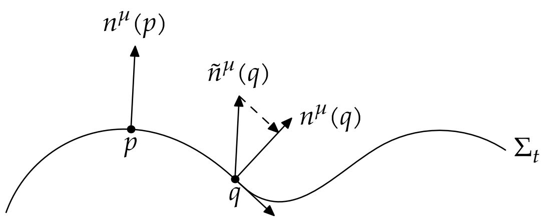

Figure 2.2 Geometric interpretation of $\tilde{K}_{\mu\nu}$.

The extrinsic curvature is a tangential tensor by construction.

The evolution of spatial tensors is naturally given by the Lie derivative along n_μ. One important case is the Lie derivative of the induced metric. From (2.7) and (2.9) it follows that

$$\begin{aligned}
(\mathcal{L}_n \gamma)_{\mu\nu} &= n^\alpha \nabla_\alpha \gamma_{\mu\nu} + \gamma_{\alpha\nu} \nabla_\mu n^\alpha + \gamma_{\mu\alpha} \nabla_\nu n^\alpha \\
&= -\varepsilon n^\alpha \nabla_\alpha (n_\mu n_\nu) + \gamma_{\alpha\nu} (\tilde{K}_\mu{}^\alpha + \varepsilon n_\mu \tilde{a}^\alpha) + \gamma_{\mu\alpha} (\tilde{K}_\nu{}^\alpha + \varepsilon n_\nu \tilde{a}^\alpha) \\
&= 2\tilde{K}_{\mu\nu}.
\end{aligned}$$

It can be seen that in this case the normal Lie derivative of a tangential tensor again produces a spatial tensor. This is true in general for cotensors: consider a tangential cotensor $\tilde{T}_{\nu_1 \ldots \nu_m}$. Then, its Lie derivative along n_μ is given by [16]

$$(\mathcal{L}_n \tilde{T})_{\nu_1 \ldots \nu_m} = n^\mu \nabla_\mu \tilde{T}_{\nu_1 \ldots \nu_m} + \sum_{i=1}^{m} \tilde{T}_{\nu_1 \ldots \mu \ldots \nu_m} \nabla_{\nu_i} n^\mu. \tag{2.10}$$

Now, contraction of (2.10) with n^{ν_k} for some $1 \le k \le m$ leads to

$$n^{\nu_k}(\mathcal{L}_n\,\tilde{T})_{\nu_1\ldots\nu_m} = n^{\nu_k}n^{\alpha}\nabla_{\alpha}\tilde{T}_{\nu_1\ldots\nu_m} + \tilde{T}_{\nu_1\ldots\alpha\ldots\nu_m}n^{\nu_k}\nabla_{\nu_k}n^{\alpha}$$
$$= n^{\alpha}\nabla_{\alpha}(n^{\beta}\tilde{T}_{\nu_1\ldots\beta\ldots\nu_m})$$
$$= 0.$$

That the tangential property is preserved under Lie differentiation only for cotensors is not a strong restriction. The relevant decompositions can be formulated purely in terms of cotensors.

2.4.1 *Decomposition of curvature*

Although tensors can generally be decomposed as a polynomial in the normal covector n_{μ}, it is nevertheless instructive to calculate the non-trivial components of the Riemann tensor and its various contractions explicitly.

The Riemann tensor To begin with, it is straightforward to give the components which are partially projected along the normal vector n^{μ}. With **(2.9)** it is found that

$$\nabla_{\mu}\nabla_{\nu}n^{\rho} = \nabla_{\mu}\tilde{K}_{\nu}{}^{\rho} + \varepsilon\tilde{a}^{\rho}\nabla_{\mu}n_{\nu} + \varepsilon n_{\nu}\nabla_{\mu}\tilde{a}^{\rho}.$$

The derivative of n_{ν} in the second term can be eliminated by repeated substitution of **(2.9)**. One then obtains

$$\nabla_{\mu}\nabla_{\nu}n^{\rho} = \nabla_{\mu}\tilde{K}_{\nu}{}^{\rho} + \varepsilon\tilde{K}_{\mu\nu}\tilde{a}^{\rho} + n_{\mu}\tilde{a}^{\rho}\tilde{a}_{\nu} + \varepsilon n_{\nu}\nabla_{\mu}\tilde{a}^{\rho}.$$

Antisymmetrisation over the indices μ and ν leads to, after the index ρ is lowered,

$$R_{\rho\sigma\mu\nu}n^{\sigma} = 2\nabla_{[\mu}\tilde{K}_{\nu]\rho} + 2\tilde{a}_{\rho}n_{[\mu}\tilde{a}_{\nu]} - 2\varepsilon n_{[\mu}\nabla_{\nu]}\tilde{a}_{\rho}. \tag{2.11}$$

The symmetries of the Riemann tensor ensure that all other contractions are related to this one by a sign. Contraction of **(2.11)** with n^{μ}, one obtains

$$R_{\rho\sigma\mu\nu}n^{\sigma}n^{\mu} = \left(\mathcal{L}_n\,\tilde{K}\right)_{\nu\rho} - \tilde{K}_{\nu\lambda}\tilde{K}_{\rho}{}^{\lambda} - \varepsilon\tilde{K}_{\nu\lambda}n_{\rho}\tilde{a}^{\lambda} + \varepsilon\tilde{a}_{\rho}\tilde{a}_{\nu}$$
$$- \nabla_{\nu}\tilde{a}_{\rho} + \varepsilon n_{\nu}(\mathcal{L}_n\,\tilde{a})_{\rho} - \varepsilon n_{\nu}\tilde{K}_{\rho}{}^{\alpha}\tilde{a}_{\alpha} - n_{\rho}n_{\nu}\tilde{a}_{\lambda}\tilde{a}^{\lambda}. \tag{2.12}$$

Further contractions with n^{μ} will vanish due to the symmetries of the Riemann tensor.

Projection of **(2.11)** onto Σ_t results in the *Codazzi-Mainardi relation*

$$\mathrm{P}_{\perp}{}^{\rho}{}_{\alpha}\,\mathrm{P}_{\perp}{}^{\mu}{}_{\beta}\,\mathrm{P}_{\perp}{}^{\nu}{}_{\gamma}R_{\rho\sigma\mu\nu}n^{\sigma} = \mathrm{D}_{\beta}\,\tilde{K}_{\gamma\alpha} - \mathrm{D}_{\gamma}\,\tilde{K}_{\beta\alpha}. \tag{2.13}$$

Projection of **(2.12)** onto Σ_t, one obtains

$$R_{\rho\sigma\mu\nu}\,\mathrm{P}_{\perp}{}^{\rho}{}_{\alpha}\,\mathrm{P}_{\perp}{}^{\nu}{}_{\beta}n^{\sigma}n^{\mu} = (\mathcal{L}_n\,\tilde{K})_{\alpha\beta} - \tilde{K}_{\alpha\lambda}\tilde{K}_{\beta}{}^{\lambda} - (\mathrm{D}_{\beta} - \varepsilon\tilde{a}_{\beta})\tilde{a}_{\alpha}.$$

The tangential components of the Riemann tensor can be found by use of the Ricci identity **(2.3)** for a spatial vector v^{μ} on Σ_t:

$$\left[\mathrm{D}_{\mu}, \mathrm{D}_{\nu}\right]v^{\rho} = {}^{(\mathrm{s})}R^{\rho}{}_{\sigma\mu\nu}v^{\sigma}.$$

Equation (2.6) can be used to express the left-hand side in terms of projections of the ambient space covariant derivatives. The result obtained in this way is the *Gauss relation*

$$\mathrm{P_\perp}{}^{\rho}{}_{\alpha}\,\mathrm{P_\perp}{}^{\sigma}{}_{\beta}\,\mathrm{P_\perp}{}^{\mu}{}_{\gamma}\,\mathrm{P_\perp}{}^{\nu}{}_{\delta}R_{\rho\sigma\mu\nu} = {}^{(\mathrm{s})}R_{\alpha\beta\gamma\delta} - 2\varepsilon\tilde{K}_{\alpha[\gamma}\tilde{K}_{\delta]\beta}. \tag{2.14}$$

This completes the derivation of the non-trivial components of the Riemann tensor. All others are either equal to zero or related the ones above by symmetry.

The Ricci tensor The Ricci tensor is formed out of the contraction of the Riemann tensor, and its normal and tangential components can therefore straightforwardly be computed from the results above.

To begin with, it can be verified that $(\mathcal{L}_n\tilde{K})_{\mu\nu}$ satisfies

$$\gamma^{\mu\nu}(\mathcal{L}_n\tilde{K})_{\mu\nu} = \mathcal{L}_n\tilde{K} + 2\tilde{K}_{\alpha\beta}\tilde{K}^{\alpha\beta}.$$

With this identity the normal-normal component of the Ricci tensor can be computed through the contraction of (2.12) with $g_{\mu\nu}$. The result is

$$R_{\mu\nu}n^{\mu}n^{\nu} = -\mathcal{L}_n\tilde{K} - \tilde{K}_{\mu\nu}\tilde{K}^{\mu\nu} + \nabla_{\mu}\tilde{a}^{\mu}. \tag{2.15}$$

Contraction of (2.13) with $\gamma_{\alpha\beta}$ results in the *contracted Codazzi equation*, which is the mixed component of the Ricci tensor:

$$\mathrm{P_\perp}{}^{\nu}{}_{\alpha}R_{\mu\nu}n^{\mu} = \mathrm{D}_{\lambda}\tilde{K}_{\alpha}{}^{\lambda} - \mathrm{D}_{\alpha}\tilde{K}_{\lambda}{}^{\lambda}. \tag{2.16}$$

Contraction of the Gauss relation with $\gamma^{\alpha\beta}$ and substitution of (2.15) results in the *contracted Gauss relation*

$$\begin{aligned}\mathrm{P_\perp}{}^{\mu}{}_{\alpha}\,\mathrm{P_\perp}{}^{\nu}{}_{\beta}R_{\mu\nu} = {}^{(\mathrm{s})}R_{\alpha\beta} - \varepsilon\tilde{K}\tilde{K}_{\alpha\beta} + 2\tilde{K}_{\alpha\lambda}\tilde{K}^{\lambda}{}_{\beta}\\ - \varepsilon(\mathcal{L}_n\tilde{K})_{\alpha\beta} + \varepsilon(\mathrm{D}_{\beta} - \varepsilon\tilde{a}_{\beta})\tilde{a}_{\alpha}.\end{aligned} \tag{2.17}$$

This completes the decomposition of the normal and tangential components of the Ricci tensor.

The Ricci scalar The Ricci scalar can now be obtained via contraction of the Ricci tensor with the inverse metric. Substitution of (2.7), (2.15) and (2.17) yields

$$R = \gamma^{\mu\nu}\,\mathrm{P_\perp}{}^{\alpha}{}_{\nu}\,\mathrm{P_\perp}{}^{\beta}{}_{\nu}R_{\alpha\beta} + \varepsilon n^{\mu}n^{\nu}R_{\mu\nu}.$$

In the literature the scalar Gauss relation is also known as the *Gauss-Codazzi equation.*

Substitution of (2.15) and (2.17) yields the *scalar Gauss relation*

$$R = {}^{(\mathrm{s})}R + \varepsilon(\tilde{K}^2 - \tilde{K}_{\alpha\beta}\tilde{K}^{\alpha\beta}) - 2\varepsilon(\mathcal{L}_n + \tilde{K})\tilde{K} + 2\varepsilon(\mathrm{D}_{\alpha} - \varepsilon\tilde{a}_{\alpha})\tilde{a}^{\alpha}, \tag{2.18}$$

which gives the full orthogonal decomposition of the scalar curvature in terms of normal and tangential objects.

2.4.2 *Decomposition of the d'Alembert operator*

Scalar operators can be decomposed in terms of orthogonal and tangential parts, similar to tensors. One useful quantity that will be used later is the d'Alembert operator $\Box = g^{\mu\nu}\nabla_\mu\nabla_\nu$ for scalar fields ϕ. Substitution of the decomposition (2.7) leads to the following expression:

$$\begin{aligned}\Box\phi &= \gamma^{\mu\nu}\nabla_\mu\nabla_\nu\phi + \varepsilon n^\mu n^\nu\nabla_\mu\nabla_\nu\phi \\ &= \gamma^{\mu\nu}\,\mathrm{P}_\perp{}^\rho{}_\mu\,\mathrm{P}_\perp{}^\sigma{}_\nu\nabla_\rho\nabla_\sigma\phi - \varepsilon\tilde{a}^\mu\,\mathrm{D}_\mu\,\phi + \varepsilon\,\mathcal{L}_n(\mathcal{L}_n\,\phi).\end{aligned} \tag{2.19}$$

The first term can be rewritten in terms of spatial derivatives and Lie derivatives along n_μ. The covariant spatial derivative is related to the covariant ambient derivative via

$$\mathrm{D}_\mu\,\mathrm{D}_\nu\,\phi = \mathrm{P}_\perp{}^\rho{}_\mu\,\mathrm{P}_\perp{}^\sigma{}_\nu\nabla_\rho\,\mathrm{P}_\perp{}^\lambda{}_\sigma\nabla_\lambda\phi.$$

Expansion of the right-hand side and yields an expression for the second covariant derivative of ϕ:

$$\mathrm{P}_\perp{}^\rho{}_\mu\,\mathrm{P}_\perp{}^\sigma{}_\nu\nabla_\mu\nabla_\nu\phi = \mathrm{D}_\mu\,\mathrm{D}_\nu\,\phi + \varepsilon\tilde{K}_{\mu\nu}\,\mathcal{L}_n\,\phi. \tag{2.20}$$

The combination (2.19) with (2.20) then results in the following decomposition in normal and tangential parts:

$$\Box\phi = (\mathrm{D}^\mu - \varepsilon\tilde{a}^\mu)\,\mathrm{D}_\mu\,\phi + \varepsilon\left(\mathcal{L}_n + \tilde{K}\right)\mathcal{L}_n\,\phi.$$

From the d'Alembert operator it can be seen how coviariant derivatives for orthogonal and tangential tensors can be constructed from the point of view of the ambient manifold, but these will not be explicitly necessary in this work.

2.5 ADM LINE ELEMENT

The ADM *decomposition* is a foliation of spacetime in terms of hypersurfaces of signature ε, where the scalar field T is identified with the time field that connects the different Σ_t for different values of t. Given a foliation in which the coordinates $X^\mu = X^\mu(t, x^a)$ describe spacetime, the tangent vectors to the leaves Σ_t are given by

$$e^\mu{}_a = \frac{\partial X^\mu}{\partial x^a}\bigg|_t,$$

as before, while

$$T^\mu = \frac{\partial X^\mu}{\partial t}\bigg|_x$$

gives the components of the vector field that generates time evolution. Generally, T^μ will not be orthogonal to the Σ_t, and hence it can be decomposed into normal and tangential parts as in **figure 2.3**:

$$T^\mu \equiv Nn^\mu + N^\mu = Nn^\mu + N^a e^\mu{}_a.$$

Here N is the lapse function, as before, while N^μ (or, equivalently, N^a) is called the *shift vector.*

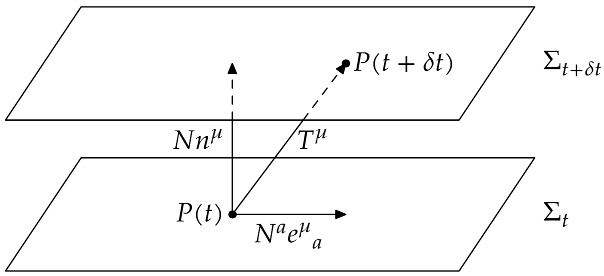

Figure 2.3 Decomposition of time evolution in terms of the lapse and shift.

The line element can then readily be decomposed into tangential and normal parts:

$$\begin{aligned} \mathrm{d}s^2 &= g_{\mu\nu}\,\mathrm{d}X^\mu \mathrm{d}X^\nu \\ &= (\gamma_{\mu\nu} + \varepsilon n_\mu n_\nu)\,(T^\mu\,\mathrm{d}t + \mathrm{d}x^\mu)\,(T^\nu\,\mathrm{d}t + \mathrm{d}x^\nu) \\ &= \varepsilon N^2 \mathrm{d}t^2 + (\mathrm{d}x^a + N^a \mathrm{d}t)\gamma_{ab}(\mathrm{d}x^b + N^b \mathrm{d}t). \end{aligned} \tag{2.21}$$

The last object that is left to decompose is the metric determinant. From the line element it follows that $g^{tt} = \mathrm{cofactor}(g_{tt})/g = h/g$. Similarly, the (00) component of $g^{\mu\nu}$ can be calculated to be $g^{00} = \varepsilon N^{-2}$, and hence the volume element is given by

$$\sqrt{\varepsilon g} = N\sqrt{\gamma}. \tag{2.22}$$

2.6 ADM ACTION AND BOUNDARY TERMS

In **section 2.1** it was stated that the Einstein-Hilbert action must be supplemented with boundary terms if the spacetime manifold has a non-trivial boundary. The nature of these boundary terms can readily be understood within the framework of hypersurface foliation. The treatment of these terms is given here for completeness.

A straightforward way to derive the boundary terms is through the use of the ADM spacetime foliation. Substitution of **(2.18)** and **(2.22)** into **(2.1)** yields the ADM form of the action:

$$\begin{aligned} S_{\mathrm{EH}} = \kappa \int_{\mathcal{M}} \Big[{}^{(\mathrm{s})}R - \varepsilon \left(\tilde{K}_{\mu\nu} \tilde{K}^{\mu\nu} - \tilde{K}^2 \right) \\ - 2\varepsilon \left(\mathcal{L}_n \tilde{K} + \tilde{K}^2 - \mathrm{D}_\mu \tilde{a}^\mu + \varepsilon \tilde{a}^\mu \tilde{a}_\mu \right) \Big] \sqrt{\gamma} N \mathrm{d}t \mathrm{d}^d x. \end{aligned}$$

Comparison with **(2.8)** and **(2.9)** shows that the last term above is a total derivative:

$$\mathcal{L}_n \tilde{K} + \tilde{K}^2 - \mathrm{D}_\mu \tilde{a}^\mu + \varepsilon \tilde{a}^\mu \tilde{a}_\mu = \nabla_\mu (n^\mu \tilde{K} - \tilde{a}^\mu).$$

Application of Stokes' theorem then yields [16]

$$\int_{\mathcal{M}} 2\varepsilon \left(\mathcal{L}_n \tilde{K} + \tilde{K}^2 - \mathrm{D}_\mu \tilde{a}^\mu + \varepsilon \tilde{a}_\mu \tilde{a}^\mu \right) \sqrt{\gamma} N \mathrm{d}t \mathrm{d}^d x = 2\varepsilon \oint_{\partial\mathcal{M}} (n^\mu \tilde{K} - a^\mu)\, \mathrm{d}\sigma_\mu, \tag{2.23}$$

where on the right-hand side the integration is over the boundary $\partial\mathcal{M}$ of $\mathcal{M}$. On the right-hand side integration is implied over the boundary $\partial\mathcal{M}$. As the hypersurface element $\mathrm{d}\sigma_\mu$ is a surface normal, it must be proportional to n_μ. Furthermore, the scalar surface element must correspond to a scalar surface patch on the boundary hypersurface. It follows that

$$\mathrm{d}\sigma_\mu = \varepsilon n_\mu \sqrt{\gamma}\, \mathrm{d}^d x, \tag{2.24}$$

where the factor of ε ensures that outward flux through timelike surfaces is positive. Insertion of **(2.24)** into **(2.23)** yields the *Gibbons-Hawking-York* (GHY) *boundary term* of the action of general relativity [17,18]:

If the manifold has additional boundaries beyond the ones induced by the spacetime foliation, each boundary should be considered separately.

$$S_{\mathrm{EH}} = S_{\mathrm{ADM}} - S_{\mathrm{GHY}}, \text{ with} \tag{2.25}$$
$$S_{\mathrm{ADM}} \equiv \kappa \int_{\mathcal{M}} \left[{}^{(s)}R - \varepsilon \left(\tilde{K}_{\mu\nu} \tilde{K}^{\mu\nu} - \tilde{K}^2 \right) \right] \sqrt{\gamma} N \mathrm{d}t \mathrm{d}^d x,$$
$$S_{\mathrm{GHY}} \equiv 2\varepsilon\kappa \oint_{\partial\mathcal{M}} \tilde{K} \sqrt{\gamma}\, \mathrm{d}^d x.$$

The presence of the GHY boundary term shows that the Einstein-Hilbert action is not differentiable from the perspective of variational calculus; it prevents the variation of **(2.25)** to be written as a functional derivative. This is straightforwardly remedied by adding this term on the left-hand side. The result is that the gravitational action is

$$S_{\mathrm{EH}} + S_{\mathrm{GHY}} = S_{\mathrm{ADM}}.$$

The variation of S_{ADM} then leads to **(2.2)**. The action, of which the variation reproduces the Einstein field equations, can be seen to be S_{ADM}.

The GHY boundary term diverges in spacetimes that are asymptotically flat. For this reason, an additional boundary term is subtracted from the action, where the integrand is equal to $2\varepsilon \tilde{K}_0 \sqrt{\gamma}$, where $\tilde{K}_0$ is the asymptotic value of the extrinsic curvature of the boundary, embedded in flat space. From the perspective of variational calculus this extra term is a constant, and therefore it does not affect the field equations **(2.2)**. The reason to include this term is that it renders the action finite.

The GHY boundary term can readily be generalised to modified theories of gravity, such as higher derivative gravity or the scalar-tensor theories that are described in **chapter 3**. However, such a generalisation goes beyond the scope of this work. An extended treatment about boundary terms in general relativity, as well as their extensions in theories of modified gravity, is presented in reference [19] and the references therein.

All the gravitational degrees of freedom can be represented by the spatial metric, the lapse function and the shift vector. The functional forms of these objects within the framework of cosmology can be inferred from physical observations, which will be done here.

It is known from observations that the earth does not occupy a privileged position within the solar system. Observations of the large-scale structure of the universe show that, at length scales of the order of approximately 100 Mpc, the universe appears to be highly homogeneous and isotropic. This is called the *cosmological principle*. This has the consequence that there is no point in space that is preferred. The metric tensor is therefore homogeneous and isotropic at these length scales.

The cosmological principle poses a requirement for the parametrisation of the metric tensor. Considering **(2.21)**, isotropy implies that the shift vector be zero everywhere. Similarly, homogeneity implies that the lapse function can only depend on time, and that the spatial metric is both homogeneous and isotropic in space and homogeneous in time. Thus, the ADM line element reduces to

$$\mathrm{d}s^2 = \varepsilon N^2(t)\mathrm{d}t^2 + a^2(t)\gamma_{ab}(\mathbf{x})\mathrm{d}x^a\mathrm{d}x^b. \tag{2.26}$$

The function a implies that proper distances change with time. For this reason it is called the *scale factor*. The restriction that γ_{ab} be homogeneous and isotropic in space fixes its geometry. This will be explicitly considered in **chapter 4**.

SCALAR-TENSOR THEORIES

3

THE THEORY OF GENERAL RELATIVITY provides the most accurate macroscopic description of gravity as a dynamical interaction between spacetime—the dynamical union of space and time—and the matter fields that pervade it. Nevertheless, new discoveries led to the belief that alternative theories, based on different fundamental principles, might provide a description of phenomena that are not straightforwardly explained by general relativity. The origin of this belief can be traced back to a mismatch between the theory of general relativity and observations in the large and the small: cosmology and quantum field theory. For example, the cosmological singularity at the beginning of the universeis generally taken to signify that the theory of general relativity is not applicable during the time of the very early universe. Furthermore, it is known that the universe underwent a period of accelerated expansion at least twice during its lifetime. Although solutions to the Einstein field equations **(2.2)** exist that describe universes that undergo such expansion, they provide neither a reason why such a theory is meaningful nor a satisfying explanation of the underlying fundamental principles. From the side of particle physics it is known that the theory of general relativity is non-renormalisable. Any attempt to treat general relativity as an effective theory of quantum gravity results in the presence of counter terms that contain higher order curvature invariants, and therefore strongly modify the theory and its subsequent predictions. Furthermore, it is known that the renormalisation of a theory that describes a scalar field in a curved spacetime results in a direct coupling between the Ricci curvature and the scalar field itself. Thus, it appears that the formulation of a quantum field theory in a curved spacetime results in a modification of the Einstein-Hilbert action **(2.1)**.

It is therefore natural to wonder in what ways Einstein's theory of general relativity can be modified in order to resolve these issues. It turns out that, as a classical field theory, these modifications can be classified into simple groups. In a 4-dimensional spacetime general relativity is the unique relativistic field theory of which its action yields equations of motion for the gravitational field that are second order in time. This is a consequence of Lovelock's theorem [20]. One must therefore find other ways to modify general relativity if one is to insist on a spacetime that is 4-dimensional in nature. Two such methods were described above:

I. introduce higher-order curvature invariants,

II. introduce additional field couplings.

Both approaches will be briefly investigated in the next sections, with an emphasis on the latter. The simplest such extension of general relativity is the introduction of a scalar field that is coupled to the gravitational field in a special way. Such a coupling is said to be *non-minimal*, as opposed to the ordinary *minimal coupling* that is exhibited by a matter field in general relativity. These models are called *scalar-tensor theories*, and have various applications that are cosmological in nature. Examples include the theory of *quintessence*, which aims to describe the accelerated expansion

Such a theory is a special case of Horndeski's theory of gravitation [21].

that is observed in the universe today, and *Brans-Dicke theory*, which was motivated by philosophical considerations of the structure of the universe.

Point II above can be realised in various ways. In addition to coupling a scalar field to gravity in a non-minimal way, one can introduce families of scalar field, or non-minimal couplings of vector or tensor fields. This work focusses on single-field scalar-tensor theories, and therefore this chapter provides a brief introduction to their general properties and relation to alternative modified theories of gravitation. For a more comprehensive overview the reader is invited to consult the review works that are available. An introduction to the field is provided by the references [22–28].

3.1 SCALAR-TENSOR THEORIES OF GRAVITATION

One of the key differences between the theory of general relativity and scalar-tensor theories can be found in their respective assumptions about the interplay between gravity and matter. General relativity can be formulated in terms of the *principle of equivalence*. This principle implies that all masses experience gravitational forces in the same way, irrespective of their internal composition, and the gravitational coupling is an absolute constant. There is no reason to assume *a priori* that this is a fundamental principle, and alternative formulations of gravitation allow for the coupling of gravity to matter to come about through a spontaneous symmetry breaking. The coupling constant of gravity would therefore vary from point to point in spacetime, and would subsequently be dynamical. In other words, the gravitational coupling would be, in effect, a field.

In the simplest theories that have such a dynamical formulation the actual value of the gravitational coupling is determined in part by the matter fields, the influence of which arises as a source in the equations of motion. In this way, the strength of gravity is determined by the mass distribution of the universe. Scalar-tensor theories are therefore based in part on *Mach's principle* [29]. The cost for Mach's principle to be realised in a theory of gravitation is that the gravitational force varies over different regions of spacetime. Therefore, these theories of gravitation violate the *strong equivalence principle*.

Scalar-tensor theories replace the gravitational constant in the theory of general relativity with a field that couples directly to the derivatives of the gravitational metric. One of the earliest examples of this can be found in *Brans-Dicke theory* [30,31], where the Planck mass is replaced by a scalar field ϕ:

$$S_{\mathrm{BD}} = \int_{\mathcal{M}} \left(\phi R - \omega\, \phi^{-1} \nabla_\mu \phi \nabla^\mu \phi \right) \sqrt{-g}\, \mathrm{d}^4 X. \tag{3.1}$$

Here, ω is the Brans-Dicke coupling constant. Scalar-tensor theories can also be motivated beyond the philosophical principles outlined above. For example, string theories have scalar-tensor theories as their low-energy limit, rather than general relativity, as they predict that the spin-2 graviton has a scalar partner, called the dilaton [32,33]. Hence, on the assumption that string theory is correct one is led to the belief that gravitation is a scalar-tensor theory, rather than the purely geometrical theory of general relativity.

Of course, one final motivation for scalar-tensor theories is that there is nothing that contradicts such a formulation of gravitation in the first place. Their structure is conceptually different from general relativity, and merits an investigation in its own right.

When the scalar field takes on the role of the inflaton field, scalar-tensor theories are able to naturally explain the accelerated expansion of the early universe [34–36]. This is in contrast to traditional models of inflation, where the scalar field has to be introduced *ad hoc*. In fact, the increasingly precise measurements of the inflationary parameters show a slight preference for non-minimally coupled models. Furthermore, the non-minimal coupling allows scalar-tensor theories to bridge the gap between the energy scale of inflation and the energy scales accessible to particle accelerators today. The dynamics of single-field scalar-tensor theories are described in this chapter, and their relation to observations is described in **chapter 4**.

3.2 DYNAMICS OF SCALAR-TENSOR THEORIES

This work considers a generalisation of the Brans-Dicke theory **(3.1)** and considers the single field scalar-tensor theory described by the action

The cosmological constant has been neglected here, as it is essentially a contribution to the potential.

$$S[g,\phi] = \int \left(UR - \tfrac{1}{2} G \nabla_\mu \phi \nabla^\mu \phi - V \right) \sqrt{-g}\, \mathrm{d}^4 X, \tag{3.2}$$

where U, G and V are model-specific functions of ϕ that describe the non-minimal coupling, normalisation of the kinetic term and potential energy of the scalar field, respectively. For later applications it will be convenient to define a new dimensionless function s, which in terms of U and G is given as

$$s \equiv \frac{U}{GU + 3U_1^2}. \tag{3.3}$$

In perturbation theory this function typically leads to a suppression of the scalar field propagator, and for this reason it is sometimes called a *suppression function* in the literature [37–39]. The subscript notation denotes derivatives of functions of the scalar field with respect to their argument:

$$f_i(\phi) \equiv \frac{\partial^i f}{\partial \phi^i}(\phi).$$

The action **(3.2)** is able to account for a large number of scalar-tensor theories that have applications in cosmology.

The field equations for $g_{\mu\nu}$ and ϕ can straightforwardly be found through the variation of **(3.2)** with respect to $g_{\mu\nu}$ and ϕ:

$$\begin{aligned} R_{\mu\nu} - \tfrac{1}{2} R g_{\mu\nu} &= \tfrac{1}{2} U^{-1} T^{\phi}_{\mu\nu}, \\ G \Box\phi + \tfrac{1}{2} G_1 \nabla_\mu \phi \nabla^\mu \phi - V_1 &= -U_1 R. \end{aligned}$$

The effective energy-momentum tensor takes the form

$$T^{\phi}_{\mu\nu} = G\left(\delta^{\alpha}_{\mu}\delta^{\beta}_{\nu} - \tfrac{1}{2}g_{\mu\nu}g^{\alpha\beta}\right)\nabla_{\alpha}\phi\nabla_{\beta}\phi - g_{\mu\nu}V + 2\nabla_{\mu}\nabla_{\nu}U - 2g_{\mu\nu}\Box U.$$

It is not difficult to see that the field equations and the energy-momentum tensor reduce to their known expressions in the case of a minimally coupled scalar field, where $U = \frac{1}{2}M_{\mathrm{P}}^2$. The field equation for ϕ can be expressed as a Klein-Gordon-type equation that is independent of the curvature of spacetime. It can be most compactly expressed in term of U, s and V:

$$\Box\phi + \tfrac{1}{2}\nabla_{\mu}\log(U/s)\nabla^{\mu}\phi - sU^2W_1 = 0, \qquad W \equiv VU^{-2}. \tag{3.4}$$

The function W has a geometrical meaning in its own right. This will become clear in **section 3.4**.

3.3 CONFORMAL TRANSFORMATIONS

Scalar-tensor theories allow a parametrisation of fields that has a similar form as Einstein-Hilbert gravity [40]. The parametrisation of the action **(3.2)** is called the *Jordan frame* (JF), while the parametrisation that corresponds to **(2.1)** is called the *Einstein frame* (EF). Both frames are related to each other by a conformal transformation of the metric tensor.

This section describes the effect of a conformal transformation that takes a metric $g_{\mu\nu}$ to a new metric $\tilde{g}_{\mu\nu}$ on a 4-dimensional manifold. The transformation is of the form

$$g_{\mu\nu} = \Omega\,\tilde{g}_{\mu\nu}. \tag{3.5}$$

Here, $\Omega(\phi)$ is the conformal factor, which is a strictly positive function of the scalar field ϕ. The transformation **(3.5)** implies

$$g^{\mu\nu} = \Omega^{-1}\,\tilde{g}^{\mu\nu}, \qquad g^{1/2} = \Omega^2\,\tilde{g}^{1/2}, \qquad \Gamma^{\rho}_{\mu\nu} = \tilde{\Gamma}^{\rho}_{\mu\nu} + \Omega^{\rho}{}_{\mu\nu}, \tag{3.6}$$

where $\Gamma^{\rho}_{\mu\nu}$ are the components of the Christoffel symbol of the metric g, $\tilde{\Gamma}^{\rho}_{\mu\nu}$ the components of the Christoffel symbol of the metric $\tilde{g}_{\mu\nu}$ and $\Omega^{\rho}_{\mu\nu}$ the difference tensor that contains the derivatives of the conformal factor:

$$\Omega^{\rho}{}_{\mu\nu} \equiv \tfrac{1}{2}\Omega^{-1}\left(\delta^{\rho}_{\mu}\partial_{\nu}\Omega + \delta^{\rho}_{\nu}\partial_{\mu}\Omega - \tilde{g}_{\mu\nu}\tilde{g}^{\rho\alpha}\partial_{\alpha}\Omega\right).$$

The Riemann tensor can straightforwardly be calculated through substitution:

$$\begin{aligned} R^{\rho}{}_{\sigma\mu\nu} = \tilde{R}^{\rho}{}_{\sigma\mu\nu} &+ \Omega^{-1}\left(\tilde{g}^{\rho\alpha}\tilde{g}_{\sigma[\mu}\delta^{\beta}_{\nu]} - \delta^{\rho}_{[\mu}\delta^{\alpha}_{\nu]}\delta^{\beta}_{\sigma}\right)\tilde{\nabla}_{\alpha}\tilde{\nabla}_{\beta}\Omega \\ &+ \tfrac{1}{2}\Omega^{-2}\left(3\delta^{\rho}_{[\mu}\delta^{\alpha}_{\nu]}\delta^{\beta}_{\sigma} - 3\tilde{g}_{\sigma[\mu}\delta^{\alpha}_{\nu]}\tilde{g}^{\rho\beta} + \delta^{\rho}_{[\mu}\tilde{g}_{\nu]\sigma}\tilde{g}^{\alpha\beta}\right)\tilde{\nabla}_{\alpha}\Omega\tilde{\nabla}_{\beta}\Omega. \end{aligned} \tag{3.7}$$

The Ricci tensor is then obtained from **(3.7)** by contraction of the first and third indices

$$R_{\mu\nu} = \tilde{R}_{\mu\nu} - \Omega^{-1}\left[\delta^{\alpha}_{\mu}\delta^{\beta}_{\nu} + \tfrac{1}{2}\tilde{g}_{\mu\nu}\tilde{g}^{\alpha\beta}\right]\tilde{\nabla}_{\alpha}\tilde{\nabla}_{\beta}\Omega + \tfrac{3}{4}\Omega^{-2}\delta^{\alpha}_{\mu}\delta^{\beta}_{\nu}\tilde{\nabla}_{\alpha}\Omega\tilde{\nabla}_{\beta}\Omega. \tag{3.8}$$

Finally, the Ricci scalar is found from the trace of (3.8):

$$R = \Omega^{-1}\tilde{R} - 3\Omega^{-2}\tilde{g}^{\alpha\beta}\tilde{\nabla}_\alpha\tilde{\nabla}_\beta\Omega + \tfrac{3}{2}\Omega^{-3}\tilde{g}^{\alpha\beta}\tilde{\nabla}_\alpha\Omega\tilde{\nabla}_\beta\Omega. \tag{3.9}$$

The transformations (3.6) and (3.9) will be used in combination with a non-linear field redefinition in order to formulate (3.2) in terms of a theory that resembles the original Einstein-Hilbert action (2.1).

3.4 FIELD PARAMETRISATIONS OF SCALAR-TENSOR THEORIES

The action (3.2), expressed in terms of the fields $(g_{\mu\nu}, \phi)$ features a non-minimal coupling to gravity. It can be brought into a form which resembles that of a scalar field minimally coupled to Einstein-Hilbert gravity, by a non-linear field redefinition $(g, \phi) \to (\tilde{g}, \tilde{\phi})$. Comparison of (3.2) with (2.1) and (3.9) shows that, for the gravitational sector, it is sufficient to perform a conformal transformation (3.5) $g_{\mu\nu} \to \tilde{g}_{\mu\nu}$ of the metric field. The conformal factor is given by

$$\Omega = \kappa U^{-1}. \tag{3.10}$$

The Ricci scalar transforms according to (3.9). Substitution of the transformed Ricci scalar into the JF action (3.2) then results in

$$S[\tilde{g}, \phi] = \int \left(\kappa\tilde{R} - \frac{1}{2}\frac{\kappa}{U}\frac{GU + 3U_1^2}{U}\tilde{\nabla}_\mu\phi\tilde{\nabla}^\mu\phi - \tilde{V} \right)\sqrt{-\tilde{g}}\,\mathrm{d}^4X,$$

where the EF potential $\tilde{V}$ is defined as

$$\tilde{V} \equiv \kappa^2 U^{-2} V = \kappa^2 W. \tag{3.11}$$

A particular feature of W is therefore that it is invariant under frame reparametrisations: $W(\phi) = \tilde{W}(\tilde{\phi})$. Tensors in the EF will always be indicated by a tilde. Indices of EF tensors are raised or lowered with respect to the metric $\tilde{g}_{\alpha\beta}$, and covariant derivatives $\tilde{\nabla}_\mu$ are defined with respect to the Christoffel symbol of the metric $\tilde{g}_{\alpha\beta}$. The kinetic term can be brought into the canonical form via the redefinition $\phi \to \tilde{\phi}$ of the JF scalar field ϕ to the EF scalar field $\tilde{\phi}$, that follows from the differential relation

$$\left(\frac{\partial\tilde{\phi}}{\partial\phi}\right)^2 = \frac{\kappa}{U}\frac{GU + 3U_1^2}{U} = \left(\frac{Us}{\kappa}\right)^{-1}, \tag{3.12}$$

where the last equality follows from (3.3). In terms of the EF parametrisation $(\tilde{g}, \tilde{\phi})$ the action (3.2) reads

$$S[\tilde{g}, \tilde{\phi}] = \int \left(\tfrac{1}{2}M_\mathrm{P}^2\tilde{R} - \tfrac{1}{2}\tilde{\nabla}_\mu\tilde{\phi}\tilde{\nabla}^\mu\tilde{\phi} - \tilde{V} \right)\sqrt{-\tilde{g}}\,\mathrm{d}^4X. \tag{3.13}$$

In the last step, the constant $\kappa = \frac{1}{2}M_\mathrm{P}^2$ was again identified with half the square of the Planck mass.

With the formalism of scalar-tensor theories in place it is instructive to consider two important cases. The first identifies the inflaton field with the standard model Higgs field, the second identifies it with the scalaron field of $f(R)$ theories. Their relevance becomes apparent within the formalism of inflationary cosmology, which will be discussed in **chapter 4**. Both models lead to almost indistinguishable predictions for the inflationary spectral observables [41–44]. They are both representatives of a larger class of inflationary attractors [45,46].

3.5.1 *Higgs inflation*

Higgs inflation proposes that the standard model Higgs field is the inflaton [37, 38, 41, 43, 47–49]. This model has several appealing characteristics. First, there is no need to postulate the inflaton field as a separate scalar field. This allows the standard model of particle physics to be combined with the standard model of cosmology into a single field theory that can be extended up to the Planck scale. Second, it obviates the above-mentioned problem that if the Higgs field were to be minimally coupled to gravity, renormalisation effects would induce a non-minimal coupling.

The need for Higgs inflation to be a scalar-tensor theory follows from inflationary reasoning. The inflationary power spectra are proportional to the conformal potential W, which in the case of a minimally coupled theory does not correctly describe the spectral amplitude. The model of Higgs inflation is described by the action

$$S_{\mathrm{H}} = \int_{\mathcal{M}} \left(\tfrac{1}{2} M_{\mathrm{P}}^2 R + \xi H^{\dagger} H R - \mathrm{D}_{\mu} H^{\dagger} \, \mathrm{D}^{\mu} H - \lambda \left(H^{\dagger} H - \tfrac{1}{2} v^2 \right)^2 \right) \sqrt{-g} \, \mathrm{d}^4 X.$$

Here, H is the complex Higgs field, λ the tree-level quartic Higgs self-coupling and $v \approx 246\,\mathrm{GeV}$ the electroweak symmetry breaking scale. The magnitude of the non-minimal coupling $\xi \approx 10^4$ follows from the observational constraints placed by the inflationary power spectra [49]. The coupling between the Higgs field and the gauge bosons is realised by the gauge covariant derivative

$$\mathrm{D}_{\mu} H = (\nabla_{\mu} - \mathrm{i} g A_{\mu}^{a} \tau^{a} - \tfrac{1}{2} \mathrm{i} g' B_{\mu}) H,$$

where A_{μ}^{a} and B_{μ} are the $SU(2)$ and $U(1)$ gauge bosons with the respective coupling constants g and g', and τ^{a} are the generators of $SU(2)$.

The Higgs field is an $SU(2)$ doublet, and hence comprises two complex (or four real) fields. The connection to single-field scalar tensor theories is made once one decomposes H in terms of four real fields ϕ and α_a:

$$H = \tfrac{1}{\sqrt{2}} \exp(\mathrm{i} \tau^{a} \alpha_{a}) (0, \phi)^{T}.$$

The Goldstone fields α_a can be removed from the formalism by the adoption of the *unitary gauge*. The action can then be expressed in terms of ϕ and the massive standard model gauge bosons

$$W_\mu^\pm \equiv \tfrac{1}{\sqrt{2}}\left(A_\mu^1 \mp \mathrm{i}A_\mu^2\right), \qquad Z_\mu^0 \equiv \left(g^2 + g'^2\right)^{-1/2}\left(gA_\mu^3 - g'B_\mu\right).$$

The result is

$$S_\mathrm{H} = \int\limits_{\mathcal{M}} \left(\tfrac{1}{2}M_\mathrm{P}^2 R + \tfrac{1}{2}\xi R\phi^2 - \tfrac{1}{2}\nabla_\mu\phi\nabla^\mu\phi - V(\phi)\right)\sqrt{-g}\,\mathrm{d}^4X + S_\mathrm{m},$$

where the action S_m accounts for the mass terms of the gauge bosons:

$$S_\mathrm{m} = -\tfrac{1}{2}\int\limits_{\mathcal{M}} v^{-2}\phi^2 g^{\mu\nu}\left(m_W^2 W_\mu^+ W_\nu^- + m_Z^2 Z_\mu^0 Z_\nu^0\right)\sqrt{-g}\,\mathrm{d}^4X.$$

The masses are explicitly given as $m_W^2 = \tfrac{1}{2}v^2g^2$ and $m_Z^2 = \tfrac{1}{2}v^2(g^2+g'^2)$. The model functions can simply be read off from this action:

$$U(\phi) = \tfrac{1}{2}M_\mathrm{P}^2 + \tfrac{1}{2}\xi\phi^2, \qquad G(\phi) = 1, \qquad V(\phi) = \tfrac{1}{4}\lambda\left(\phi^2 - v^2\right)^2.$$

The model described above neglects the Higgs coupling to the fermion fields of the standard model, but it is sufficient to describe the caveats associated with the connection between particle physics and cosmology. It will be shown in **chapter 4** that the inflationary epoch depends on the flatness of the inflaton potential. It is important that this flatness is not spoiled by quantum corrections. There are two sources of quantum corrections:

I. radiative corrections that originate from the standard model fields (in this example limited to the gauge bosons),

II. renormalisation group running of the model functions U, G and V.

For Higgs inflation, these radiative corrections and the renormalisation group improvement turned out to be important for the consistency with particle physics experiments. A discussion of these analyses goes beyond the scope of this work. Instead, the interested reader is referred to the investigations of references [37,38,41,43,48–51].

3.5.2 *$f(R)$-theory as a scalar-tensor theory*

Scalar-tensor theories form a quite general modification of general relativity. One of the most successful inflationary models, $f(R)$ theory, can actually be realised as a scalar-tensor theory [23]. Cosmological models such as $f(R)$ theory considers the Einstein-Hilbert action as the low-curvature limit of a general functional of the Ricci scalar [24,53,54]:

The equivalence of scalar-tensor theories and $f(R)$ theories holds for classical theories. An investigation of perturbatively quantised $f(R)$ theories is given in [52].

$$S_{f(R)} = \int f(R)\sqrt{-g}\,\mathrm{d}^4X. \tag{3.14}$$

From the variation of this action the field equations can be found to be

$$f(R)R_{\mu\nu} - \tfrac{1}{2}f(R)g_{\mu\nu} + \left(g_{\mu\nu}\Box - \nabla_\mu\nabla_\nu\right)f(R) = T_{\mu\nu}.$$

In contrast to general relativity, $f(R)$ theory yields equations of motion which are up to fourth order. It will be shown in the next section that this implies that these theories have an additional scalar degree of freedom, which can be identified with the inflaton [55].

The action **(3.14)** can be formulated as a scalar-tensor theory by the introduction of an auxiliary field α via

$$S_{f(R)} = \int [f(\alpha) + f_1(\alpha)(R - \alpha)] \sqrt{-g}\, \mathrm{d}^4 X. \tag{3.15}$$

Recall that $f_1(\alpha) = \partial f / \partial \alpha$. The variation of **(3.15)** with respect to α results in the equation of motion $\alpha = R$, provided f_2 does not vanish. Therefore **(3.15)** is equivalent to **(3.14)**. With the assumption that f_1 is invertible, the inflaton field can be defined by the relation

If $f_1 < 0$, one instead defines the left-hand side of **(3.16)** to be $-\phi^2$.

$$\phi^2 = 2 f_1(\alpha). \tag{3.16}$$

This differential relation allows the auxiliary field α to be expressed in terms of ϕ. The action **(3.15)** can then be formulated as

$$S_{f(R)} = \int \left[\tfrac{1}{2}\phi^2 R - V(\phi)\right] \sqrt{-g}\, \mathrm{d}^4 X,$$

where the inflaton potential is given by

$$V(\phi) = \tfrac{1}{2}\phi^2\, \alpha(\phi) - f(\alpha(\phi)). \tag{3.17}$$

A peculiar feature of these scalar-tensor theories is that the JF can be identified by the absence of a kinetic term. This is a consequence of the origin of ϕ as a function of the non-dynamical auxiliary field α.

To see how this works in practice it is instructive to consider what a reasonable $f(R)$ theory might be. Any such theory must predict general relativity in some limit. As the universe appears to be flat on cosmological scales it is natural to consider a series expansion of the Ricci scalar R:

$$f(R) = c_0 + c_1 R + c_2 R^2. \tag{3.18}$$

The correspondence with general relativity in the low-curvature limit fixes the proportionality constant $c_0 = -M_\mathrm{P}^2 \Lambda$ in the first term to be a cosmological constant, while the proportionality constant $c_1 = \frac{1}{2} M_\mathrm{P}^2$ in the second term must be half the square of the Planck mass. The coupling constant in the second term c_2 must be determined through observations.

One motivation for the model is the observation that renormalisation effects induce a term that is quadratic in R.

The model described in **(3.18)** is called *Starobinsky inflation* [54], and was one of the earliest models of cosmological inflation. The model functions for Starobinsky inflation are

$$U(\phi) = \tfrac{1}{2}\phi^2, \qquad G(\phi) = 0, \qquad V(\phi) = (16c)^{-1} M_\mathrm{P}^2 \left(1 - \frac{\phi^2}{M_\mathrm{P}^2}\right)^2.$$

The above discussion shows that there is no preferred frame *a priori*, and all frames are classically equivalent. Given that it is always possible to perform a conformal transformation to transition from the JF to the EF, it seems natural to wonder whether it is necessary to consider the JF in the first place. After all, once the model functions in **(3.2)** have been specified, one is free to transform to the EF, and subsequently perform all calculations there.

However, frame transformations are not guaranteed to commute with quantisation procedures. This is investigated in reference [52].

The answer to this question is one mostly of convenience, as, for a given set of model functions, it is not guaranteed that the scalar field ϕ can be expressed as a function of $\tilde{\phi}$ in a closed form. Therefore, an analytical treatment of the EF is not always possible or practical.

A more fundamental objection arises once one considers the imprint of perturbations in an inflationary context. The expansion of the universe during its inflationary epoch should not continue overly long if it is to correctly describe the formation of large-scale structure that is observed in the universe today. Since the expansion of space is inextricably tied to the strength of the gravitational field, it is generally a function of the scalar field and therefore frame-dependent. The conditions that define the end of inflation in its conventional formulation must therefore be carefully specified, for otherwise one would make frame-dependent predictions. Once these conditions are put into place, it is more natural to work with the JF. This will be illustrated in **section 4.5.5**.

4 COSMOLOGY

Cosmology provides a systematic study of the structure and evolution of the universe. The theoretical description of modern cosmology rests on the formulation of general relativity. Since its first presentation in 1915, general relativity has caused an explosive growth in the understanding of the structure of the cosmos. It forms the basis for the six-parameter *Lambda cold dark matter* (ΛCDM) model. This is the simplest theory that provides an understanding of, among others, the physics that followed shortly after the hot Big Bang origin of the universe, the *cosmic microwave background* (CMB) radiation, the large-scale structure of the distribution of galaxies, the abundances of hydrogen, helium and lithium, and the accelerated expansion of the universe.

Despite its empirical success, the bare ΛCDM model is left with several unanswered questions. For example, it does not provide an explanation for the apparent flatness of the universe. Although it is only fairly recently that the curvature of the universe at cosmological scales has been reliably measured, the equations of motion for the gravitational field suggest that the degree of flatness that is observed today would require very precise initial conditions of the cosmological parameters. There is no physical principle that predicts why this should be so. The resolution to this paradox is, as shall be seen, provided by cosmic inflation.

This chapter gives an overview of the conventional approach to cosmology. After that, it provides a description of cosmic inflation as an extension of the ΛCDM model in the specific context of cosmological scalar-tensor theories. These themselves are modifications—or, depending on the point of view, extensions—of general relativity. The reader is invited to consult the extensive references on the subject, such as references [14,56–60].

4.1 THE COSMOLOGICAL PRINCIPLE

In **chapter 2**, the cosmological principle was invoked in order to reduce the line-element to the simple form

$$\mathrm{d}s^2 = -N^2(t)\mathrm{d}t^2 + a^2(t)\gamma_{ab}\mathrm{d}x^a\mathrm{d}x^b,$$

where γ_{ab} is a homogeneous and isotropic spatial metric. Again, homogeneity in spacetime implies that γ_{ab} can only be a function of the spatial coordinates. The condition that the spatial metric should be homogeneous and isotropic in space places a strong restriction on what the geometry of space can be. There are just three possible geometries that have this kind of symmetry:

I. flat space,

II. spaces of constant positive curvature,

III. spaces of constant negative curvature.

In all these cases the spatial metric can be derived from the case of constant positive curvature, where space has the topology of a 3-dimensional hypersphere of radius a. The embedding of this hypersphere in a 4-dimensional Euclidean space results in the familiar constraint equation

$$\delta_{ab}x^a x^b + z^2 = a^2, \tag{4.1}$$

where z is an auxiliary variable. Differentiation of this relation results in the constraint

$$z\,\mathrm{d}z = -\delta_{ab}x^a \mathrm{d}x^b.$$

A spatial length interval can then be obtained by insertion of this constraint into the Euclidean metric:

$$\mathrm{d}\ell^2 = \delta_{ab}\mathrm{d}x^a\mathrm{d}x^b + \frac{\delta_{ab}\delta_{cd}x^a x^c \mathrm{d}x^b \mathrm{d}x^d}{a^2 - \delta_{ab}x^a x^b}. \tag{4.2}$$

This relation expresses the distance between two spatial points entirely by the three independent coordinates x^a. However, a given set of coordinates (x^1, x^2, x^3) corresponds to two points on the hypersphere. The metric in this representation is therefore degenerate. A more convenient representation is given in terms of (hyper)spherical coordinates, where the x^a are parametrised by the radial distance r and the two angles θ and ϕ:

$$\begin{aligned} x^1 &= r\sin\theta\cos\phi, \\ x^2 &= r\sin\theta\sin\phi, \\ x^3 &= r\cos\theta. \end{aligned}$$

The equation for the hypersphere in this coordinate system takes the form

$$\delta_{ab}x^a\,\mathrm{d}x^b = r\,\mathrm{d}r, \qquad \delta_{ab}\mathrm{d}x^a\mathrm{d}x^b = \mathrm{d}r^2 + r^2\mathrm{d}\Omega_2^2,$$

where $\mathrm{d}\Omega_2$ is the metric on the unit sphere. The spatial metric (4.2) then takes the form

Spaces with constant negative curvature are given by the 3-dimensional pseudosphere, which cannot be embedded in Euclidean space.

$$\mathrm{d}\ell^2 = \frac{\mathrm{d}r^2}{1 - a^{-2}r^2} + r^2\mathrm{d}\Omega_2^2.$$

The metric for flat space can be obtained from (4.2) by the limit $a \to \infty$, while the metric for spaces of constant negative curvature can be obtained by the formal replacement $a \to ib$. The spatial metric can be brought in a form compatible with (2.26) after the rescaling $r = |a|r$, after which one obtains

$$\mathrm{d}\ell^2 = |a|^2\left(\frac{\mathrm{d}r^2}{1 - \varkappa r^2} + r^2\mathrm{d}\Omega_2^2\right), \tag{4.3}$$

The pseudosphere is also known as the hyperboloid.

where $\varkappa$ is the normalised curvature of the spatial manifold. It takes the values $\varkappa = 0$ for the plane, $\varkappa = 1$ for the sphere and $\varkappa = -1$ for the pseudosphere. After insertion of this expression into **(2.26)**, one obtains the line element for a *Friedmann-Lemaître-Robertson-Walker universe* (FLRW).

The coordinate system **(4.3)** on the hypersphere covers half of space. This can be seen from **(4.1)**, which has two solutions for the auxiliary variable z. It is often desirable to express the line element in a coordinate system that covers the whole of space. This can be done by the introduction of a radial coordinate χ, that is the solution to the differential equation

$$\left(\frac{\mathrm{d}\chi}{\mathrm{d}r}\right)^2 = \frac{1}{1-\varkappa r^2}. \tag{4.4}$$

The geometric interpretation of this new radial coordinate is shown in **figure 4.1** for the case of a two-dimensional sphere and pseudosphere.

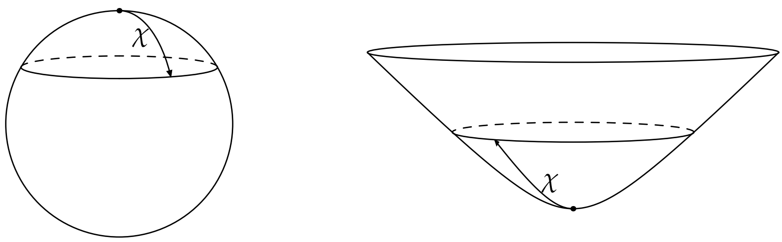

Figure 4.1 Geometric meaning of χ for curved spaces.

The line element **(4.3)** then takes the form

$$\mathrm{d}\ell^2 = |a|^2\left(\mathrm{d}\chi^2 + f(\chi)^2\mathrm{d}\Omega_2^2\right), \qquad f(\chi) = \begin{cases} \sinh\chi & \text{for } \varkappa = -1, \\ \chi & \text{for } \varkappa = 0, \\ \sin\chi & \text{for } \varkappa = 1. \end{cases} \tag{4.5}$$

4.2 THE FLRW UNIVERSE

The combination of **(2.26)** and **(4.3)** leads to the line element for the homogeneous and isotropic universes:

$$\mathrm{d}s^2 = -N^2\mathrm{d}t^2 + a^2\gamma_{ab}\mathrm{d}x^a\mathrm{d}x^b, \tag{4.6}$$

where the components of γ_{ab} can be read off from **(4.3)**. The non-vanishing components of the Christoffel symbol for this spacetime can be computed to be

$$\Gamma^0_{00} = \mathrm{D}_t N, \qquad \Gamma^0_{ab} = \frac{a^2}{N}H\gamma_{ab},$$
$$\Gamma^a_{0b} = NH\delta^a_b, \qquad \Gamma^a_{bc} = \Gamma^{\mathrm{s}a}{}_{bc}.$$

Here, D_t is the reparametrisation invariant time derivative that was introduced in **chapter 2**, $H = \mathrm{D}_t a/a$ is the Hubble parameter and $\Gamma^{\mathrm{s}a}{}_{bc}$ are the components of the Christoffel symbol calculated from the spatial metric γ_{ab}. The non-vanishing components of the Riemann tensor are

$$R_{0a0b} = -N^2H^2\left(1+\frac{\mathrm{D}_t H}{H^2}\right)g_{ab}, \qquad R_{abcd} = a^{2\,(\mathrm{s})}R_{abcd} + 2H^2 g_{a[c}g_{d]b}.$$

${}^{(\mathrm{s})}R_{abcd}$ is the Riemann tensor calculated from γ_{ab}. The Ricci tensor can be found by contraction of the Riemann tensor with the metric. The non-vanishing components are

$$R_{00} = -3N^2H^2\left(1+\frac{\mathrm{D}_t H}{H^2}\right), \qquad R_{ab} = {}^{(\mathrm{s})}R_{ab} + H^2\left(3+\frac{\mathrm{D}_t H}{H^2}\right)g_{ab}.$$

Finally, the Ricci scalar is

It will be convenient to express later results in terms of the normalised spatial curvature $\varkappa = \frac{1}{6}{}^{(\mathrm{s})}R$.

$$R = a^{-2\,(\mathrm{s})}R + 12H^2\left(1+\frac{1}{2}\frac{\mathrm{D}_t H}{H^2}\right).$$

Meanwhile, the energy-momentum tensor describes a perfect fluid:

$$T_{\mu\nu} = (\rho + p)\, u_\mu u_\nu + p\, g_{\mu\nu}. \tag{4.7}$$

The four-velocity u_μ is normalised such that $u_\mu u^\mu = -1$, and, in the context of the ADM decomposition, can be identified with the normal vector n_μ. Insertion of these relations into the Einstein field equations (2.2) results in the *Friedmann equations*

$$H^2 = \tfrac{1}{6}U^{-1}(\rho - a^{-2}U^{(\mathrm{s})}R),$$
$$\mathrm{D}_t H = -\tfrac{1}{4}U^{-1}(\rho + p - \tfrac{2}{3}a^{-2}U^{(\mathrm{s})}R).$$

From the Friedmann equations one can derive the following useful identity:

$$3H^2 + \mathrm{D}_t H = \tfrac{1}{4}U^{-1}\left(\rho - p - \tfrac{2}{3}a^{-2}U^{(\mathrm{s})}R\right). \tag{4.8}$$

Finally, the Friedmann equations lead to the continuity equation

$$\mathrm{D}_t\left(U^{-1}\rho\right) + 3HU^{-1}(\rho + p) = 0.$$

The continuity equation is a result of the Einstein equations and the Bianchi identity. In the case that U is constant it can be obtained from the time component of the conservation of the energy-momentum tensor (4.7).

4.2.1 *Time reparametrisations*

The Friedmann equations imply that the only dynamical degree of freedom in the FLRW metric is the scale factor. The lapse function becomes non-dynamical. The time derivative D_t is invariant under time reparametrisations, and it is therefore frequently convenient to pick parametrisations (gauges) that differ from the coordinate time t. These gauges correspond to different prescriptions of the lapse function.

The simplest possible time parametrisation that is often encountered in the literature corresponds to the gauge $N = 1$. This defines *Friedmann time*.

Another gauge, that turns out to be useful for inflationary cosmology, is the gauge $N = H^{-1}$. It will be shown in **section 4.5** that this gauge measures time in terms of the number of e-folds the scale factor has increased with respect to some reference value. For this reason, this time parametrisation will be denoted by N_e. It is a useful time gauge when one wants to solve the Friedmann equations numerically.

Another gauge that is particularly useful for cosmological purposes is *conformal time* τ, for which $N = a$. When expressed in terms of conformal time, the line element **(4.6)** takes the form

$$\mathrm{d}s^2 = a^2(\tau)\left(-\mathrm{d}\tau^2 + \gamma_{ab}\,\mathrm{d}x^a\mathrm{d}x^b\right).$$

The usefulness of conformal time is apparent when one considers lightlike geodesics, for which the line element vanishes: $\mathrm{d}s^2 = 0$. From the symmetry of the FLRW universe it follows that $\mathrm{d}\Omega_2 = 0$, and therefore

$$\mathrm{d}s^2 = a^2(\tau)\left(-\mathrm{d}\tau^2 + \mathrm{d}\chi^2\right),$$

where χ is the radial coordinate defined in **(4.4)**. It then follows that lightlike geodesics trace out straight lines in the (τ, χ) plane that travel at an angle of $\pi/4$.

4.3 DE SITTER SPACE

One of the most important discoveries of the previous century is that space expands at an accelerated rate [61,62]. Observations similarly indicate that an increased rate of expansion also took place during the early universe. In order to study spacetimes that have this property it is useful to temporarily extend the conditions imposed on spacetime by the cosmological principle. In particular, in this section the cosmological principle is considered to be a symmetry of both space and time, rather than space alone. Such a space is called a *De Sitter space*, and has a special role in inflationary cosmology. Due to its homogeneity and isotropy it has a constant curvature. Its important properties are summarised here.

De Sitter space is a solution to the Einstein field equations **(2.1)** with the inclusion of the cosmological constant. The cosmological constant can be considered as a vacuum contribution to the energy density; its energy-momentum tensor is

Alternatively, De Sitter space can be defined as the embedding of a 4-dimensional hypersphere in 5-dimensional Minkowski space, analogous to the procedure in **section 4.1**.

$$T^{\Lambda}_{\mu\nu} = -\Lambda g_{\mu\nu}.$$

Comparison with the energy-momentum tensor of a perfect fluid **(4.7)** leads to the conclusion that a vacuum energy density can be described as a perfect fluid of which the energy density and momentum are

$$\rho^{\Lambda} = -p^{\Lambda} = \Lambda.$$

Substitution of these expressions into the continuity equation leads to the conclusion that ρ^{Λ}/U is constant. It then follows from the Friedmann equations that

$$\mathrm{D}_t H + H^2 = \frac{\mathrm{D}_t{}^2 a}{a} = \frac{\rho^{\Lambda}/U}{6} \equiv H_{\Lambda}^2.$$

The solution to this equation is

$$a = a_1 \exp(\int H_\Lambda N \mathrm{d}t) + a_2 \exp(-\int H_\Lambda N \mathrm{d}t), \tag{4.9}$$

where a_1 and a_2 are constants of integration. They can be fixed by insertion of **(4.9)** into the first Friedmann equation, after which one obtains

$$4H_\Lambda^2 a_1 a_2 = \varkappa. \tag{4.10}$$

Of particular interest here is flat space, for which the right-hand side vanishes. In that case, one must set a_2 to zero in order to obtain a solution that corresponds to an expanding space. Notice that this feature is general; any non-zero curvature will require that $a_1 \neq 0$. The scale factor will therefore always increase approximately exponentially, given enough time. Substitution of the solution of **(4.10)** into the line-element yields

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + H_\Lambda^{-2} \begin{pmatrix} \sinh^2(H_\Lambda t) \\ \exp(2H_\Lambda t) \\ \cosh^2(H_\Lambda t) \end{pmatrix} \left[\mathrm{d}\chi^2 + \begin{pmatrix} \sinh^2\chi \\ \chi^2 \\ \sin^2\chi \end{pmatrix} \mathrm{d}\Omega_2^2 \right] \qquad \begin{matrix} \varkappa = -1, \\ \varkappa = \ \ 0, \\ \varkappa = \ \ 1, \end{matrix}$$

where for convenience the lapse function has been set to 1. The three different solutions each represent a section of the same spacetime, as one can be transformed into another via a suitable coordinate transformation.

4.4 EVOLUTION OF THE UNIVERSE

The Friedmann equations and the continuity equation provide three equations in order to determine four unknowns. Therefore, in order to obtain a unique solution for these equations one would have to provide a fourth equation. This additional equation relates the energy density ρ and the pressure p in an *equation of state*:

$$\rho = wp. \tag{4.11}$$

The equation of state is enough to express the energy density ρ as a function of the scale factor. After substitution of the equation of state in the continuity equation it is straightforward to verify that

$$\rho(a) \propto a^{-3(1+w)} \tag{4.12}$$

is a solution. The solution **(4.12)** can be inserted into the first Friedmann equation to yield a differential equation for the scale factor alone. This fixes, for a given equation of state, the behaviour of the scale factor as a function of time:

$$a(t) \propto t^{\frac{2}{3}(1+w)^{-1}}.$$

An exception to these scaling behaviours is provided by De Sitter space, for which $w = -1$. The equations of state and some of their corresponding physical consequences have been gathered in **table 4.1**. In particular, it lists the *comoving particle horizon* d_H, which is defined as

$$d_{\mathrm{H}} = \int_{t_0}^{t} a^{-1}(t)\, N(t)\mathrm{d}t.$$

The particle horizon is the distance light could have travelled from some initial time t_0 until some time t. Particles that are at distances smaller than the particle horizon with respect to each other are said to be causally connected. Also of particular importance is the *comoving Hubble radius* $(aH)^{-1}$. The physical particle horizon and the physical Hubble radius are typically equal in cosmological models in which the *strong energy condition* holds, that is, if $\rho + 3p > 0$, and both typically increase in time.

The above discussion assumes that the energy density of the universe is completely determined by the energy density of a single kind of matter. If different kinds of matter are present, each would have its own equation of state and the total energy density is formed by their sum. The energy density for each kind of matter depends on the scale factor according to **(4.12)**. The behaviour of the scale factor can then be determined by substitution of the total energy density into the first Friedmann equation.

Table 4.1 Behaviour of the scale factor and associated quantities for different equation of state parameters.

	w	$\rho(a)$	$a(t)$	d_{H}	$(aH)^{-1}$
radiation	$\frac{1}{3}$	a^{-4}	$t^{\frac{1}{2}}$	$t^{\frac{1}{2}}$	$t^{\frac{1}{2}}$
matter	0	a^{-3}	$t^{\frac{2}{3}}$	$t^{\frac{1}{3}}$	$t^{\frac{1}{3}}$
curvature	$-\frac{1}{3}$	a^{-2}	t	$\log t$	constant
vacuum	-1	a^{0}	$\mathrm{e}^{H_\Lambda t}$	$\mathrm{e}^{-H_\Lambda t}$	$\mathrm{e}^{-H_\Lambda t}$

4.4.1 *Epochs of the universe*

As each kind of matter has its own characteristic dependence on the scale factor it is reasonable to expect that the universe went through different phases called *epochs*, in which the energy density was dominated by one particular kind of matter. In the very early universe, the energy was dominated by a constant vacuum energy, since the universe underwent a period of accelerated expansion. This exponential expansion was followed by a period in which the energy content was dominated by radiation, followed by a period in which the energy density was dominated by matter.

The first Friedmann equation serves as a constraint equation for the energy content of the universe. Division by H^2 results in

$$\Omega_{\mathrm{r}} + \Omega_{\mathrm{m}} + \Omega_{\varkappa} + \Omega_{\Lambda} = 1. \tag{4.13}$$

with $\Omega_i = \rho_i/\rho_0$ and the critical energy density defined as $\rho_0 \equiv 6UH^2$. At present, the universe has a small left-over vacuum energy density Λ_{de}, which is known as *dark energy*. The origin of this vacuum energy is unknown. Constructions that generate this energy density dynamically through the introduction of another scalar field are known as quintessence.

This section is concluded with the definition of the redshift z. Photons that are emitted at some time t are redshifted by the expansion of the universe when these photons are observed today. The redshift z quantifies this change as

$$1 + z \equiv \frac{a_{\text{today}}}{a(t)}.$$

One usually chooses a normalisation in which $a_{\text{today}} = 1$.

4.5 INFLATION

Although the bare ΛCDM model is successful in describing and explaining various properties of the universe at large, it still leaves various questions unanswered. Three of these are the so-called *flatness*, *horizon* and *monopole problems*. Inflation was first proposed independently by Starobinsky, Guth and Sato [54,63–65] and subsequently worked out by Linde, Albrecht, Steinhardt and others [66–68] in an attempt to find a resolution to these three problems. Since then, inflation has been developed as a cornerstone of cosmology, and successfully predicts the anisotropies in the CMB and the formation of large-scale structure out of small quantum fluctuations [69–75]. In this section the basic idea of inflation will be presented, why it solves the flatness and horizon problems (and the monopole problem as an extension), and how it is modelled through the dynamics of a scalar field. Furthermore, the slow-roll formalism of inflation will be introduced and used to derive the cosmological parameters for the quantum fluctuations that are observed in the CMB today.

4.5.1 *The cosmic microwave background radiation*

The conventional picture proposed by the ΛCDM model is that, in the early universe, the various kinds of energy density that comprise **(4.13)** were in thermal equilibrium with each other. Electrons, protons and photons were tightly coupled in an ionised plasma that filled the universe. The energy density dropped as the universe expanded, and as a result the universe cooled down. The electrons and protons were able to form hydrogen once the universe had cooled down sufficiently, after which the mean free path of the photons diverged to infinity. This moment of decoupling defines a surface in time that is measured on earth as the *surface of last scattering*, which fills the universe as the cosmic microwave background (CMB). The CMB was emitted at a redshift of $z \approx 1100$, which implies that its original temperature of $T_{\text{CMB}} \approx 3000\,\text{K}$ has been redshifted to its current measured value of $T \approx 2.73\,\text{K}$.

The energy density scales as a^{-4} for radiation. From the Stefan-Boltzmann law it then follows that T scales as a^{-1}.

The CMB is the most perfect realisation of a blackbody that is observed in nature, with temperature anisotropies of the relative order $\Delta T/T \approx 10^{-5}$.

4.5.2 *The flatness and horizon problems*

The flatness problem Observation of (4.13) leads to a peculiar puzzle. Observations indicate that the curvature contribution to the energy density today is negligibly small, and effectively absent. In the original ΛCDM model, the Big Bang origin of the universe was followed by an epoch where the energy content of the universe was dominated by radiation, followed by an epoch where it was dominated by matter. Comparison of (4.13) with table 4.1 then shows that, as the scale factor increases with time, the curvature term should dominate, such that $\Omega_\varkappa \approx 1$. However, this is not the case, as the curvature of the observable universe appears to be zero.

A universe where $\rho_\varkappa$ is the dominant contribution to the Friedmann equation is known as a *Milne universe*.

Since (4.13) predicts that any deviation from $\Omega_\varkappa = 1$ should increase with time, it appears as if the initial conditions of the universe were greatly fine-tuned for the curvature to be as flat as it is. This is known as the flatness problem.

As Ω_r and Ω_m decrease faster than $\Omega_\varkappa$.

The horizon problem It was mentioned before that the conventional *hot Big Bang* scenario of the ΛCDM model proposes that, after the Big Bang, the energy content of the universe was first dominated by radiation, and afterwards by matter. From table 4.1 it can be seen that, during the radiation and matter epochs, the particle horizon grew in direct proportion to a positive power of the scale factor. The result is that, since the Big Bang, the scale factor has increased monotonically. This, in turn, implies that length scales that enter the horizon today have been far outside of the horizon at the time that the surface of last scattering was emitted. This is sketched in figure 4.2.

That the CMB nevertheless has a near homogeneous temperature suggests that different patches in the universe were in thermal contact with each other during the time of decoupling while they were, in actuality, causally disconnected. This paradox is called the *horizon problem*.

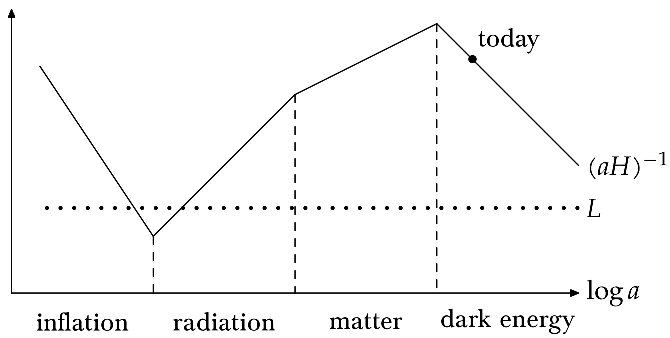

Figure 4.2 Evolution of the comoving length scales L over different epochs of the universe.

Resolution through inflation Both the horizon and flatness problem can be solved with the formalism of *cosmic inflation*. Inflation is a phase of accelerated expansion in the early universe.

From figure 4.2 it follows that the universe must have undergone an epoch during which the particle horizon, and indeed the Hubble horizon, shrank. Comparison with the Friedmann equations indicates that this yields the following condition on the evolution of the scale factor:

$$\mathrm{D}_t(aH)^{-1} < 0 \qquad \leftrightarrow \qquad \frac{\mathrm{D}_t{}^2 a}{a} > 0.$$

This is usually taken as the formal definition of inflation, and will be adopted in what follows.

From the Friedmann equations it follows that, in order for a period of inflation to be realised, the matter content of the universe must be dominated by a substance that violates the strong energy condition: $\rho + 3p < 0$. No known form of ordinary matter has this property. It can be seen from **table 4.1** that the only way to have an accelerated expansion of the scale factor is for the energy content of the universe to be dominated by vacuum energy.

To see how this solves the flatness and horizon problems, consider **figure 4.2** and **figure 4.3**. In the former figure it can be seen that the comoving length scales are constant in time (indeed, this is so by definition), but the dynamical Hubble horizon changes. In particular, it decreases during the inflationary epoch. Therefore, any given comoving length scale L is at some point entirely enclosed by the Hubble horizon, given enough time.

The flatness problem is resolved similarly by the paradigm of inflation. In **(4.13)** it can be seen that Ω_Λ grows during the inflationary epoch, as $\Omega_\varkappa$ decreases. After inflation ends, the vacuum decays into radiation and matter. This resulted in a universe that is dominated by radiation and matter, while the curvature is negligible.

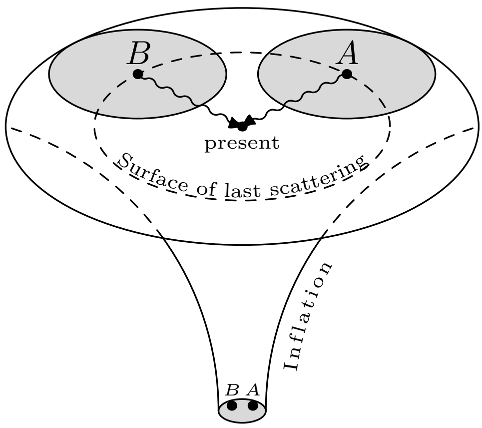

Figure 4.3 Inflation expands small uniform patches of space to seemingly causally disconnected distances.

If the model of inflation is correct, it is necessary to determine how long the inflationary epoch should last. If it does not last long enough it cannot resolve the horizon and flatness problems, but if it lasts too long it cannot explain the large-scale structure of the universe. A convenient parameter that quantifies the expansion of the universe is the number of e-folds N_e that the scale factor increased with respect to the start of inflation at time t_i. It is defined as

$$N_e = \log \frac{a_f}{a_i} = \int_{t_i}^{t_f} H N \mathrm{d}t, \tag{4.14}$$

with t_f the time at which inflation ends. The number of e-folds in most models of inflation that is necessary to describe the observed flatness of the universe is $N_e \approx 60$.

4.5.3 *Scalar-field inflation*

A simple method to implement inflation is relies on the introduction of a scalar field ϕ, called the *inflaton*. The energy density of the early universe is then assumed to be dominated by the energy density of the inflaton. In this way, the inflaton field is able to drive an accelerated expansion of space. A simple model is described by the action

$$S_\phi[\phi] = -\int \left(\tfrac{1}{2}\nabla_\mu\phi\nabla^\mu\phi + V(\phi)\right)\sqrt{-g}\,\mathrm{d}X^4. \tag{4.15}$$

From the homogeneity of the CMB it follows that ϕ itself is homogeneous, and is only a function of time. It will be shown later that, under certain conditions, the vacuum energy Λ is then provided by the potential $V(\phi)$. In this regime, where the energy density is dominated by the potential energy of the inflaton, the universe is approximately a De Sitter universe.

The presence of structure, both in the large and small scales, indicate that inflation must at some point come to an end. This is accommodated by the fact that the inflaton field is dynamical, and as a result the vacuum energy will change with time. The early universe therefore cannot be a perfect De Sitter universe. Deviations from the De Sitter universe form the basis of the *slow-roll approximation*. In this regime, the potential of the inflaton field can be said to be flat, such that the inflaton slowly rolls down the potential.

The basis of this section will be an extension of the minimally coupled action in **(4.15)**. The inflaton will be described by a generic scalar-tensor theory, as described in **chapter 3**. This has two advantages:

I. Current observations slightly favour models that exhibit a non-minimal coupling.

II. A non-minimal coupling could potentially give insights into the fundamental nature of the inflaton field.

In recent years a plethora of models has been developed for scalar-tensor theories [76], but the results in this work will be quite general.

The energy-momentum tensor for a homogeneous inflaton field takes the simple form

$$T^\phi_{\mu\nu} = \rho^\phi n_\mu n_\nu + p^\phi \gamma_{\mu\nu}.$$

The normal vector is defined as $n_\mu = -\nabla_\mu\phi/\sqrt{-\nabla_\mu\phi\nabla^\mu\phi}$. The energy density ρ^ϕ and pressure p^ϕ for a scalar-tensor theory are

$$\rho^\phi = \tfrac{1}{2}G(\mathrm{D}_t\,\phi)^2 + V - 6H\,\mathrm{D}_t\,U,$$
$$p^\phi = \tfrac{1}{2}G(\mathrm{D}_t\,\phi)^2 - V + 4H\,\mathrm{D}_t\,U + 2\,\mathrm{D}_t{}^2\,U.$$

These two quantities appear in the Friedmann equations. Thus, a scalar-tensor theory is able to produce the condition where the strong energy condition is violated, and subsequently drive an accelerated expansion of space. Additionally, note that it is not just the strong energy condition that is violated in scalar-tensor theories, since the energy density is not necessarily positive. Scalar-tensor theories therefore potentially violate the *weak energy condition.*

4.5.4 *The slow-roll approximation*

In order to have a successful phase of inflation the potential of the inflaton must dominate its energy density. Paradoxically, this does not necessarily imply that the scalar field moves slowly through its potential, just that the change of the potential relative to the potential itself is small. This is sketched in **figure 4.4**. Once the inflaton reaches the minimum of its potential it triggers a period of reheating, where it loses energy due to friction. This then leads to the production of radiation and matter that is observed today.

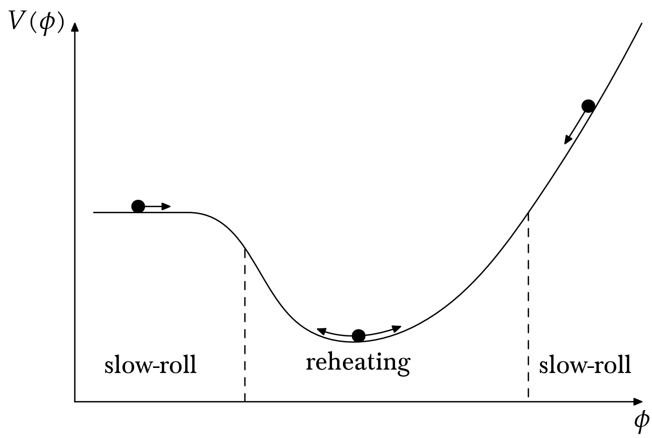

Figure 4.4 Typical slow-roll regimes for a given potential in inflation.

The slow-roll approximation for scalar-tensor theories can be implemented by the condition that, for any smooth function f of ϕ, the following inequality holds [77]:

$$\mathrm{D}_t{}^2 f \ll H\,\mathrm{D}_t f \ll H^2 f.$$

One recovers the usual notion of slow-roll inflation when f is the identity [78]:

$$\mathrm{D}_t{}^2\,\phi \ll H\,\mathrm{D}_t\,\phi \ll H^2\phi.$$

The slow-roll condition and the equations of motion motivate the definition of the following quantities:

$$\varepsilon_{1,\mathrm{H}} = -\frac{\mathrm{D}_t\,H}{H^2}, \qquad \varepsilon_{2,\mathrm{H}} = -\frac{\mathrm{D}_t{}^2\,\phi}{H\,\mathrm{D}_t\,\phi}, \tag{4.16.a}$$

$$\varepsilon_{3,\mathrm{H}} = \frac{1}{2}\frac{\mathrm{D}_t\,U}{HU}, \qquad \varepsilon_{4,\mathrm{H}} = \frac{1}{2}\frac{\mathrm{D}_t\,s}{Hs}, \qquad \varepsilon_{5,\mathrm{H}} = \frac{1}{2}\frac{\mathrm{D}_t{}^2\,U}{H^2U}. \tag{4.16.b}$$

The first two quantities describe the dynamics of the Hubble parameter and the inflaton, while the remaining quantities describe the effects of the model function U, s and V. Taken together, they define the *Hubble slow-roll formalism*, as they completely determine the dynamics of the Hubble parameter and the inflaton field. The Klein-Gordon equation, the second Friedmann equation and equation **(4.8)** can respectively be written as

$$-\frac{sU^2W_1}{3H\,\mathrm{D}_t\,\phi} = 1 - \tfrac{1}{3}\varepsilon_{2,H} + \tfrac{1}{3}\varepsilon_{3,H} - \tfrac{1}{3}\varepsilon_{4,H}, \tag{4.17}$$

$$\frac{(\mathrm{D}_t\,\phi)^2}{Us} = 4\left[\varepsilon_{1,H} + \varepsilon_{3,H} - \varepsilon_{5,H} + 3\varepsilon_{3,H}^2\right]H^2, \tag{4.18}$$

$$UW = 6(1 - \tfrac{1}{3}\varepsilon_{1,H} + \tfrac{5}{3}\varepsilon_{3,H} + \tfrac{1}{3}\varepsilon_{5,H})H^2. \tag{4.19}$$

The spatial curvature has been set to zero, as inflation generically flattens any spacetime. During the inflationary epoch it is necessary, not only for the slow-roll parameters to be small, but to remain small for a sufficiently long time. It is therefore useful to introduce a hierarchy of slow-roll parameters as follows:

$$\varepsilon_{j,H}^{(i+1)} = \frac{\mathrm{D}_t\,\varepsilon_{j,H}^{(i)}}{H\varepsilon_{j,H}^{(i)}}, \qquad i = 0, 1, \ldots, \qquad j = 1, 2, 3, 4, \tag{4.20}$$

with $\varepsilon_{j,H}^{(0)}$ the slow-roll parameters defined in **(4.16)**. In this manner one can extend the slow-roll approximation to the *slow-roll expansion* [78], where one can expand up to the order of the slow-roll parameters that is desired. The slow-roll approximation then amounts to a truncation at the linear order.

From this discussion it follows that $\varepsilon_{5,H}$ plays no role in the slow-roll approximation, since straightforward substitution shows that

$$\varepsilon_{5,H} = (2\varepsilon_{3,H} - \varepsilon_{1,H} + \varepsilon_{3,H}^{(1)})\varepsilon_{3,H}.$$

Thus $\varepsilon_{5,H}$ is quadratic in the slow-roll parameters.

There exists an additional formulation of the slow-roll approximation, which is approximately equal to the previous one when certain conditions hold. This is the *potential formulation* of the slow-roll approximation. The *potential slow-roll parameters* are defined as

$$\varepsilon_{1,W} = sU\frac{W_1}{W}\frac{(UW)_1}{UW}, \qquad \varepsilon_{2,W} = \varepsilon_{1,W} + 2\left(\frac{sUW_1}{W}\right)_1, \tag{4.21.a}$$

$$\varepsilon_{3,W} = -sU\frac{U_1W_1}{UW}, \qquad \varepsilon_{4,W} = -sU\frac{s_1W_1}{sW}. \tag{4.21.b}$$

These definitions are based on the observation that, as long as the slow-roll approximation is valid, $sU^2W_1 \approx -3H\,\mathrm{D}_t\,\phi$, which can be seen by truncation of (4.17) at the leading order. With this approximation it is not difficult to check that $\varepsilon_{1,H} \approx \varepsilon_{1,W}$. Similar relations hold for the other slow-roll parameters. Their approximate equality makes it convenient to exchange one formulation in favour of the other. It is for this reason that the subscripts H or W will be omitted in what follows.

The potential formulation enables one to make direct inflationary predictions, once the behaviour of the inflaton field is specified. The number of e-folds generated by inflation can be expressed in terms of the inflaton field. When the leading contribution of (4.17) is inserted into (4.14), one obtains

$$N_\mathrm{e} = \int_{\phi_\mathrm{i}}^{\phi_\mathrm{f}} \frac{H^2}{H\,\mathrm{D}_t\,\phi}\,\mathrm{d}\phi \approx -\frac{1}{2}\int_{\phi_\mathrm{i}}^{\phi_\mathrm{f}} \left(\frac{sUW_1}{W}\right)^{-1} \mathrm{d}\phi,$$

where ϕ_i and ϕ_f denote the field values at the beginning and end of inflation, respectively.

4.5.5 *Frame invariance of cosmological observables*

A complication of cosmological scalar-tensor theories over their minimally coupled counterparts is that scalar-tensor theories have to deal with the question of whether results are truly physical, or merely an artefact of the frame parametrisation. The issue of frame parametrisations is described in **chapter 3**. This section will briefly outline how the slow-roll formalism can be formulated such that it is covariant with respect to the transformation from the JF to the EF, and vice versa.

Since the transition from the JF to the EF involves a conformal transformation, it is no surprise that the expansion produced by inflation is a frame-dependent concept. It can be shown exactly that, under a frame-transition as described in **section 3.5**, the number of e-folds (4.14) transforms as

$$N_\mathrm{e} = \int_{a_\mathrm{i}}^{a_\mathrm{e}} \mathrm{d}\log a = \int_{\tilde{a}_\mathrm{i}}^{\tilde{a}_\mathrm{e}} \mathrm{d}\log \tilde{a} + \int_{\Omega_\mathrm{i}}^{\Omega_\mathrm{e}} \mathrm{d}\log\Omega = \tilde{N}_\mathrm{e} + \log(\Omega_\mathrm{e}/\Omega_\mathrm{i}).$$

Similarly, the instance at which a comoving length scale enters the Hubble radius becomes frame dependent.

Thus, N_e is not invariant under a frame transition. Similarly, the size of the comoving Hubble radius will be a frame-dependent statement. Both of these quantities are important for the formulation of inflationary cosmology.

The above is not necessarily a problem, as the duration of inflation itself becomes a frame-dependent concept. One must specify a precise condition for inflation to end, such that this condition is invariant under frame transformations. It turns out that the frame-independent condition is when

$$\max(\varepsilon_1 + \varepsilon_3, |\varepsilon_1 + \varepsilon_2 + 3\varepsilon_3 + \varepsilon_4|) = 1. \tag{4.22}$$

Equation **(4.22)** ensures that cosmological observables are invariant under frame transitions [79].

4.6 COSMOLOGICAL PERTURBATIONS

Observations indicate that the universe is largely homogeneous and isotropic. Deviations from perfect homogeneity and isotropy are small, of the relative order of 10^{-5}. This justifies a linear parametrisation of the perturbations, where the metric and scalar field are decomposed as

$$g_{\mu\nu}(t,\mathbf{x}) = \bar{g}_{\mu\nu}(t) + \delta g_{\mu\nu}(t,\mathbf{x}), \qquad \phi(t,\mathbf{x}) = \bar{\phi}(t) + \delta\phi(t,\mathbf{x}). \tag{4.23}$$

The background values $\bar{g}_{\mu\nu}$ and $\bar{\phi}$ can be taken to be form-invariant under coordinate transformations [80]. That is, the background values are invariant under coordinate transformations, and generally only the perturbations are affected. It is therefore not clear beforehand whether or not the inhomogeneities in **(4.23)** are physical inhomogeneities, or whether they are the result of some coordinate system. The effect of coordinate transformations on the perturbations will be considered here. Under a coordinate transformation $x^\mu \to \tilde{x}^\mu = \tilde{x}^\mu(x)$ the metric and inflaton field transform as

$$\tilde{g}_{\mu\nu}(\tilde{x}) = \frac{\partial x^\alpha}{\partial \tilde{x}^\mu}\frac{\partial x^\beta}{\partial \tilde{x}^\nu} g_{\alpha\beta}(x), \qquad \tilde{\phi}(\tilde{x}) = \phi(x). \tag{4.24}$$

Of particular importance are the infinitesimal coordinate transformations

$$x^\mu \to \tilde{x}^\mu(x) = x^\mu - \xi^\mu. \tag{4.25}$$

Substitution of **(4.25)** into **(4.24)** then yields the dependence of the perturbations on the coordinate transformation up to first order in the gauge parameter:

$$\widetilde{\delta g}_{\mu\nu} = \delta g_{\mu\nu} + \mathcal{L}_\xi g_{\mu\nu}, \qquad \widetilde{\delta\phi} = \delta\phi + \mathcal{L}_\xi \phi. \tag{4.26}$$

Here $\mathcal{L}_\xi$ denotes the Lie derivative along ξ^μ.

4.6.1 *Gauge dependence of the perturbations*

Equation **(4.26)** can be used to determine the dependence of the perturbations of the metric and inflaton field on coordinate transformations. In order to do so it is convenient to characterise the degrees of freedom in terms of their irreducible representations of the group of spatial rotations. The perturbations of the spatial part of the metric $\delta\gamma_{ab}$ are then decomposed into scalars, vectors and tensors:

$$\delta\gamma_{ab} = 2h\bar{\gamma}_{ab} + \bar{\mathrm{D}}_a\bar{\mathrm{D}}_b h' + \bar{\mathrm{D}}_{(a}h^{\mathrm{T}}_{b)} + h^{\mathrm{TT}}_{ab}. \tag{4.27}$$

The scalar degrees of freedom are parametrised by the trace h and the longitudinal vector mode h'. The h_a^{T} and h_{ab}^{TT} respectively denote the transverse vector degrees of freedom and the transverse traceless tensor degrees of freedom. They satisfy

$$\bar{\mathrm{D}}^a h_a^{\mathrm{T}} = \bar{\gamma}^{ab} h_{ab}^{\mathrm{TT}} = \bar{\mathrm{D}}^a h_{ab}^{\mathrm{TT}} = 0.$$

The vector degrees of freedom of the metric can similarly be decomposed in terms of a scalar and a transverse vector:

$$\delta g_{0a} = \bar{\mathrm{D}}_a N' + N_a^{\mathrm{T}}.$$

Finally, the vector component of the gauge parameter can be decomposed into 1 scalar and 1 transverse vector:

$$\xi_a = \bar{\mathrm{D}}_a \xi + \xi_a^{\mathrm{T}}.$$

It is then not difficult to show that, with the homogeneous and isotropic FLRW background **(4.6)**, the perturbation of the lapse transforms as

$$\delta\tilde{N} = \delta N - N\bar{\mathrm{D}}_t(N^{-1}\xi_0), \tag{4.28}$$

the perturbation of the shift vector transforms as

$$\bar{\mathrm{D}}_a\tilde{N}' = \bar{\mathrm{D}}_a(N' + Na^2[\bar{\mathrm{D}}_t(a^{-2}\xi) + a^{-2}\xi_0]), \tag{4.29.a}$$

$$\tilde{N}_{\mathrm{a}}^{\mathrm{T}} = N_a^{\mathrm{T}} + Na^2\bar{\mathrm{D}}_t(a^{-2}\xi_a^{\mathrm{T}}), \tag{4.29.b}$$

the perturbations of the components of the spatial metric transform as

$$\tilde{h} = h - \tfrac{1}{3}\bar{\bar{K}}\xi_0, \tag{4.30.a}$$

$$\bar{\mathrm{D}}_a\bar{\mathrm{D}}_b\tilde{h}' = \bar{\mathrm{D}}_a\bar{\mathrm{D}}_b(h' + 2\xi), \tag{4.30.b}$$

$$\bar{\mathrm{D}}_{(a}\tilde{h}_{b)}^{\mathrm{T}} = \bar{\mathrm{D}}_{(a}(h_{b)}^{\mathrm{T}} + 2\xi_{b)}^{\mathrm{T}}), \tag{4.30.c}$$

$$\tilde{h}^{\mathrm{TT}} = h_{ab}^{\mathrm{TT}}. \tag{4.30.d}$$

and finally, the perturbation of the inflaton field transforms as

$$\delta\tilde{\phi} = \delta\phi - \xi_0\bar{\mathrm{D}}_t\bar{\phi}. \tag{4.31}$$

4.6.2 *Gauge invariant perturbations*

General relativity is special among field theories in the sense that it does not *a priori* favour any single coordinate system. Occasionally, as is the case in the FLRW universe, a coordinate system can be considered to be preferable over others due to the symmetry properties of spacetime. However, no such preferred coordinate system exists for cosmic perturbations, and hence this freedom in the choice of coordinates is problematic. It is not immediately obvious whether a given inhomogeneity is a genuine perturbation, or merely an artefact of the chosen coordinate system.

Also known as *gauge freedom* or *diffeomorphism invariance.*

As an example, consider a universe that is homogeneous and isotropic, which is filled with some quantity $q(t,\mathbf{x}) = q(t)$. The freedom in the choice of coordinates ensures that the time variable $\tilde{t} = t + \delta t(t,\mathbf{x})$ is equally valid. It follows that, in general, q depends on the spatial coordinates $\mathbf{x}$. Under the assumption that $\delta t \ll t$, it can be seen that

$$q(t) = q(\tilde{t} - \delta t) \approx q(\tilde{t}) - \partial_t q \delta t.$$

The first term on the right-hand side must be interpreted as the background value of the quantity q, while the second term defines a perturbation $\delta q(\tilde{t},\mathbf{x})$ that breaks the homogeneity. This perturbation is not a physical one, but a fictitious perturbation that results from the chosen coordinate system. This is illustrated in **figure 4.5**. One can in an analogous fashion ensure that any physical perturbation vanishes.

Given that the perturbations depend on the choice of coordinate system one would expect that any given fluctuation must be treated with great care. It is, however, possible to construct objects which are manifestly the same in each coordinate system, and therefore represent perturbations that are genuinely physical. In this section it will become clear that these *gauge invariant* objects are not unique, but can nevertheless be given representations such that their equations of motion are—at least in form—simple. These representations will be derived here.

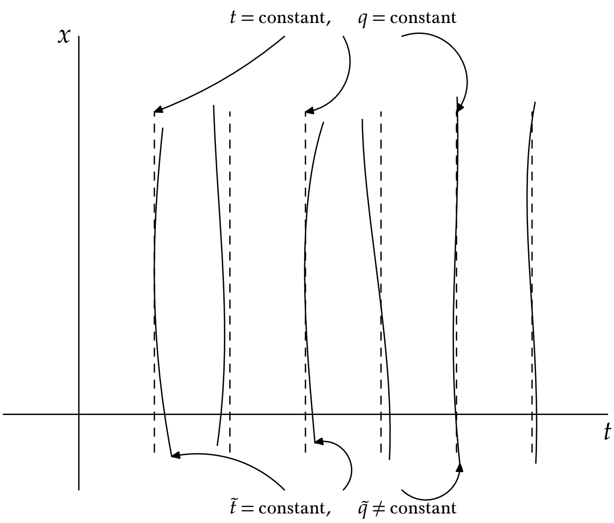

Figure 4.5 Distribution of some quantity q in different coordinate systems.

The simplest such object combines the trace perturbations of the metric and scalar field perturbations. With **(4.30.a)** and **(4.31)** it is straightforward to check that

$$\varpi = \delta\phi - 3(\bar{\mathrm{D}}_t\bar{\phi}/\tilde{K})\, h, \tag{4.32}$$

with $\tilde{K} = 3H$, is gauge invariant. This defines the *comoving density perturbations.* A related object is the *comoving curvature perturbation*, which is defined as

$$\zeta = h - \tfrac{1}{3}(\bar{K}/\bar{\mathrm{D}}_t\bar{\phi})\delta\phi. \tag{4.33}$$

The vector modes can be combined with the lapse perturbation to form the quantities

$$w \ = \delta N + N\bar{\mathrm{D}}_t(N^{-1}N' - \tfrac{1}{2}a^2\bar{\mathrm{D}}_t(a^{-2}h')), \tag{4.34.a}$$

$$w_a = N^{\mathrm{T}}_a - \tfrac{1}{2}a^2 N\bar{\mathrm{D}}_t(a^{-2}h^{\mathrm{T}}_a). \tag{4.34.b}$$

It can straightforwardly be checked, by use of **(4.28)** and **(4.29)**, that these too are gauge invariant. Note that it is not necessary to use the perturbation of the lapse function in the construction above. One could instead have used the scalar field perturbation, or the trace perturbation. However, the above construction has the satisfying property that all perturbations are used exactly once.

Although **(4.34)** is gauge invariant, the longitudinal mode is non-dynamical, while the transverse vector modes decay during inflation. They will therefore be neglected from now on.

The transverse traceless tensor perturbations are themselves gauge invariant, and need no auxiliary variables.

Gauge invariant perturbations are not unique, as any linear combination or rescaling of gauge invariant objects is again gauge invariant, as long as the coefficients of said combination or rescaling depend only on the background variables. However, the comoving curvature perturbation and the tensor perturbation are the simplest, and their use is most widespread in the literature.

4.6.3 *Quantised scalar perturbations*

In order to perform a canonical quantisation of the gauge invariant perturbations one would need an action for the perturbations. This can be derived from the general action **(3.2)**, with the insertion of the decomposition **(4.23)**. It is sufficient to expand the result up to second order in the perturbations, since these are assumed to be small. The gauge invariant action for the scalar perturbations can then be derived by expressing the perturbations by their gauge invariant combination **(4.33)**. Although the explicit calculation is straightforward, it requires time and care. This goes beyond the scope of this work. For this reason, the results are briefly summarised. The explicit calculation for the action of the comoving curvature perturbation has been calculated in detail in reference [81] and the references therein.

The free action for the comoving curvature perturbation with a FLRW background **(4.6)** is

$$S_{\mathrm{S}} = \int z_{\mathrm{S}}^2 \left(\mathrm{D}_t\,\zeta\,\mathrm{D}_t\,\zeta - a^{-2}\,\mathrm{D}_a\,\zeta\,\mathrm{D}^a\,\zeta\right) a^3\,N\mathrm{d}t\mathrm{d}^3x,$$

where

$$z_{\mathrm{S}}^2 \equiv \left(\frac{\mathrm{D}_t\,\phi}{H}\right)^2 s^{-1}\left(1 + \tfrac{1}{2}H^{-1}\,\mathrm{D}_t\log U\right)^{-2}. \tag{4.35}$$

It is convenient to analyse the scalar perturbations in the parametrisation $N = a$ (conformal time), and to introduce the *Mukhanov-Sasaki* (MS) *variable* v through

$$v \equiv a z_S \zeta. \tag{4.36}$$

The quantised Fourier components of the MS variable are given by

$$v(\tau, x) = \frac{1}{(2\pi)^{3/2}} \int \left[v_k(\tau) a_{\mathbf{k}} e^{i\mathbf{k}\cdot\mathbf{x}} + v_k^*(\tau) a_{\mathbf{k}}^\dagger e^{-i\mathbf{k}\cdot\mathbf{x}} \right] d^3k, \tag{4.37}$$

where the canonical commutation relations $\left[a_{\mathbf{k}'}, a_{\mathbf{k}}^\dagger\right] = \delta(\mathbf{k} - \mathbf{k}')$ of the creation and annihilation operators $a_{\mathbf{k}}$ and $a_{\mathbf{k}}^\dagger$ imply that the coefficients satisfy the relation

$$W\left[v_k(\tau), v_k^*(\tau)\right] \equiv v_k (v_k^*)' - v_k' v_k^* = i. \tag{4.38}$$

Variation of the MS action with respect to v_k one finds that the Fourier modes satisfy the relation

Note that the Fourier modes depend on the magnitude of $\mathbf{k}$ only, which follows from **(4.39)**.

$$(v_k)'' + \omega_k^2 v_k = 0, \quad \omega_k^2 = k^2 - \frac{(a z_S)''}{a z_S}. \tag{4.39}$$

Two boundary conditions must be imposed in order to uniquely solve this equation. The first boundary condition is given by the Wronskian condition **(4.38)**. The second boundary condition is conventionally taken to be the demand that, at early times, the modes v_k enter a Lorentz invariant vacuum state. This is called the *Bunch-Davies* boundary condition [82]. It states that during early times, before the onset of inflation, the Fourier modes were simple harmonic oscillators, unaffected by curvature. During early times the second term in ω_k^2 can be neglected, and the solutions to **(4.39)** are simple plane waves. The Wronskian boundary condition then fixes these solutions to satisfy

$$\lim_{-k\tau \to \infty} v_k = (2k)^{-\frac{1}{2}} \exp(-i k \tau), \tag{4.40}$$

which fixes the second boundary condition.

Comparison of **(4.40)** with **(4.37)** yields the observation that the Bunch-Davies condition selects the positive frequency mode functions. For these functions the vacuum is the lowest energy state, since it is annihilated by the annihilation operators $a_{\mathbf{k}}$:

$$a_{\mathbf{k}} |0\rangle = 0. \tag{4.41}$$

This vacuum state is called the *Bunch-Davies vacuum*.

In the slow-roll approximation the frequencies can be explicitly evaluated to first order in the slow-roll parameters:

$$\frac{(a z_S)''}{a z_S} \approx (2 - 3\mathcal{E}_S)\, \tau^{-2}. \tag{4.42}$$

Here, $\mathcal{E}_{\mathrm{S}} \equiv 2\varepsilon_1 - \varepsilon_2 - \varepsilon_4$. As an intermediate step in the derivation of (4.42) it is necessary to express the quantity aH in terms of conformal time. With the assumption that the slow-roll parameters are constant, the comoving Hubble radius can be expressed in terms of conformal time as

$$aH = -(1 - \varepsilon_1)^{-1}\tau^{-1}. \tag{4.43}$$

Under these simplifications (4.39) can be solved analytically in terms of Bessel functions J_α and Y_α, respectively of the first and second kind. Equation (4.39) can be written in terms of the rescaled time coordinate $x = -k\tau$, which leads to

The solution is frequently represented in terms of Hankel functions in the literature.

$$v_k(x) = \sqrt{x}\left[A_k J_{2\alpha}(x) + B_k Y_{2\alpha}(x)\right], \qquad \alpha = \tfrac{3}{4}\sqrt{1 + \tfrac{4}{3}\mathcal{E}_{\mathrm{S}}}. \tag{4.44}$$

The solution that satisfies the boundary conditions (4.38) and (4.40) is [83]

$$A_k = \tfrac{1}{2}(\pi/k)^{\frac{1}{2}} \exp\left[\pi \mathrm{i}(\tfrac{1}{4} + \alpha)\right], \quad B_k = \mathrm{i}A_k. \tag{4.45}$$

The asymptotic boundary condition (4.40) fixes the behaviour of the MS variable in the infinite past, before the onset of inflation.

Now that the exact solution to (4.39) has been found, it is now appropriate to make some comments about its future-directed evolution, when $x \to 0$. In this regime, the time-dependent part (4.42) dominates the Mukhanov-Sasaki equation, and by inspection it can be seen that

$$v_k \propto a z_{\mathrm{S}}, \qquad x \ll 1.$$

Equation (4.36) then implies that the momentum modes of the original variable, the comoving curvature perturbation ζ, become constant at small values of x. From (4.43) it follows that the momentum modes of ζ become constant once their associated comoving wavelength exits the comoving Hubble horizon.

From the discussion in section 4.4 it follows that for these momentum modes their wavelength is larger than the comoving Hubble radius, or even (in the inflationary stage) the particle horizon. The spectrum therefore becomes time-independent at these *superhorizon scales*, and it is sufficient to calculate the value of the spectrum at the time τ_* when the mode k_* exited the horizon.

4.6.4 *Quantised tensor perturbations*

Tensor perturbations are generated during inflation. The process closely follows that of the scalar perturbations. More details can be found in reference [81].

Since the tensor perturbations h^{TT}_{ab} are gauge invariant by themselves, it is convenient to treat them separately. Expansion of the action (3.2) in the tensor perturbations results in

$$S_{\mathrm{T}} = \tfrac{1}{2}\int a^3 z_{\mathrm{T}}^2 \left(\mathrm{D}_t h^{\mathrm{TT}}_{ab}\, \mathrm{D}_t h^{\mathrm{TT}ab} - a^{-2}\partial_a h^{\mathrm{TT}}_{ab}\partial^a h^{\mathrm{TT}ab}\right) N \mathrm{d}t\mathrm{d}^3x,$$

where now

$$z_{\mathrm{T}}^2 = \tfrac{1}{2}U.$$

The tensor perturbations can be decomposed in terms of its Fourier modes

$$h_{ab}^{\mathrm{TT}} = \sum_I e_{ab}^I h_I,$$

where e_{ab}^I is a polarisation tensor. It satisfies the completeness relation

$$e_{ab}^I e^{Jab} = \delta^{IJ}.$$

The derivation of the equations of motion for the tensor perturbations largely follows the derivation for the scalar perturbations.

The action for the tensor perturbations can be written in its canonical form by the introduction of the MS variable

$$u_I \equiv a z_{\mathrm{T}} h_I,$$

and after a decomposition of u_I in terms of Fourier modes

$$u_I = \frac{1}{(2\pi)^{3/2}} \int \left(u_{I,k} a_k \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{x}} + u_{I,k}^* a_k^\dagger \mathrm{e}^{-\mathrm{i}\mathbf{k}\cdot\mathbf{x}} \right) \mathrm{d}^3 k,$$

the equations of motion for the Fourier modes in conformal time is then seen to be

$$u_{I,k}'' + \omega_k^2 u_{I,k} = 0, \qquad \omega_k^2 = k^2 - \frac{(a z_{\mathrm{T}})''}{a z_{\mathrm{T}}},$$

which is identical to **(4.39)** with a different generalised mass z_{T}. If the slow-roll approximation holds, one can derive that

$$\frac{(a z_{\mathrm{T}})''}{a z_{\mathrm{T}}} \approx \frac{2 + 3\mathcal{E}_{\mathrm{T}}}{\tau^2},$$

where $\mathcal{E}_{\mathrm{T}} = \varepsilon_1 + \varepsilon_3$. Given that the equations of motion for the tensor perturbation account for two copies of the scalar perturbations with a modified frequency, it is not necessary to perform the quantisation procedure from first principles, but it suffices to apply the results from **section 4.6.3**.

4.7 INFLATIONARY OBSERVABLES

The quantum fluctuations are taken to be random, and observations indicate that their probability distribution is, to a high extent, Gaussian. This section outlines how these Gaussian perturbations leave their imprints on the CMB power spectrum.

4.7.1 *Scalar power spectrum*

Since the cosmic perturbations are Gaussian, their probability distribution is determined entirely in terms of their two-point correlation function and its Fourier transform P_v (called the *power spectrum*):

Scalar-tensor theories can induce non-Gaussianities, but these are neglected in this work.

$$\xi_v(\tau,\mathbf{x}-\mathbf{x}') \equiv \langle v(\tau,\mathbf{x}), v(\tau,\mathbf{x}')\rangle, \qquad P_v(t,\mathbf{k}) \equiv \int \xi_v(\tau,\mathbf{r}) e^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}}\,\mathrm{d}^3 r.$$

The quantum mechanical treatment follows from the quantisation (4.37), together with the Bunch-Davies vacuum (4.41).

The averages are understood to be statistical averages. With these definitions the two-point correlator of the Fourier modes can be written down as

$$\langle v_{\mathbf{k}}(\tau), v_{\mathbf{k}'}(\tau)\rangle = P_v(\tau,\mathbf{k})\delta^3(\mathbf{k}-\mathbf{k}').$$

In the inflationary scenario the v is identified with perturbations in the homogeneous and isotropic FLRW background. Isotropy of the FLRW universe implies that the correlation function $\xi_v(\mathbf{r})$ can only depend on the magnitude of the separation vector $\mathbf{r}$. Consequently, the power spectrum $P_v(\tau,\mathbf{k})$ is only a function of the magnitude of the wave vector $\mathbf{k}$. The two-point correlation function and the power spectrum can therefore be expressed in terms of each other as

$$\xi_v(\tau,r) = \int_0^\infty \frac{P_v(\tau,k)}{2\pi^2}\frac{\sin(kr)}{kr}k^2\mathrm{d}k, \qquad P(\tau,k) = 4\pi\int_0^\infty \xi_v(\tau,r)\frac{\sin(kr)}{kr}r^2\mathrm{d}r.$$

In the literature the variance of the MS variable is frequently used instead of the power spectrum. It can be obtained from the correlation function in to limit $r \to 0$:

$$\xi_v(\tau,0) = \langle v(\tau,\mathbf{x}), v(\tau,\mathbf{x})\rangle = (2\pi^2)^{-1}\int_0^\infty \Delta^2(\tau,k)\,\mathrm{d}\log k,$$

where the dimensionless power spectrum Δ^2 is defined as

$$\Delta^2(\tau,k) \equiv (2\pi^2)^{-1}k^3 P_v(\tau,k).$$

The above notation can be used in order to find the approximate power spectrum Δ^2_{S} of the comoving curvature perturbation at the moment of horizon crossing, under the assumption the slow-roll approximation is holds. From (4.36) one obtains

$$\Delta^2_{\mathrm{S}}(\tau,k) = \frac{1}{2\pi^2}\frac{|v_k|^2}{z^2_{\mathrm{S}}}.$$

With (4.35), the late-time behaviour of (4.44), together with (4.45), the dimensionless power spectrum can then be written as

$$\Delta^2_{\mathrm{S}}(\tau,k) = \frac{H^2/U}{16\pi^2(\varepsilon_1+\varepsilon_3)}\left[1-2(\varepsilon_1-\varepsilon_3)+2c_\gamma\mathcal{E}_{\mathrm{S}}-2\mathcal{E}_{\mathrm{S}}\log\left(\frac{k}{aH}\right)\right], \tag{4.46}$$

where $c_\gamma \equiv 2-\gamma_{\mathrm{E}}-\log 2 \approx 0.7296$ is a numerical constant, and γ_{E} is the Euler-Mascheroni constant. The power spectrum is typically parametrised via the power law ansatz

$$\Delta^2_{\mathrm{S}} = A_{\mathrm{S}}(k_*)\left(\frac{k}{k_*}\right)^{n_{\mathrm{S}}(k_*)+\frac{1}{2}\alpha_{\mathrm{S}}(k_*)+\dots}. \tag{4.47}$$

Here, k_* is an arbitrary pivot scale, which is typically identified with the mode that enters the Hubble radius: $k_* = aH$.

The spectra are characterised by an *amplitude* A_{S}, a *scalar spectral index* n_{S} and a *running of the scalar spectral index* α_{S}. With the use of **(4.19)**, the spectral amplitude can be extracted from **(4.46)** and expressed in terms of the inflaton potential:

$$A_{\mathrm{S}} \approx \frac{W}{72\pi^2(\varepsilon_1 + \varepsilon_3)}.$$

The scale dependence is parametrised by the scalar spectral index

$$n_{\mathrm{S}} \equiv 1 + \frac{\mathrm{d}\log \Delta^2}{\mathrm{d}\log k}\Big|_{k=k_*} \approx 1 - 2\mathcal{E}_{\mathrm{S}}.$$

Lastly, the spectral index itself may depend on the momentum scale. This scale dependence is parametrised by the running of the scalar spectral index

$$\alpha_{\mathrm{S}} \equiv \frac{\mathrm{d}n_{\mathrm{S}}}{\mathrm{d}\log k}\Big|_{k=k_*} \approx -2H^{-1}\frac{\mathrm{D}_t\,\mathcal{E}_{\mathrm{S}}}{1-\varepsilon_1}.$$

Notice that the running of the spectral index is itself second order in the slow-roll parameters.

4.7.2 *Tensor power spectrum*

The derivation of the tensor power spectrum closely follows that of the scalar power spectrum. The main difference is that the tensor power spectrum accounts for both tensor polarisations. The result is

$$\Delta^2(\tau,k) = \frac{H^2}{6\pi U}\left[1 - 2\varepsilon_1 + 2c_\gamma\mathcal{E}_{\mathrm{T}} - 2\mathcal{E}_{\mathrm{T}}\log\left(\frac{k}{aH}\right)\right]. \tag{4.48}$$

The *tensor amplitude* is

$$A_{\mathrm{T}} \approx \frac{W}{36\pi}. \tag{4.49}$$

The *tensor spectral index* n_{T} and *running of the tensor spectral index* α_{T} are

$$n_{\mathrm{T}} \equiv \frac{\mathrm{d}\Delta^2}{\mathrm{d}\log k}\Big|_{k=k_*} = -2\mathcal{E}_{\mathrm{T}}, \qquad \alpha_{\mathrm{T}} \equiv \frac{\mathrm{d}n_{\mathrm{T}}}{\mathrm{d}\log k}\Big|_{k=k_*} = -2H^{-1}\frac{\mathrm{D}_t\,\mathcal{E}_{\mathrm{T}}}{1-\varepsilon_1}.$$

The tensor-to-scalar ratio r is defined as

$$r \equiv A_{\mathrm{S}}/A_{\mathrm{P}} = 16\mathcal{E}_{\mathrm{T}},$$

which is a direct measure for the energy scale of inflation. Lastly, notice that the tensor-to-scalar ratio satisfies the generic consistency condition

$$r = -8n_{\mathrm{T}}.$$

4.7.3 *Empirical values*

During the past decades the CMB power spectrum has been measured with increasing precision, by experiments such as the *Cosmic Background Explorer* (COBE), the *Wilkinson Microwave Anisotropy Probe* (WMAP) and—most recently—*Planck* [84]. With the assumption of the power law parametrisation **(4.47)**, the *Planck* collaboration measured the parameters of the power spectrum with respect to a pivot scale $k_* = 5 \times 10^{-2}\,\mathrm{Mpc}^{-1}$. This pivot point is typically taken to be the scale of the perturbation mode that entered the horizon at $N_{\mathrm{e}} = 60$ e-folds before the end of inflation, but the choice for the pivot point is arbitrary. The pivot scale must be chosen within the window of scales observable in the CMB [84]:

$$k_*^{\mathrm{min}} < k_* < k_*^{\mathrm{max}}, \quad k_*^{\mathrm{min}} = 1 \times 10^{-4}\,\mathrm{Mpc}^{-1}, \quad k_*^{\mathrm{max}} = 1 \times 10^{-1}\,\mathrm{Mpc}^{-1}. \tag{4.50}$$

Measurements of the CMB constrain A_{S} and n_{S} and give an upper bound on the tensor-to-scalar ratio r. Their most recently measured values are given here for the pivot scale $k_* = 5 \times 10^{-2}\,\mathrm{Mpc}^{-1}$ [84]:

$$A_{\mathrm{S},*}^{\mathrm{obs}} = (2.099 \pm 0.014) \times 10^{-9} \qquad 68\,\%\ \mathrm{CL}, \tag{4.51}$$

$$n_{\mathrm{S},*}^{\mathrm{obs}} = 0.9649 \pm 0.0042, \qquad 68\,\%\ \mathrm{CL}, \tag{4.52}$$

$$r_*^{\mathrm{obs}} < 0.11 \qquad 95\,\%\ \mathrm{CL}. \tag{4.53}$$

The upper bound on r_*^{obs} can be converted into an upper bound on the energy scale during inflation. This corresponds to an upper bound on the energy density, given by the EF potential $\tilde{V} = M_{\mathrm{P}}^4 W/4$, as follows from **(3.11)** and **(4.49)**.

5 QUANTUM GEOMETRODYNAMICS

THE EXISTENCE OF THE COSMOLOGICAL PERTURBATIONS can be explained by the appealing to quantum theory. The uncertainty principle predicts that the homogeneous inflaton field will have small, quantised fluctuation. These fluctuations will lead to deviations in the energy density of the universe, and as a result different parts of space will inflate by different amounts. Similarly, the zero-point fluctuations of the gravitational field will yield small, quantised deviations from the classical homogeneous and isotropic FLRW metric.

The approach of cosmological perturbation theory, where quantised fluctuations propagate on a classical background, is justified experimentally; there are no indications so far that the gravitational interaction exhibits quantum mechanical effects. The hope has been expressed that a quantisation of the gravitational field is unnecessary conceptually, and that an exact semiclassical theory is enough to describe the fundamental interactions. The quantised degrees of freedom would then couple with the classical degrees of freedom of the gravitational field and result in the *semiclassical Einstein field equations* [85]

$$R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} = 2\kappa^{-1}\langle\psi|\hat{T}_{\mu\nu}|\psi\rangle.$$

The object $\hat{T}_{\mu\nu}$ represents the energy-momentum tensor operator and ψ is the quantum state of the matter fields. Cosmological perturbation theory is one particular application of semiclassical gravity. However, it is uncertain whether such a semiclassical theory is consistent. The approach has been subject to scrutiny, most prominently in reference [86]. The question whether semiclassical gravity, or its criticism, can be placed on a firm, irrefutable basis continues to this day. Recent discussion have, for example, been published in the references [87–89].

However, as quantum mechanics seems to be a universal description of nature, it is natural to wonder what whether it is possible to consistently quantise theories of gravity, and whether any of these descriptions are realised in nature.

One of the motivations for the quantisation of gravity is the possibility of its unification with the other three fundamental interactions: the electromagnetic force and the strong and weak nuclear forces. Since all three of these interactions are known to be quantised, and it has been shown that this quantisation is realised in nature. It may be hoped that such a unification leads to a non-perturbative theory of the fundamental interactions, which is free from the need to renormalise the divergences that occur in quantum field theory.

Another motivation originates in the study of cosmology and black holes. A seemingly necessary feature therein is the presence of singularities, which leads to the breakdown of current descriptions of gravity [90, 91]. Similar considerations in the other fundamental interactions have given rise to quantum theories, and as a result it seems reasonable that, for example, the initial singularity of the Big Bang in the early universe eventually leads to a consistent theory of quantum gravity.

An overview of different approaches to quantum gravity can be found in reference [92]. One such approach is based on the Hamiltonian formulation of the theory of general relativity. This follows the non-relativistic quantisation of classical mechanics, where one imposes that the wave function of the system satisfies the *Schrödinger equation*. In the context of general relativity, the canonical variables are the spatial metrics, which are obtained by a foliation of spacetime as described in **chapter 2**, and their conjugate momenta. In this way the theory of general relativity can be formulated as a dynamical theory of spatial geometries, called *geometrodynamics*.

The Hamiltonian formalism gives rise to constraints. These constraints are a general feature of theories that are invariant under a reparametrisation of time [93]. In the case of general relativity—or in the present case, scalar-tensor theories—these constraints impose certain relations between the phase space variables. The most important of these relations is known as the *Wheeler-DeWitt* (WDW) *equation*.

The quantisation procedure of geometrodynamics is not without complications of its own, and for this reason the treatment of the full WDW equation typically does not go beyond the formal level. The WDW equation itself composes an infinite number of second order partial functional differential equations which, barring highly symmetric cases, are extremely difficult to solve. In addition, the terms in the WDW equation consist of products of non-commuting operators that are evaluated at the same point. These terms lead firstly to ambiguities in the operator ordering, and secondly to singular Dirac delta distributions. The regularisation procedure for quantum geometrodynamics goes beyond the scope of this work. A general overview can be found in reference [94].

Although the regularisation of operators is as of yet an open problem in physics, the question of operator ordering can be partially addressed. For example, one can demand that the quantum theory is invariant under reparametrisations of the phase space variables. This would fix the operator ordering to be the familiar Laplace-Beltrami ordering. Furthermore, conditions can occasionally be imposed under which the final results do not depend on the operator ordering. This was first shown in reference [95]. The applicability of these conditions will be investigated in **chapter 6**.

Finally, an important hurdle in the canonical formulation of gravity is the question whether or not it admits a probabilistic interpretation. Related to this question is the existence of a conserved probability that is positive and normalisable. This issue will be partially addressed in this chapter.

This chapter summarises the main results from the study of the quantisation of constrained systems that are necessary for the analysis that is presented in the next chapters. For a comprehensive overview of the topic of quantum geometrodynamics—and whether or not a quantised theory of gravity is needed at all–the reader is invited to consult the references [92,94,96–99] and the references contained therein.

5.1 HAMILTONIAN FORMULATION

The Hamiltonian formulation of general relativity and similar dynamical theories of spacetime, such as scalar-tensor theories, leads to a non-perturbative quantised theory of gravity called *quantum geometrodynamics* [100,101]. The preliminary—that is, classical—work leading up to this quantised theory is presented here. The Lagrangian of scalar-tensor theories is presented in **(3.2)**. It is convenient to write the action in the familiar form using the reparamatrisation invariant time derivative $\mathrm{D}_t = \mathcal{L}_n$, which is the Lie derivative along the normal vector on the hypersurfaces Σ_t:

$$L = \int N\left(\tfrac{1}{2}\mathcal{M}_{AB}\,\mathrm{D}_t q^A\,\mathrm{D}_t q^B - \mathcal{P}(q)\right)\,\mathrm{d}^3x. \tag{5.1}$$

The indices A, B and so on are abstract configuration space indices, and the q^A indicate the configuration space coordinates. The quantities $\mathcal{M}_{AB}$ and $\mathcal{P}$ are respectively the configuration space metric and potential. Their explicit form will be derived in **chapter 6**, but for the present discussion this form is not relevant. In the case of pure Einstein-Hilbert gravity the dynamical variables are the components of the spatial metric $q^A = \gamma_{ab}$. For more general theories, such as scalar-tensor theories, the q^A comprise all degrees of freedom. This will be considered in detail in **chapter 6**.

Although the spatial metrics are defined with covariant indices, they are contravariant configuration space coordinates.

The momenta conjugate to the configuration space variables can straightforwardly be calculated from **(5.1)** with the help of D_t:

$$p_A = \frac{\delta L}{\delta \partial_0 q^A} = \mathcal{M}_{AB}\,\mathrm{D}_t q^B,$$

$$p_N = \frac{\delta L}{\delta \partial_0 N} = 0, \qquad p_a^N = \frac{\delta L}{\delta \partial_0 N^A} = 0. \tag{5.2}$$

The vanishing of the momenta p_N and p_a^N, which are canonically conjugate to respectively the lapse function and the shift vector, means that the system is subject to constraints. They are called *primary constraints.* They imply that the lapse function and shift vector are not dynamical, and their value is arbitrary. This reflects the gauge freedom of general relativity. The canonical Hamiltonian H_c can then be obtained by a Legendre transformation:

$$H_\mathrm{c} = \int N p_a\,\mathrm{D}_t q^A\,\mathrm{d}^3x - L = \int \left(N\mathcal{H}_\perp + N^a\mathcal{H}_a\right)\,\mathrm{d}^3x.$$

The Hamiltonian densities which are introduced here are defined by

$$N\mathcal{H}_\perp = \tfrac{1}{2}N\mathcal{M}^{AB}p_A p_B + N\mathcal{P}, \tag{5.3.a}$$

$$N^a\mathcal{H}_a = p_A\,\mathcal{L}_\mathbf{N}\,q^A, \tag{5.3.b}$$

with $\mathcal{L}_\mathbf{N}$ the Lie derivative along the shift vector. The evolution of any function f of the coordinates and their momenta is determined by the *Poisson bracket* of f with the canonical Hamiltonian H_c:

$$\partial_0 f = \{f, H_\mathrm{c}\} \equiv \sum_{i\in I}\int\left(\frac{\delta f}{\delta q^i}\frac{\delta H_\mathrm{c}}{\delta p_i} - \frac{\delta f}{\delta p_i}\frac{\delta H_\mathrm{c}}{\delta q^i}\right)\,\mathrm{d}^3x. \tag{5.4}$$

In particular, the only non-vanishing Poisson bracket of the phase space coordinates is

$$\{q^A(x), p_B(x')\} = \delta^A_B\,\delta^{(3)}(x,x').$$

The Poisson bracket has the defining property that it satisfies the *Jacobi identity*

$$\{A,\{B,C\}\} + \{B,\{C,A\}\} + \{C,\{A,B\}\} = 0, \tag{5.5}$$

for any three arbitrary functions A, B and C of the phase space coordinates. The index i simply ensures that all configuration space variables q^i and their corresponding canonical momenta are considered. It is not difficult to see that

$$\{p_N, H_c\} = -\int N\mathcal{H}_\perp \,\mathrm{d}^3x \equiv -H_\perp[N], \tag{5.6.a}$$

$$\{p_a^N, H_c\} = -\int N^a\mathcal{H}_a \,\mathrm{d}^3x \equiv -H_\parallel[\mathbf{N}]. \tag{5.6.b}$$

Consistency with (5.2) then demands that the Hamiltonians $H_\perp$ and H_a vanish. As these constraints arise from the consistency of primary constraints they are called *secondary constraints*. Together, they form the *Dirac hypersurface deformation algebra*

This is technically not an algebra, as the brackets involve the phase space variables.

$$\{H_\parallel[\mathbf{N}], H_\parallel[\mathbf{M}]\} = H_\parallel[[\mathbf{N},\mathbf{M}]], \tag{5.7.a}$$

$$\{H_\parallel[\mathbf{N}], H_\perp[N]\} = H_\perp[\mathcal{L}_\mathbf{N} N], \tag{5.7.b}$$

$$\{H_\perp[N], H_\perp[M]\} = -H_\parallel[\mathbf{L}]. \tag{5.7.c}$$

where $L^a = M\,\mathrm{D}^a N - N\,\mathrm{D}^a M$. The primary and secondary constraints therefore completely specify all conditions that are placed on the system. From their definitions it is clear that the phase space variables q^A and p_A satisfy

All the constraints are *first class* in the parlance of Dirac [96].

$$\{q^A, H_\parallel[\mathbf{N}]\} = (\mathcal{L}_\mathbf{N} q)^A,$$

$$\{p_A, H_\parallel[\mathbf{N}]\} = (\mathcal{L}_\mathbf{N} p)_A,$$

and therefore $H_\parallel$ is the generator of spatial diffeomorphisms. However, the constraint $H_\perp$ cannot be so straightforwardly be identified with temporal diffeomorphisms. The origin of this complication can be traced back to the hypersurface deformation algebra, as the third bracket involves a dependency of L^a on the spatial metric γ_{ab}. It can be checked that

$$\{q^A, H_\perp[N]\} = N\,\mathrm{D}_t q^A,$$

$$\{p_A, H_\perp[N]\} = N\,\mathrm{D}_t p_A,$$

only on the hypersurfaces where the constraints and the equations of motion are satisfied [102].

In many applications it is more convenient to formulate the constraints in terms of the Hamiltonian densities $\mathcal{H}_\perp$ and $\mathcal{H}_a$, instead of the integrated quantities $H_\perp$ and $H_\parallel$. Once the secondary constraints are implemented to hold identically, it follows from (5.6) that

$$\mathcal{H}_\perp = 0, \quad \mathcal{H}_a = 0.$$

The first of these constraints is called the classical *Hamiltonian constraint*, while the second is called the classical *momentum constraint*. The momenta can be expressed as derivatives of the action through a canonical transformation. The Hamiltonian constraint can then, by (5.3.a), be rewritten as

$$\tfrac{1}{2}\mathcal{M}^{AB}S_{,A}S_{,B}+\mathcal{P}=0,\qquad p_A=S_{,A}. \tag{5.8}$$

The comma denotes a functional derivative with respect to the configuration space coordinates. Equation (5.8) is known as the *Hamilton-Jacobi equation* of the theory. The Hamilton-Jacobi equation completely determines the behaviour of the configuration space coordinates q^A, although it does so in a way that is independent of time. In fact, the Hamilton-Jacobi equation can be taken as a classical definition of time.

It is useful to see how this works in detail, as it will later be useful to show that the WDW equation predicts the classical theory in its semiclassical limit. The Hamilton-Jacobi equation can be varied with respect to q^A to find

$$\mathcal{M}^{AB}S_{,A}S_{,BC}+\tfrac{1}{2}\mathcal{M}^{AB}{}_{,C}S_{,A}S_{,B}+\mathcal{P}_{,C}=0.$$

The above equation can be simplified by the introduction of the vector D_s through

$$\mathrm{D}_s\,S_{,C}+\tfrac{1}{2}\mathcal{M}^{AB}{}_{,C}S_{,A}S_{,B}+\mathcal{P}_{,C}=0,\qquad \mathrm{D}_s=\mathcal{M}^{AB}S_{,A}\delta_B. \tag{5.9}$$

Notice that the action of D_s on q^A results in

$$\mathrm{D}_s\,q^A=\mathcal{M}^{AB}S_{,B}. \tag{5.10}$$

Comparison with (5.8) then leads to the conclusion that, provided that D_s is identified with the reparametrisation invariant time derivative D_t, this coincides with the ordinary Hamilton equation of motion for the configuration space coordinates. Equation (5.10) can then be inserted into (5.9) to find the Euler-Lagrange equations of motion

$$\mathrm{D}_t{}^2\,q^A+\Gamma^A_{BC}\,\mathrm{D}_t\,q^B\,\mathrm{D}_t\,q^C+\mathcal{M}^{AB}\mathcal{P}_{,B}=0,$$

where Γ is the Christoffel symbol calculated from the configuration space metric $\mathcal{M}$. It can immediately be verified that these are the Euler-Lagrange equations of motion by performing the variation of (5.1). In summary: a solution to the Hamilton-Jacobi equation completely determines the behaviour of the dynamical fields. A notion of time evolution can then be obtained by the gradient of the classical action.

5.2 QUANTISATION OF CONSTRAINED SYSTEMS

The quantisation of constrained systems is carried out in the Schrödinger representation. The quantisation is realised by the imposition of the canonical commutation relations on the coordinates and their conjugate momenta:

$$\left[q^A(x),p_B(x')\right]=\mathrm{i}\,\delta^A_B\,\delta^{(3)}(x,x').$$

In the quantised theory the Poisson brackets are therefore replaced by commutators. The constraints (5.2) and (5.6) therefore cannot be considered to be quantum equations—if there were, any dynamics would be trivially absent. The quantised constraints are instead taken to be operators that annihilate the wave function Ψ of the system. The algebra (5.7) ensures that these are all the conditions that are placed on the wave function. Of particular importance are the quantised momentum constraints and the Hamiltonian constraint:

In cosmological scenarios Ψ is frequently called the *wave function of the universe.*

$$\mathcal{H}_a\Psi = 0, \qquad \mathcal{H}_\perp\Psi = 0. \tag{5.11}$$

The quantised Hamiltonian constraint is of such importance that it has its own name, and is called the *Wheeler-DeWitt* (WDW) *equation*. The equations (5.11) are the central objects in canonical quantum gravity. The WDW equation is equivalent to a Schrödinger equation, and therefore the wave function is independent of time. The momentum constraint implies that the wave function is invariant under spatial diffeomorphisms [103]. Recall that the wave function depends on the q^A, which are coordinates in configuration space. These coordinates represent the dynamical degrees of freedom of the gravitational theory, and can be represented in terms of the γ_{ab} and ϕ. The momentum constraint then implies that the wave function does not depend on the particular values of γ_{ab} and ϕ, but on the geometry on which they are defined. This can be accomplished by the identification of the points q^A and $q^{A'}$ that are the same up to a coordinate transformation. If the set of all q^A on the spacetime manifold $\mathcal{M}$ is then temporarily denoted by $\mathcal{C}$ and the set of all coordinate transformations on $\mathcal{M}$ by $\mathrm{C}(\mathcal{M})$, then the wave function depends on the elements of the space

$$\mathcal{S} = \mathcal{C}/\mathrm{C}(\mathcal{M}).$$

The space $\mathcal{S}$ is called *superspace*.

The time-independence of the wave function has no analogue in classical mechanics. This can be illustrated with a simple example. Consider the dynamical representation of the q^A:

$$q^A(t, \mathbf{x}) = \mathrm{e}^{\mathrm{i}H_\mathrm{c}t} q^A(0, \mathbf{x}) \mathrm{e}^{-\mathrm{i}H_\mathrm{c}t},$$

which is the familiar quantum-mechanical evolution of the operator q^A from an initial hypersurface Σ_0 to some arbitrary hypersurface Σ_t. The constraints imply that

$$H_\mathrm{c}\Psi = 0, \quad \Psi^\dagger H_\mathrm{c} = 0$$

and therefore any quantum-mechanical expectation value does not depend on time:

$$\Psi^\dagger q^A(t, \mathbf{x})\Psi = \Psi^\dagger q^A(0, \mathbf{x})\Psi.$$

Similar expressions hold for other (products of) operators. Since observables are defined as expectation values taken over ensembles, it therefore seems that no observable in quantum gravity changes with time.

From this example one comes to the conclusion that nothing changes in quantum geometrodynamics, which in the literature is known as the *problem of time*. If the canonical approach to quantum gravity is valid, then time is a classical concept that has no meaning in quantum gravity. At best, time emerges in semiclassical domains of superspace.

The quantisation in curved spacetimes is ambiguous, as the ordering of the classically commuting variables becomes non-trivial. The factor ordering must be imposed by hand by some physical principle. For example, one might argue that the theory is invariant under reparametrisations of the configuration space coordinates. In this case the position representation of the momentum and the factor ordering of the Hamiltonian take the symmetric forms

$$p_A = -\mathrm{i}|\mathcal{M}|^{-\frac{1}{4}}\delta_A|\mathcal{M}|^{\frac{1}{4}}, \qquad \mathcal{H} = \tfrac{1}{2}|\mathcal{M}|^{-\frac{1}{4}}p_A\sqrt{|\mathcal{M}|}\mathcal{M}^{AB}p_B|\mathcal{M}|^{-\frac{1}{4}} + \mathcal{P}.$$

This is the *Laplace-Beltrami operator ordering* of the WDW equation, and it is the operator ordering that will be used in **chapter 6**, though other operator orderings are common. From the above discussion it becomes clear that an additional complication of the WDW equation is the presence of singular delta functions that arise from second order functional derivatives that are evaluated at the same spacetime point. The treatment of these divergent terms is subtle and not well understood. In this work these terms will be largely neglected, and the analysis kept formal.

Finally, any particular solution of the WDW equation requires initial conditions. Although the main results obtained in **chapter 6** and **chapter 7** do not greatly depend on the boundary conditions, a full discussion of the initial conditions goes beyond the scope of this work. However, since the classical boundary conditions naturally arise from the initial conditions imposed on the solution of the WDW equation, it is appropriate to mention two prominent attempts to explain the initial conditions of the wave function of the universe. These are the *no-boundary* condition of Hartle and Hawking and the *tunneling* condition of Vilenkin and Linde [104–106].

5.3 THE SEMICLASSICAL LIMIT OF QUANTUM GEOMETRODYNAMICS

From the example in the previous section it follows time has no meaning in the quantised theory. The only dynamical degrees of freedom are the spatial metrics and the scalar field evaluated at constant time. Time can only be recovered as a (semi)classical notion, where the dynamical coordinates are classical. Solutions of the Hamilton-Jacobi equation trace out the classical field equations. Therefore the Hamilton-Jacobi formalism provides a notion of time in the semiclassical approximation.

Any theory of quantum gravity has to be able to reproduce known classical results in some limit. This raises the question whether the WDW equation is able to reproduce the classical equations of motion of general relativity (or any of its extensions) in some semiclassical limit. The answer to this question is affirmative. It is therefore instructive see how the classical field equations can, at least formally, be derived from the WDW equation. This can be demonstrated in the Laplace-Beltrami factor ordering, such that the WDW equation can be written as

$$-\tfrac{1}{2}\mathcal{M}^{AB}\nabla_A\nabla_B\Psi + \mathcal{P}\Psi = 0. \tag{5.12}$$

Here the ∇_A denotes the covariant functional derivative with respect to the coordinate q^A. The WDW equation is a second-order functional partial differential equation of which the complexity is enormous. The full WDW equation is therefore, in practice, unsolvable. However, perturbative solutions to the WDW equation exists, and perturbative analyses of the WDW equation form the foundation of the field of ***quantum cosmology***. A semiclassical solution of the WDW equation can be found by the ***Wentzel-Kramers-Brillioun*** (WKB) ansatz

The momentum constraint equation can be expanded in a similar way and ensures the invariance of the wave functional under spatial 3-dimensional diffeomorphisms order by order [107].

$$\Psi(q) = \exp\big(\mathrm{i}\, S(q)\big), \tag{5.13}$$

with S a slowly varying function. Substitution of **(5.13)** into **(5.12)** results in

$$-\tfrac{1}{2}\mathrm{i}\mathcal{M}^{AB}\nabla_A\nabla_B S + \tfrac{1}{2}\mathcal{M}^{AB}\nabla_A S\nabla_B S + \mathcal{P} = 0.$$

Since S is slowly varying the first term on the left-hand side will be negligible in comparison with the rest. What remains is the Hamilton-Jacobi equation for the classical theory. From the discussion of **section 5.1** one can then define a ***semiclassical time*** variable $\mathrm{D}_{t_\mathrm{s}}$ by the directional derivative along the gradient of S. The semiclassical Hamilton equations can be obtained in the same manner as before:

$$\mathrm{D}_{t_\mathrm{s}} q^A = p_A = \nabla_A S, \qquad \mathrm{D}_{t_\mathrm{s}} \equiv \mathcal{M}^{AB}\nabla_A S\,\nabla_B. \tag{5.14}$$

In the derivation of this relation it is assumed that the wave function is strongly peaked about a set of classical trajectories in configuration space. Unfortunately, this does not necessarily have to be the case. However, there may be a hierarchy of different degrees of freedom, some of which are approximately classical, whereas others are to be considered as fully quantum. This has important consequences for the discussion presented here, and allows one to connect the WDW equation to quantum field theory in curved spacetimes.

5.3.1 *The semiclassical expansion*

In semiclassical approaches to quantum theories there frequently is a natural distinction between two sets of configuration space variables, say Q^A and q^i, in the sense that the configuration space metric and the configuration space potential decompose into irreducible blocks

$$\mathcal{M}_{AB}(Q,q) \to M_{AB}(q) + m_{AB}(Q,q), \qquad \mathcal{P}(Q,q) \to \mathcal{P}(Q) + p(Q,q). \tag{5.15}$$

The submetrics M_{AB} and m_{AB} are frequently metrics on the subspaces of the variables Q^A and q^i, respectively, which means that

$$M^{AB} m_{BC} = 0.$$

This condition will be relaxed in **chapter 6**, but it will be assumed in the discussion here for presentational clarity. The distinction between the Q^A and q^A is usually evident in the presence of a small parameter λ. This parameter induces a suppression of the kinetic term of the Q^A, and as a result these variables are frequently called *heavy* or *slow*. The q^A face no such suppression, and are therefore frequently called *light* or *fast*. Furthermore, the metric m_{AB} in the context of the semiclassical expansion typically does not depend strongly on the Q^A, and it can be assumed that its derivatives with respect to the Q^A can be neglected. The function $\mathcal{p}$ acts as a potential for the q^i, but is allowed to have some parametric dependence on the Q^A. Under certain conditions it is then possible to recover a notion of quantum mechanics from the WDW equation.

Examples of a grouping of degrees of freedom into heavy and light can be found in the *Born-Oppenheimer approximation*, where λ is determined by the ratio of the atomic electron and the nuclear masses. In minimally coupled inflationary models, λ is naturally determined by the ratio of the inflaton mass and the Planck mass. In scale-invariant models, such as scalar-tensor theories, the parameter is not manifestly present, although it can be introduced by resorting to physical principles. One can perform a systematic semiclassical approximation of the WDW equation. This is done in detail in **chapter 6** and **chapter 7**. Here a qualitative sketch is provided.

Before proceeding, it is necessary to introduce some additional notation. Because the configuration space metric has the block structure of **(5.15)** it is useful to speak of different components of the covariant derivative ∇_A. When discussing the abstract semiclassical expansion of the WDW equation it is convenient to denote the components ∇_A that are compatible with the metric $\mathcal{M}_{AB}$ by $\bar{\nabla}_A$, while the components that are compatible with the metric m_{AB} by $\tilde{\nabla}_A$.

With the decomposition **(5.15)** of the metric the WDW equation becomes

$$\tfrac{1}{2}\bar{\nabla}_A\bar{\nabla}^A\Psi + \tfrac{1}{2}\tilde{\nabla}_i\tilde{\nabla}^i\Psi + \mathcal{P}\Psi + \mathcal{p}\Psi = 0. \tag{5.16}$$

The form of the Hamiltonian suggests that approximate solutions to this equation can be found via the use of a combined WKB-Born-Oppenheimer ansatz

$$\Psi(Q,q) = \psi(Q,q)\exp[\mathrm{i}\,S(Q)], \tag{5.17}$$

where S is a slowly varying function of the Q^A, as before. The function ψ is slowly varying with respect to the Q^A, but is a swiftly varying function of the q^A. After substitution of the ansatz into **(5.16)** one can divide the result by Ψ to obtain

$$\begin{aligned}&-\tfrac{1}{2}\mathrm{i}\bar{\nabla}_A\bar{\nabla}^A S + \tfrac{1}{2}\bar{\nabla}_A S\,\bar{\nabla}^A S + \mathcal{P}\\ &-\mathrm{i}\bar{\nabla}_A S\,\bar{\nabla}^A\log\psi - \tfrac{1}{2}\psi^{-1}\bar{\nabla}_A\bar{\nabla}^A\psi - \tfrac{1}{2}\psi^{-1}\tilde{\nabla}_i\tilde{\nabla}^i\psi + \mathcal{p} = 0.\end{aligned}$$

The first three terms depend only on the Q^A, while the remaining terms also depend on the q^i. There must therefore exist a function f such that

$$-\tfrac{1}{2}\mathrm{i}\,\bar{\nabla}_A\bar{\nabla}^A S + \tfrac{1}{2}\bar{\nabla}_A S\,\bar{\nabla}^A S + \mathcal{P} = f(Q), \tag{5.18.a}$$

$$-\mathrm{i}\,\bar{\nabla}_A S\,\bar{\nabla}^A\psi - \tfrac{1}{2}\bar{\nabla}_A\bar{\nabla}^A\psi - \tfrac{1}{2}\tilde{\nabla}_i\tilde{\nabla}^i\psi + \mathcal{p}\psi = -f(Q)\psi. \tag{5.18.b}$$

The function $f(Q)$ parametrises the interaction between the Q^A and the q^A. To a first approximation in the semiclassical expansion it is natural to assume that f

A discussion that involves different choices of f is provided in reference [108].

vanishes identically. One can then assume that the higher order derivatives of S can be neglected. Higher order derivatives of ψ with respect to the Q^A can similarly be neglected. One then finds that S satisfies the Hamilton-Jacobi equation for the Q^A alone. The Q^A are therefore classical variables in the semiclassical expansion, and S defines a set of trajectories that coincide with the classical equations of motion for the Q^A. With respect to these variables the semiclassical time **(5.14)** can thus be defined. Substitution of the solution of **(5.18.a)** into **(5.18.b)** then yields

$$\mathrm{i}\,\mathrm{D}_{t_\mathrm{S}}\,\psi = -\tfrac{1}{2}m^{ij}\tilde{\nabla}_i\tilde{\nabla}_j\psi + p\psi \equiv \mathcal{H}_\mathrm{m}\psi, \tag{5.19}$$

where $\mathcal{H}_\mathrm{m}$ is interpreted as a Hamiltonian for the q^A. Thus, to leading order in the semiclassical approximation, the wave function ψ for the fast degrees of freedom satisfies an effective Schrödinger equation. The emergence of a Schrödinger equation connects the formalism of quantum geometrodynamics to quantum physics in curved spacetimes. This is shown explicitly in **chapter 7** in the context of quantum cosmology, where the inflationary power spectra are derived directly from the WDW equation. How this Schrödinger picture relates to the conventional Heisenberg picture presented in **chapter 4** is shown in **appendix C**.

The semiclassical time is sometimes called WKB time or Tomonaga-Schwinger time.

Note that a systematic semiclassical expansion of the WDW equation would expand S and ψ in powers of λ. One can continue this expansion to arbitrary precision. This leads to corrections to the Hamilton-Jacobi equation and the effective Schrödinger equation. The corrections for the Schrödinger equation are due to the influence of the slowly evolving background geometry on the fast degrees of freedom and can therefore naturally be associated with quantum gravitational effects. This was calculated for Einstein-Hilbert gravity with a minimally coupled scalar field in [95], and will be generalised to scalar-tensor theories in **chapter 6**. The influence of quantum gravity on the inflationary power spectra was first calculated in [109,110] for Einstein-Hilbert gravity and will be generalised for scalar-tensor theories in **chapter 7**.

5.4 UNITARITY IN QUANTUM GEOMETRODYNAMICS

One of the main complications in the canonical approach to quantum gravity is the apparent lack of a Hilbert space structure. For example, the indefinite and second-order structure of the WDW equation can result in negative probabilities. It is therefore not obvious how the wave function can be assigned a probabilistic interpretation. Such an interpretation is necessary in order for the effective Schrödinger equation **(5.19)** to be interpreted as the emergence of quantum field theory in a curved spacetime. Since the effective Schrödinger equation arises in the semiclassical approximation of the WDW equation it is therefore natural to expect that a probabilistic interpretation of quantum geometrodynamics can be defined at least semiclassically. It is appropriate to investigate this before proceeding.

Consider the full WDW equation **(5.12)**. All degrees of freedom are treated equally, so no assumption about semiclassical behaviour has been made. It can be verified by substitution that a solution to the WDW equation generates a current $\mathcal{J}^A$ that is conserved in superspace. To wit, one obtains

$$\mathcal{J}^A = \tfrac{1}{2}\mathrm{i}\mathcal{M}^{AB}\left(\Psi^\dagger\nabla_A\Psi - \Psi\nabla_A\Psi^\dagger\right), \qquad \nabla_A\mathcal{J}^A = 0. \tag{5.20}$$

One can use the current $\mathcal{J}^A$ to construct conserved probabilities. To do so, one can imagine hypersurfaces Σ in superspace. The probability density $\mathrm{d}P$ for the system to be in a particular configuration on this hypersurface is then

$$\mathrm{d}P = \mathcal{J}^A \mathrm{d}\Sigma_A, \tag{5.21}$$

where $\mathrm{d}\Sigma_A$ is a normal hypersurface element. This probability is conserved and invariant under coordinate transformations on superspace. It can be seen from (5.20) that this probability is not positive definite; if it is positive in some region of superspace for some wave function Ψ, then it is negative for the wave function $\Psi' = \Psi^\dagger$. It is therefore necessary to consider under which conditions, if any, it is possible to construct a probability density that is non-negative.

If, for the sake of illustration, all degrees of freedom are semiclassical, the Hamilton-Jacobi formalism defines a family of classical trajectories through the function S (which is defined through (5.13)). The dynamics of the different degrees of freedom are then parametrised by the semiclassical time t_s through (5.14). This semiclassical time, and therefore the function S, is defined wherever the semiclassical approximation holds. From this is clear that the hypersurfaces Σ cannot be arbitrary. The classical trajectories are determined by (5.14), and from (5.13) and (5.20) it follows that the current $\mathcal{J}^A$ is proportional to the velocities $\mathrm{D}_{t_\mathrm{s}} Q^A$, which are tangent vectors along the classical trajectories. It follows that an appropriate choice of hypersurfaces Σ has the property that the trajectories in configuration space intersect the Σ exactly once. If they intersect more than once, it follows from (5.21) that the probability is indefinite. Negative probabilities are therefore a result of a wrong choice of hypersurfaces. A straightforward way to ensure that the trajectories intersect the hypersurfaces once and only once is to identify them with surfaces of constant semiclassical time. A hypersurface normal is then proportional to $\nabla_A S$. Since $\mathcal{J}^A$ is proportional to the classical velocities (5.14), it is similarly proportional to $\nabla_A S$. It then follows that

$$\mathrm{d}P = \mathcal{J}^A \mathrm{d}\Sigma_A \propto \nabla^A S \nabla_A S$$

has a constant sign, which can be chosen to be positive. These two different choices of hypersurfaces, where the classical trajectories intersect, are sketched in figure 5.1.

The emergence of a positive probability from the semiclassical expansion of the WDW equation is not much different if some of the degrees of freedom are quantum. If there is at least one semiclassical coordinate of configuration space, there is exists a family of classical trajectories that flows through configuration space. Therefore, there is still a notion of time, and one can still define equal-time hypersurfaces on which probability is well-defined [111].

However, the situation is complicated when one considers regions of configuration space where none of the degrees of freedom are semiclassical. It is then no longer obvious if one can define equal-time hypersurfaces such that the wave function can be interpreted in a probabilistic way. However, since it can be expected that quantum

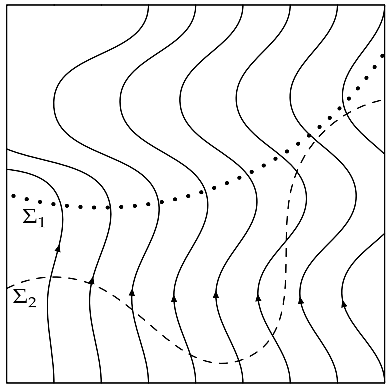

Figure 5.1 Different choices of hypersurfaces. Σ_1 is an acceptable hypersurface, whereas Σ_2 leads to negative probabilities.

fluctuations become small when the universe is large, it seems reasonable that at least some of the degrees of freedom must be semiclassical [111]. When at least one semiclassical degree of freedom is present the semiclassical Hamilton-Jacobi equation arises naturally, as was illustrated in **section 5.3.1**.

5.5 EXAMPLE: MINIMALLY COUPLED THEORIES

The above formalism is perhaps best understood by means of an example. In anticipation of the next chapters, it is therefore instructive to consider the explicit semiclassical expansion of the WDW equation for general relativity with a minimally coupled scalar field. The approach closely follows reference [95], and will be generalised in **chapter 6**. The issue of operator ordering will be neglected for the sake of illustration.

The gravitational field is described by the action S_{ADM} from **section 2.6**. This action is in the canonical form **(5.1)** if the gravitational potential is identified with $-\kappa\sqrt{\gamma}^{(s)}R$. The kinetic term depends on the extrinsic curvature $\tilde{K}_{ab}$. A minimally coupled scalar field ϕ with a potential V is described by the action **(3.2)** with $U = \kappa$ and $G = 1$. The model as a whole is therefore described by the action

$$\begin{aligned} S &= S_{\text{ADM}} + S_\phi, \text{ with} \\ S_{\text{ADM}} &= \int_M \kappa\sqrt{\gamma}\left[{}^{(s)}R + \tilde{K}_{ab}\tilde{K}^{ab} - \tilde{K}^2\right] N\mathrm{d}t\mathrm{d}^3x, \\ S_\phi &= \int_M \sqrt{\gamma}\left[\tfrac{1}{2}(\mathrm{D}_t\,\phi)^2 - \tfrac{1}{2}\,\mathrm{D}_a\,\phi\,\mathrm{D}^a\,\phi - V(\phi)\right] N\mathrm{d}t\mathrm{d}^3x. \end{aligned} \tag{5.22}$$

From this it can readily be found that

$$\mathrm{D}_t\, q^A = \begin{pmatrix} 2\tilde{K}_{ab} \\ \mathrm{D}_t\,\phi \end{pmatrix}, \qquad \mathcal{M}_{AB} = \sqrt{\gamma}\begin{pmatrix} \frac{1}{2}\kappa G^{abcd} & 0 \\ 0 & 1 \end{pmatrix},$$

where $G^{abcd} \equiv \gamma^{a(c}\gamma^{d)b} - \gamma^{ab}\gamma^{cd}$ is an object known as the *DeWitt metric*. It corresponds to the metric on the subspace of the spatial metrics γ_{ab}. The canonical momenta and Hamiltonian can be found following the procedure from **section 5.1**. The momenta are

$$p_\gamma^{ab} = \kappa\sqrt{\gamma}G^{abcd}\tilde{K}_{cd},$$
$$p_\phi = \mathrm{D}_t\,\phi,$$

while the Hamiltonian is

$$H = \int_{\mathcal{M}} N\left(\kappa^{-1}G_{abcd}p_\gamma^{ab}p_\gamma^{cd}/\sqrt{\gamma} + \tfrac{1}{2}p_\phi^2/\sqrt{\gamma} - \kappa\sqrt{\gamma}{}^{(s)}R + \sqrt{\gamma}V(\phi)\right)\mathrm{d}^3x$$
$$+ \int_{\mathcal{M}} N^a(-2\gamma_{ab}\,\mathrm{D}_c\,p_\gamma^{bc} + p_\phi\,\mathrm{D}_a\,\phi)\,\mathrm{d}^3x,$$

from which the Hamiltonian and momentum constraints can readily be identified. The constraints can now be quantised by the assumption of the canonical commutation relations

$$\left[\gamma_{ab}(\mathbf{x}), p_\gamma^{cd}(\mathbf{y})\right] = \mathrm{i}\delta^{(3)}(\mathbf{x}-\mathbf{y})\delta^c_{(a}\delta^d_{b)}, \qquad \left[\phi(\mathbf{x}), p_\phi(\mathbf{y})\right] = \mathrm{i}\delta^{(3)}(\mathbf{x}-\mathbf{y}),$$

with all other commutators equal to zero. In the absence of operator ordering ambiguities the momentum operator takes the form of the ordinary functional derivative $p_\gamma^{ab} = -\mathrm{i}\delta/\delta\gamma_{ab}$. The quantised momentum constraint is therefore

$$-2\gamma_{ab}\,\mathrm{D}_c\,\frac{\delta\Psi}{\delta\gamma_{bc}} + \mathrm{D}_a\,\phi\frac{\delta\Psi}{\delta\phi} = 0,$$

while the WDW equation is

$$-\frac{G_{abcd}}{\kappa\sqrt{\gamma}}\frac{\delta^2\Psi}{\delta\gamma_{ab}\delta\gamma_{cd}} - \frac{1}{2\sqrt{\gamma}}\frac{\delta^2\Psi}{\delta\phi^2} - \kappa\sqrt{\gamma}{}^{(s)}R\Psi + \sqrt{\gamma}V\Psi = 0. \tag{5.23}$$

The WDW equation is of the form **(5.16)**, where the small parameter can now be identified as $\lambda = \kappa^{-1}$. The spatial metrics can therefore be identified as the slow variables and the scalar field as the fast variable, in accordance with the discussion of **section 5.3.1**.

The structure of the WDW equation invites an ansatz for the wave functional of the form

$$\Psi = \psi(\gamma,\phi)\exp[\mathrm{i}S(\gamma)]. \tag{5.24}$$

Substitution of this ansatz into **(5.23)** results in the two separate equations **(5.18.a)** and **(5.18.b)**, which read explicitly

$$- \tfrac{1}{2}\mathrm{i}\kappa \frac{G_{abcd}}{\sqrt{\gamma}} \frac{\delta^2 S}{\delta\gamma_{ab}\delta\gamma_{cd}} + \tfrac{1}{2}\kappa \frac{G_{abcd}}{\sqrt{\gamma}} \frac{\delta S}{\delta\gamma_{ab}} \frac{\delta S}{\delta\gamma_{cd}} - \kappa^{-1}\sqrt{\gamma}^{(\mathrm{s})}R = 0, \tag{5.25.a}$$

$$- \tfrac{1}{2}\mathrm{i}\kappa \frac{G_{abcd}}{\sqrt{\gamma}} \frac{\delta S}{\delta\gamma_{ab}} \frac{\delta\psi}{\delta\gamma_{cd}} - \tfrac{1}{2}\kappa \frac{G_{abcd}}{\sqrt{\gamma}} \frac{\delta^2\psi}{\delta\gamma_{ab}\delta\gamma_{cd}} + \frac{1}{2\sqrt{\gamma}} \frac{\delta^2\psi}{\delta\phi^2} + \sqrt{\gamma}V(\phi)\psi = 0. \tag{5.25.b}$$

The separation function f has been set to zero.

The parameter λ occurs in both **(5.25.a)** and **(5.25.b)**, and can be used as a natural expansion parameter. It is readily seen that, in order to find an approximate solution to the WDW equation **(5.23)**, one must first solve **(5.25.a)**. The solution S can be substituted into **(5.25.b)** to find the solution ψ.

5.5.1 *Expansion of the heavy sector*

The function S can be decomposed as a power series in λ:

$$S = \lambda^{-1}S_0 + S_1 + \lambda S_2 + \ldots, \tag{5.26}$$

where the dots indicate terms of higher power in λ. The semiclassical expansion of the WDW equation divides the relation between quantum mechanics and general relativity into different levels, which can be placed in a relative hierarchy. This hierarchy is reflected in the semiclassical expansion. The lowest order (λ^{-1}) concerns the classical dynamics of general relativity. The second-to-lowest order (λ^{0}) concerns quantum mechanics in an external, classical background. The order obtained after that (λ^{1}) concerns interactions between the quantised degrees of freedom and the gravitational field. It is at this order that the quantum gravitational corrections can be expected to occur. One can therefore truncate the semiclassical expansion at terms of order λ in order to get an estimate of the leading quantum gravitational effects. This will be done in what follows.

Substitution of **(5.26)** into **(5.25.a)** then results in a polynomial in λ, of which each coefficient must vanish separately. This results in a family of equations, which can be solved iteratively.

Order $\mathcal{O}(\lambda^{-1})$: At this order in the expansion one obtains the Hamilton-Jacobi equation for general relativity:

$$\frac{G_{abcd}}{2\sqrt{\gamma}} \frac{\delta S_0}{\delta\gamma_{ab}} \frac{\delta S_0}{\delta\gamma_{cd}} - \sqrt{\gamma}^{(\mathrm{s})}R = 0. \tag{5.27}$$

It was discussed in **section 5.3** that the Hamilton-Jacobi equation gives rise to a semiclassical notion of time. A time derivative $\mathrm{D}_{t_\mathrm{s}}$ can be defined to be

$$\mathrm{D}_{t_\mathrm{s}} = \frac{G_{abcd}}{\sqrt{\gamma}} \frac{\delta S_0}{\delta\gamma_{ab}} \frac{\delta}{\delta\gamma_{cd}}. \tag{5.28}$$

It can be concluded that S_0 describes a classical spacetime geometry. The semiclassical expansion enables one to go beyond this classical regime, and allows one to compute quantum corrections to the classical background geometry.

Order $\mathcal{O}(\lambda^0)$*:* At this order in the expansion one obtains an equation for S_1.

$$\frac{G_{abcd}}{\sqrt{\gamma}}\left(\frac{\delta S_0}{\delta\gamma_{ab}}\frac{\delta S_1}{\delta\gamma_{cd}} - \tfrac{1}{2}\mathrm{i}\frac{\delta^2 S_0}{\delta\gamma_{ab}\delta\gamma_{cd}}\right) = 0. \tag{5.29}$$

This equations has the solution $S_1 = \mathrm{i}\log\Delta_1$, where Δ_1 satisfies

$$\frac{G_{abcd}}{\sqrt{\gamma}}\frac{\delta}{\delta\gamma_{ab}}\left(\Delta^{-2}\frac{\delta S_0}{\delta\gamma_{cd}}\right) = 0.$$

This is the continuity equation for the first order WKB correction for general relativity.

Order $\mathcal{O}(\lambda^1)$*:* With **(5.28)** this equation can be written to be

$$\mathrm{D}_{t_\mathrm{s}} S_2 = \frac{G_{abcd}}{2\sqrt{\gamma}}\left(\mathrm{i}\frac{\delta^2 S_1}{\delta\gamma_{ab}\delta\gamma_{cd}} - \frac{\delta S_1}{\delta\gamma_{ab}}\frac{\delta S_1}{\delta\gamma_{cd}}\right).$$

The replacement of S_1 by Δ_1 then results in

$$\mathrm{D}_{t_\mathrm{s}} S_2 = \frac{G_{abcd}}{2\sqrt{\gamma}}\left(\frac{1}{\Delta_1^2}\frac{\delta\Delta_1}{\delta\gamma_{ab}}\frac{\delta\Delta_1}{\delta\gamma_{cd}} - \frac{1}{\Delta_1}\frac{\delta^2\Delta_1}{\delta\gamma_{ab}\delta\gamma_{cd}}\right).$$

This is the second order WKB correction for general relativity.

5.5.2 *Expansion of the fast sector*

The semiclassical solution S can be substituted into **(5.25.b)**. This results in an evolution equation for ψ. The truncation of S at the leading term S_0 yields

$$\mathrm{i}\,\mathrm{D}_{t_\mathrm{s}}\psi = -\frac{1}{2\sqrt{\gamma}}\frac{\delta^2\psi}{\delta\phi^2} + \sqrt{\gamma}V\psi \equiv \mathcal{H}_\phi\psi.$$

This is the Schrödinger equation for ψ, as was found in **section 5.3.1**. It describes a quantised scalar field in a curved classical space.

The systematic expansion of the WDW equation leads to subleading contributions in S that go beyond the Schrödinger equation, in contrast to the qualitative discussion of **section 5.3.1**. These contributions lead to a coupling between the semiclassical background geometry and the scalar field ϕ, and can be ascribed to quantum gravitational effects.

The inclusion of these subleading terms in the semiclassical expansion of S into **(5.25.b)** leads to

$$\mathrm{i}\,\mathrm{D}_{t_\mathrm{s}}\,\psi = \mathcal{H}_\phi\psi + \tfrac{1}{2}\lambda\frac{G_{abcd}}{\sqrt{\gamma}}\left(\frac{\delta^2\psi}{\delta\gamma_{ab}\delta\gamma_{cd}} + \mathrm{i}\frac{\delta S_1}{\delta\gamma_{ab}}\frac{\delta\psi}{\delta\gamma_{cd}}\right). \tag{5.30}$$

There are new structures compared to the uncorrected Schrödinger equation. It is worthwhile to note that these extra structures originate entirely from the heavy degrees of freedom, and therefore indicate a non-trivial influence of the gravitational field on the light degrees of freedom. It is also worthwhile to mention that this equation is effective only in the sense that the expansion of S was truncated at the linear order term in λ. It is therefore straightforward to see how these corrections can be extended to arbitrary orders of λ.

5.5.3 *Unitarity in the semiclassical expansion*

It can be expected that a semiclassical notion of probability can be defined, as (5.30) arises as an extension of the uncorrected Schrödinger equation. It would then be possible to define a Schrödinger inner product for the light degree of freedom whenever this notion holds. That is, for given solutions Ψ and Φ of the WDW equation, it must be possible to define an inner product of the form

$$\langle\Psi|\Phi\rangle = \int \bar{\Psi}\Phi\,\mathrm{d}\phi.$$

The full WDW equation does not allow such a structure, and it can be expected that the conventional Schrödinger product breaks down once quantum gravitational effects become relevant. In particular, the correction terms in (5.30) have, in the past, been subject to considerable discussion, as their presence has been associated with violations of unitarity. This could either indicate that the formalism is wrong, or that the effective Schrödinger inner product breaks down. The later case would imply that it would have to be replaced by a more general inner product. This section therefore concludes with a brief consideration of the construction of a probability measure in superspace.

It was discussed in **section 5.4** that a solution Ψ to the WDW equation (5.23) gives rise to a conserved current $\mathcal{J}^A$. From this current a probability density $\mathrm{d}P$ can be defined according to (5.21). One has to ensure that the hypersurfaces of (5.21) are chosen properly, such that the probability is positive definite. This can be guaranteed by considering the hypersurfaces Σ of constant S_0. A normal vector n_A to these surfaces can be found to be

$$n_A = \varepsilon N \nabla_A S_0,$$

where the Hamilton-Jacobi equation (5.27) fixes the normalisation εN to be

$$N = \left(\varepsilon\nabla_A S_0 \nabla^A S_0\right)^{-1/2}, \qquad \varepsilon = \mathrm{sgn}(\sqrt{\gamma}^{(\mathrm{s})}R).$$

Coordinates x^B on the hypersurface can be constructed through the integral curves

$$\frac{\delta Q^A}{\delta x^B} = T^A{}_B,$$

where the $T^A{}_B$ are a linear independent set of vector fields that satisfy $n_A T^A{}_B = 0$ and $T^A{}_B T^B{}_C = \delta^A_C$. The metric on Σ is then

$$m_{AB} \equiv \mathcal{M}_{CD} T^C{}_A T^D{}_B.$$

The surface element $\mathrm{d}\Sigma$ then follows from (2.24):

$$\mathrm{d}\Sigma = \varepsilon n_A \sqrt{m} \prod_B \mathrm{d}x^B.$$

It is assumed in this example that space is not flat. It was argued in **chapter 4** that this assumption does not hold in inflationary cosmology. This issue will be addressed in **chapter 7**.

The components of the current can be found by substitution of the solution Ψ of the WDW equation into the definition of the current. With the ansatz (5.24) it is found that

$$\mathcal{J}_{ab} = \lambda \frac{G_{abcd}}{\sqrt{\gamma}} \left(2|\psi|^2 \frac{\delta S}{\delta \gamma_{cd}} + \mathrm{i}\psi \frac{\delta \psi^\dagger}{\delta \gamma_{cd}} - \mathrm{i}\psi^\dagger \frac{\delta \psi}{\delta \gamma_{cd}} \right),$$

$$\mathcal{J}^\phi = \mathrm{i} \left(\psi \frac{\delta \psi^\dagger}{\delta \phi} - \psi^\dagger \frac{\delta \psi}{\delta \phi} \right).$$

The probability density is therefore

$$\mathrm{d}P = \mathcal{J}^A \mathrm{d}\Sigma_A = \lambda N \left[|\psi|^2 \operatorname{Re}(\mathrm{D}_{t_\mathrm{s}} S) - \operatorname{Im}(\bar{\psi} \mathrm{D}_{t_\mathrm{s}} \psi) \right] \sqrt{m} \prod_A \mathrm{d}x^A.$$

This expression can be combined with (5.26), (5.27), (5.29) and (5.30) to give a formal expression for the conserved probability up to linear order in λ:

$$P = \int_\Sigma \lambda \bar{\psi} \left[1 + \lambda N^2 \mathcal{H}_\phi \right] \psi N^{-1} \sqrt{m} \prod_A \mathrm{d}x^A.$$

The probability provides an inner product between two states Ψ and Φ as a natural generalisation of the Schrödinger-type inner product:

$$\langle \Psi | \Phi \rangle = \int \bar{\Psi} \left[1 + \lambda N^2 \mathcal{H}_\phi \right] \Phi N^{-1} \sqrt{m} \prod_A \mathrm{d}x^A. \tag{5.31}$$

This inner product is conserved (by virtue of (5.21)) and it is positive definite (the spectrum of $\mathcal{H}_\phi$ is positive). It can be seen that, up to a normalisation factor originating from the heavy degrees of freedom, this inner product reduces to the familiar Schrödinger-type inner product on the space of the light degree of freedom at the level of the uncorrected Schrödinger equation, which corresponds to the formal limit $\lambda \to 0$. Furthermore, new structures appear at the level of the uncorrected Schrödinger equation. It is therefore no longer possible to define a Schrödiner-type inner product on the space of the light degree of freedom alone. Instead, one has to consider all the configuration space variables—both light and heavy.

It bears repeating that the results derived here are valid only when the semiclassical approximation holds. The construction relies on the existence of the Hamilton-Jacobi equation, without which it is not clear whether it is possible to define appropriate hypersurfaces on which a notion of probability can be defined. However, this question goes outside the scope of this work. The next chapters focus exclusively on the semiclassical regions of superspace, in which **(5.31)** is perfectly valid.

6 QUANTUM GEOMETRODYNAMICS IN SCALAR-TENSOR THEORIES

THE GOAL OF THIS CHAPTER is perform a semiclassical expansion of the Wheeler-DeWitt (WDW) equation for scalar-tensor theories. The motivation for this is two-fold. First, the semiclassical expansion of the WDW equation for Einstein-Hilbert gravity with a minimally coupled scalar field was performed in reference [95]. The quantum gravitational corrections were found to be suppressed by inverse powers of the square of the Planck mass. Scalar-tensor theories replace the Planck mass with a dynamical scalar field. It may therefore be expected that for certain scalar-tensor models the quantum gravitational corrections are enhanced relative to the minimally coupled theories. Second, it was shown in reference [95] that the effective Schrödinger equation, which arises in the semiclassical expansion of the WDW equation for minimally coupled theories, is independent of the operator ordering ambiguities that were discussed in **chapter 5**. It is natural to ask whether such a result would still hold in scalar-tensor theories, where the scalar field can be identified with a dynamical Planck mass.

The semiclassical expansion of the WDW equation for minimally coupled theories relied on the presence of the Planck mass as an expansion parameter. Scalar-tensor theories generally do not have a manifest presence of energy scales that can be used to serve as a natural expansion parameter. In fact, the non-minimal coupling of the scalar field to gravity induces a derivative coupling between the scalar field and the gravitational degrees of freedom, which prevents a direct application of the expansion scheme. This technical difficulty is addressed here by the use of the correspondence of the theory in the Jordan frame (JF) and the Einstein frame (EF). It is found that a large non-minimal coupling can have strong effects on the quantum gravitational correction terms. These effects are briefly discussed in the context of the specific model of Higgs inflation.

The chapter is structured as follows: the Arnowitt-Deser-Misner (ADM) decomposition of a general scalar-tensor theory is presented in **section 6.1**. The theory contains an explicit non-minimal coupling in the JF parametrisation. The classical theory is formulated in the Hamiltonian framework in **section 6.2**. Hence the explicit momentum and Hamiltonian constraints are derived. The constraints are subsequently quantised in **section 6.3**. The quantisation then gives rise to the WDWequation for scalar-tensor theories. In **section 6.4**, the WDW equation is expressed in the EF parametrisation, where it becomes diagonal, and perform the weighting required for the application of the Born-Oppenheimer approximation scheme. Finally, the weighted WDW operator can be re-expressed in the JF parametrisation. In **section 6.5**, the semiclassical approximation will be calculated up to the order where the first quantum gravitational corrections arise and discuss the impact of the non-minimal coupling. The results are summarised and listed with a brief discussion of possible applications **section 6.6**.

Furthermore, some technical details, that were not crucial for the main points of this chapter, were relegated to the appendices. Detailed expressions for objects related to the geometry of configuration space are presented in **appendix A**.

The action functional of a general scalar-tensor theory in four dimensions, with a scalar field ϕ non-minimally coupled to gravity, can be parametrised in terms of three arbitrary functions $U(\phi)$, $G(\phi)$ and $V(\phi)$:

$$S[g,\phi] = \int_{\mathcal{M}} \left(UR - \tfrac{1}{2}G\nabla_\mu\phi\nabla^\mu\phi - V\right)\sqrt{-g}\,\mathrm{d}^4X. \tag{6.1}$$

Here, U is the non-minimal coupling, G parametrises a non-canonically normalised kinetic term and V is the scalar field potential. It is assumed that the manifold $\mathcal{M}$ is globally hyperbolic and endowed with the 4-dimensional metric $g_{\mu\nu}$ and a metric compatible affine connection ∇_μ. The signature of $\mathcal{M}$ is taken to be Lorentzian. The Riemannian curvature $R^\rho{}_{\sigma\mu\nu}$ is defined by

$$R^\rho{}_{\sigma\mu\nu}v^\sigma = [\nabla_\mu, \nabla_\nu]v^\rho, \qquad v \in T\mathcal{M}.$$

A point X in $\mathcal{M}$ can be described by local coordinates X^μ. In order to express the action **(6.1)** in the Hamiltonian formalism the 4-dimensional ambient space $\mathcal{M}$ is foliated by a one-parameter family of 3-dimensional hypersurfaces Σ_t of constant time t. Thus, the hypersurfaces Σ_t are the level surfaces of a globally defined smooth scalar time field t. The gradient of t defines a natural unit covector field

$$n_\mu \equiv -\frac{\nabla_\mu t}{\sqrt{-g^{\mu\nu}\nabla_\mu t\nabla_\nu t}}, \qquad g^{\mu\nu}n_\mu n_\nu = -1.$$

At each point in Σ_t, the normal vector field

$$n^\mu = g^{\mu\nu}n_\nu,$$

is orthogonal to Σ_t and allows an orthogonal decomposition of tensor fields with respect to n^μ. In particular, the ambient metric decomposes as

$$g_{\mu\nu} = \gamma_{\mu\nu} - n_\mu n_\nu.$$

Here, $\gamma_{\mu\nu}$ is the tangential part of $g_{\mu\nu}$, which implies $\gamma_{\mu\nu}n^\mu = 0$. The hypersurfaces Σ_t can be considered as the embeddings of an intrinsically 3-dimensional manifold $\widehat{\Sigma}_t$ into the ambient space $\mathcal{M}$. A point x in $\widehat{\Sigma}$ can be described by local coordinates x^a. Thus, the 4-dimensional coordinate $X^\mu = X^\mu(t, x^a)$ can be parametrised in terms of the time field t and the 3-dimensional coordinates x^a. The change of X^μ with respect to t and x^a can be described by the coordinate one-form

$$\mathrm{d}X^\mu = \frac{\partial X^\mu(t,x)}{\partial t}\mathrm{d}t + \frac{\partial X^\mu(t,x)}{\partial x^a}\mathrm{d}x^a = t^\mu \mathrm{d}t + e^\mu{}_a\mathrm{d}x^a, \tag{6.2}$$

where the time vector field t^μ and the soldering form $e^\mu{}_a$ as

$$t^\mu \equiv \frac{\partial X^\mu(t,x^i)}{\partial t} \equiv N\,n^\mu + N^\mu, \qquad e_a{}^\mu \equiv \frac{\partial X^\mu(t,x)}{\partial x^a}. \tag{6.3}$$

This orthogonal decomposition shows the component $N = N(t, x)$ in the normal direction n^μ, which is called the *lapse function*, and the components $N^\mu = N^\mu(t, x)$ tangential to Σ_t, which form the *shift vector*. The soldering form $e^\mu{}_a$ can be thought of as a tangential vector with respect to the μ index, which implies that $e^\mu{}_a n_\mu = 0$, and a 3-dimensional vector with respect to the a index. It can be used to pull back tangential tensors in $\mathcal{M}$ to tensors in $\widehat{\Sigma}_t$. These properties can be summarised for the spatial metric and shift vector as

$$\gamma_{ab} \equiv e^\mu{}_a e^\nu{}_b \gamma_{\mu\nu}, \qquad N^a \equiv e_\mu{}^a N^\mu,$$
$$\delta^\mu_\nu = e^\mu{}_a e_\nu{}^a, \qquad \delta^a_b = e^a{}_\mu e^\mu{}_b.$$

The ambient space coordinate one-form (6.2) can therefore be written as

$$\mathrm{d}X^\mu = N n^\mu \mathrm{d}t + e_a{}^\mu \left(N^a \mathrm{d}t + \mathrm{d}x^a\right).$$

The ambient space line element then acquires the ADM form

$$\begin{aligned} \mathrm{d}s^2 &= g_{\mu\nu} \mathrm{d}X^\mu \mathrm{d}X^\mu \\ &= \left(-N^2 + \gamma_{ab} N^a N^b\right) \mathrm{d}t^2 + 2\gamma_{ab} N^a \mathrm{d}x^b \mathrm{d}t + \gamma_{ab} \mathrm{d}x^a \mathrm{d}x^b. \end{aligned} \tag{6.4}$$

The volume element can be decomposed as

$$\sqrt{-g} = N\sqrt{\gamma}. \tag{6.5}$$

The affine connection D_a on $\widehat{\Sigma}_t$ that is compatible with the metric γ_{bc} defines the 3-dimensional curvature

$${}^{(\mathrm{s})}R^a{}_{bcd} v^b = [\mathrm{D}_c, \mathrm{D}_d]\, v^a, \qquad v \in T\widehat{\Sigma}_t.$$

The relation between the 4-dimensional Ricci scalar R and the 3-dimensional Ricci scalar ${}^{(\mathrm{s})}R$ is given by the Gauss-Codazzi equation (2.18)

$$R = {}^{(\mathrm{s})}R - \left(\tilde{K}^2 - \tilde{K}_{ab}\tilde{K}^{ab}\right) - 2\left(\mathrm{D}_b + \tilde{a}_b\right)\tilde{a}^b + 2\left(\mathrm{D}_t + \tilde{K}\right)\tilde{K}. \tag{6.6}$$

Here, ${}^{(\mathrm{s})}R$ is the intrinsic 3-dimensional curvature, which is calculated from the induced metric γ_{ab}, while $\tilde{K}_{ab}$ is the extrinsic curvature and $\tilde{K}$ its trace:

$$\tilde{K}_{ab} \equiv \tfrac{1}{2} N^{-1}\left[\partial_0 \gamma_{ab} - \left(\mathcal{L}_{\mathbf{N}}\gamma\right)_{ab}\right] = \tfrac{1}{2}\,\mathrm{D}_t\,\gamma_{ab}\,, \qquad K \equiv \gamma^{ab}\tilde{K}_{ab}.$$

The covariant reparametrisation invariant time derivative is defined by

$$\mathrm{D}_t \equiv N^{-1}\left(\partial_0 - \mathcal{L}_{\mathbf{N}}\right), \tag{6.7}$$

where $\mathcal{L}_{\mathbf{N}}$ is the Lie derivative along the spatial shift vector $\mathbf{N} = N^a\partial_a$. Finally, the decomposition (6.6) involves the acceleration vector a_b, which is defined as

$$a_b \equiv \mathrm{D}_b \log N. \tag{6.8}$$

The action (6.1) can be expressed in terms of intrinsic 3-dimensional tensors via (6.5)-(6.8):

$$S[N, \mathbf{N}, \gamma, \phi] = \int L \, \mathrm{d}t = \int \mathcal{L} \, \mathrm{d}t\mathrm{d}^3x. \tag{6.9}$$

Here, L and $\mathcal{L}$ are the Lagrangian and the Lagrangian density, respectively. Up to boundary terms, the Lagrangian density for (6.1) is explicitly given by

$$\mathcal{L} = N\sqrt{\gamma}\Big[UG^{abcd}\tilde{K}_{ab}\tilde{K}_{cd} + U\,{}^{(\mathrm{s})}R - 2U_1\tilde{K}\,\mathrm{D}_t\,\phi + 2\Delta U$$
$$+ \tfrac{1}{2}G\left(\mathrm{D}_t\,\phi\,\mathrm{D}_t\,\phi - \mathrm{D}_a\,\phi\,\mathrm{D}^a\,\phi\right) - V\Big].$$

Here, $\Delta \equiv -\gamma^{ab}\,\mathrm{D}_a\,\mathrm{D}_b$ denotes the 3-dimensional Laplacian. The tensor G^{abcd} is the *DeWitt metric* G^{abcd} which together with its inverse G_{abcd} is defined by

In the literature one frequently encounters the DeWitt metric with an additional density factor $\sqrt{\gamma}$.

$$G^{abcd} \equiv \gamma^{a(c}\gamma^{d)b} - \gamma^{ab}\gamma^{cd}, \qquad G_{abcd} \equiv \gamma_{a(c}\gamma_{d)b} - \tfrac{1}{2}\gamma_{ab}\gamma_{cd}, \tag{6.10}$$

which satisfy

$$G_{abkl}G^{klcd} = \delta^{cd}_{ab} \equiv \delta^{c}_{(a}\delta^{d}_{b)} \equiv \tfrac{1}{2}\left(\delta^c_a\delta^d_b + \delta^c_b\delta^d_a\right).$$

In addition, recall that derivatives of a function $f(\phi)$ with respect to its argument are denoted by

$$f_n(\phi) \equiv \frac{\partial^n f(\phi)}{\partial\phi^n}.$$

6.2 CANONICAL FORMULATION AND HAMILTONIAN CONSTRAINT

It is convenient to use a compact notation for the dynamical configuration space variables q^A and their velocities $\partial_0 q^A$, where the superindex A labels the corresponding components

Note that the lapse function and shift vector are not dynamical quantities.

$$(q^A) = \begin{pmatrix} \gamma_{ab} \\ \phi \end{pmatrix}, \qquad (\partial_0 q^A) = \begin{pmatrix} \partial_0\gamma_{ab} \\ \partial_0\phi \end{pmatrix}.$$

In this compact notation the Lagrangian density in (6.9) takes the form

$$\mathcal{L} = \tfrac{1}{2}\partial_0 q^A \mathcal{M}_{AB}\,\partial_0 q^B + \dots,$$

where the configuration space metric $\mathcal{M}_{AB}$ can be read off from the terms quadratic in the velocities and the dots indicate lower order time derivatives terms. In components $\mathcal{M}_{AB}$ has the block matrix structure

$$(\mathcal{M}_{AB}) = \frac{\sqrt{\gamma}}{N}\begin{pmatrix} \frac{U}{2}G^{abcd} & -U_1\gamma^{ab} \\ -U_1\gamma^{cd} & G \end{pmatrix}. \tag{6.11}$$

Note the somewhat unorthodox inclusion of the inverse lapse function into the definition of the configuration space metric (6.11). In principle, time reparametrisation invariance suggests that with each factor of time t a factor of the lapse function should be associated, such as the inverse powers of N in the covariant time derivative (6.7). Similarly, one would associate a factor of N with the time differential $\mathrm{d}t$ in (6.9). The inclusion of the lapse function in (6.11) will become clear in section 6.4.2, where the transition between two particular parametrisations of the fields is discussed. In general, the configuration space is considered formally as a differentiable manifold. A list of the associated geometrical objects is provided in appendix A.

In terms of the covariant time derivative (6.7) the Lagrangian density (6.9) acquires the compact form

$$\mathcal{L} = \tfrac{1}{2}N^2 \, \mathrm{D}_t q^A \mathcal{M}_{AB} \, \mathrm{D}_t q^B - \mathcal{P}, \tag{6.12}$$

where D_t acts componentwise on the q^A. The potential $\mathcal{P}$ is defined as

$$\begin{aligned} \mathcal{P} &\equiv \mathcal{P}_\gamma + \mathcal{P}_\phi, \\ \mathcal{P}_\gamma &\equiv N P_\gamma \equiv -N\sqrt{\gamma} U \left[{}^{(s)}R + 2U^{-1}\Delta U + \tfrac{3}{2}\,\mathrm{D}_a \log U \, \mathrm{D}^a \log U \right], \\ \mathcal{P}_\phi &\equiv N P_\phi \equiv N\sqrt{\gamma} \left[\tfrac{1}{2} s^{-1}\, \mathrm{D}_a \phi \, \mathrm{D}^a \phi + V \right], \end{aligned}$$

The tensor γ_{ab}, when viewed as metric field is defined with covariant spatial indices, despite its contravariant superindex when viewed as a configuration space coordinate.

where the suppression function s is defined to be

$$s \equiv \frac{U}{GU + 3U_1^2}. \tag{6.13}$$

The momenta can be calculated directly from (6.12):

$$p_A = \frac{\partial \mathcal{L}}{\partial(\partial_t q^A)} = N \mathcal{M}_{AB} \, \mathrm{D}_t q^B. \tag{6.14}$$

In components, the momenta read

$$(p_A) = \begin{pmatrix} p_\gamma^{ab} \\ p_\phi \end{pmatrix} = \sqrt{\gamma} \begin{pmatrix} U G^{abcd} \tilde{K}_{cd} - U_1 \gamma^{ab} \, \mathrm{D}_t \phi \\ -2U_1 \gamma^{ab} \tilde{K}_{ab} + G \, \mathrm{D}_t \phi \end{pmatrix}.$$

Relation (6.14) can be inverted to yield

$$\mathrm{D}_t q^A = N^{-1} \mathcal{M}^{AB} p_B,$$

where the inverse of the configuration space metric $\mathcal{M}^{AB}$ is defined by

$$\mathcal{M}_{AC}\mathcal{M}^{CB} = \delta_A^B, \qquad (\delta_A^B) = \begin{pmatrix} \delta_{ab}^{cd} & 0 \\ 0 & 1 \end{pmatrix}.$$

Here, the (δ_B^A) denote the components of the identity matrix on the configuration space. The components of the inverse configuration space metric $\mathcal{M}^{AB}$ are given explicitly by

The inverse metric exists whenever γ_{ab} is regular and s is defined.

$$(\mathcal{M}^{AB}) = \frac{N}{\sqrt{\gamma}} \begin{pmatrix} 2U^{-1}G_{abcd} + s\left(\frac{U_1}{U}\right)^2 \gamma_{ab}\gamma_{cd} & -s\frac{U_1}{U}\gamma_{ab} \\ -s\frac{U_1}{U}\gamma_{cd} & s \end{pmatrix}. \tag{6.15}$$

The Hamiltonian density $\mathcal{H}$ is obtained by the Legendre transformation of **(6.12)**:

$$\mathcal{H} = p_A \partial_t q^A - \mathcal{L} = \tfrac{1}{2} p_A \mathcal{M}^{AB} p_B + \mathcal{P} + p_A \, \mathcal{L}_{\mathbf{N}} \, q^A. \tag{6.16}$$

The total Hamiltonian is given by the spatial integral of **(6.16)**

$$H \equiv \int \mathcal{H} \, \mathrm{d}^3 x \equiv \int (N\mathcal{H}_\perp + N^a \mathcal{H}_a) \, \mathrm{d}^3 x. \tag{6.17}$$

The total Hamiltonian **(6.17)** is constrained to vanish due to the 4-dimensional diffeomorphism invariance of **(6.1)**. The constraint character becomes manifest in the last equality of **(6.17)**, where $\mathcal{H}$ was written as the sum of the Hamiltonian constraint $\mathcal{H}_\perp$ and the momentum constraint $\mathcal{H}_a$ together with the lapse function N and shift vector N^a. The latter two act as Lagrange multipliers. Explicitly, the constraints are given by

$$\mathcal{H}_\perp = \frac{1}{U\sqrt{\gamma}} G_{abcd} \, p_\gamma^{ab} p_\gamma^{cd} + P_\gamma + \frac{1}{2} \frac{s}{\sqrt{\gamma}} \left(p_\phi - \frac{U_1}{U} p_\gamma \right)^2 + P_\phi, \tag{6.18}$$

$$\mathcal{H}_a = -2\gamma_{a(b} \, \mathrm{D}_{c)} \, p_\gamma^{bc} + p_\phi \, \mathrm{D}_a \, \phi. \tag{6.19}$$

Note that $p_\gamma = \gamma_{ab} \, p_\gamma^{ab}$ denotes the trace of the gravitational momentum. The momentum constraint $\mathcal{H}_a$ is the generator of 3-dimensional diffeomorphism, while the dynamical evolution is controlled by the Hamiltonian constraint $\mathcal{H}_\perp$. The expressions **(6.18)** and **(6.19)** coincide with those obtained in reference [81].

6.3 QUANTUM THEORY AND WHEELER-DEWITT EQUATION

As mentioned in the **chapter 5**, it is unclear whether a Hilbert space structure is a prerequisite for canonical quantum gravity. However, in semiclassical regions of superspace one can still expect an auxiliary underlying Hilbert space with Schrödinger type inner product

$$\langle \Phi | \Psi \rangle \equiv \int \bar{\Phi}[q] \sqrt{|\mathcal{M}|} \Psi[q] \, \mathrm{d}q. \tag{6.20}$$

Here, the wave functional $\Psi[q] = \langle q | \Psi \rangle$ corresponds to the Schrödinger representation of the state $|\Psi\rangle$ and $\mathcal{M}$ is the determinant of the configuration space metric $\mathcal{M}_{AB}$. Note that the naive definition **(6.20)** involves the integration over all configurations q^A, which includes the unphysical ones, as **(6.20)** is not an inner product on the space of the solutions to the constraints [92]. Related to the inner product is the question of unitarity. This is a complicated problem in the context of quantum gravity and can be discussed at various levels [97]. In particular, if quantum theory is a universal concept, a probabilistic interpretation would require a unitary evolution at the most fundamental level, which includes the gravitational degrees of freedom. However, this chapter focusses on semiclassical regions of superspace, in which an effective probabilistic interpretation holds. The fundamental questions above are therefore beyond the scope of this work.

In the quantum theory, the conjugated phase space variables q^A and p_B are promoted to operators $\hat{q}^A$ and $\hat{p}_B$, which satisfy the canonical commutator relations

$$[\hat{q}^A, \hat{p}_B] = \mathrm{i}\, \delta^A_B, \qquad [\hat{q}^A, \hat{q}^B] = [\hat{p}_A, \hat{p}_B] = 0. \tag{6.21}$$

In the position space representation, the position operator $\hat{q}^A$ acts multiplicatively, while the momentum operator $\hat{p}_B$ acts as a derivative operator

$$\hat{p}_A \equiv -\mathrm{i}\, \mathcal{M}^{-1/4} \frac{\delta}{\delta q^A} \mathcal{M}^{1/4}.$$

This representation of the momentum operator is formally self-adjoint with respect to the inner product (6.20) and satisfies the canonical commutation relation (6.21) [112]. The operator versions of the classical constraints (6.18) are defined by replacements of the classical phase space variables by their quantum operators:

$$\hat{\mathcal{H}}_\perp \equiv \mathcal{H}_\perp(\hat{q}, \hat{p}), \qquad \hat{\mathcal{H}}_a \equiv \mathcal{H}_a(\hat{q}, \hat{p}).$$

This procedure is ambiguous due to factor ordering problems, which arise because of (6.21). In particular, for the transition from the classical Hamiltonian constraint

$$\mathcal{H}_\perp = N^{-1} \left(\tfrac{1}{2} p_A \mathcal{M}^{AB} p_B + \mathcal{P} \right) \tag{6.22}$$

to the quantum Hamiltonian constraint, this factor ordering ambiguity can be traced back to the non-commutativity of the configuration space metric with the momentum operator:

$$[\mathcal{M}_{AB}(\hat{q}), \hat{p}_C] \neq 0.$$

The factor ordering ambiguity does not affect the principal part of the Hamiltonian constraint operator $\hat{\mathcal{H}}_\perp$—only its lower derivative terms. It can be partially addressed by the adoption of the covariant Laplace-Beltrami factor ordering, which effectively corresponds to the replacement the quadratic form $p_A \mathcal{M}^{AB}(q) p_B$ in (6.22) by the symmetric combination

The Laplace-Beltrami ordering makes it is possible to add an additional configuration space curvature term to $\Box$. This introduces an arbitrary parameter in the potential.

$$\mathcal{M}^{-1/4} \hat{p}_A \mathcal{M}^{1/4}\, \mathcal{M}^{AB}\, \mathcal{M}^{1/4} \hat{p}_B \mathcal{M}^{-1/4} = -\mathcal{M}^{AB} \nabla_A \nabla_B \equiv -\Box. \tag{6.23}$$

Here, ordinary functional derivatives are abbreviated by δ_A and introduced the covariant functional derivatives ∇_A, defined with respect to the Christoffel connection of the configuration space metric:

$$\delta_A \Psi = \frac{\delta \Psi}{\delta q^A}, \qquad \Gamma^C_{AB} = \tfrac{1}{2} \mathcal{M}^{CD} \left(\delta_A \mathcal{M}_{DB} + \delta_B \mathcal{M}_{AD} - \delta_D \mathcal{M}_{AB} \right).$$

The Hamiltonian constraint operator with the factor ordering (6.23) can be written compactly as

$$\hat{\mathcal{H}}_\perp = N^{-1}\left(-\tfrac{1}{2}\Box + \mathcal{P}\right). \tag{6.24}$$

Equation (6.24) will be referred to as the WDW operator. With the geometrical quantities that are provided in (A.1)-(A.6) the explicit form of the Laplace-Beltrami operator (6.22) and hence the WDW operator (6.24) can be found to be

$$\begin{aligned}\hat{\mathcal{H}}_\perp = &-\frac{1}{\sqrt{\gamma}U}G_{abcd}\frac{\delta^2}{\delta\gamma_{ab}\delta\gamma_{cd}} - \frac{7}{8\sqrt{\gamma}U}\gamma_{ab}\frac{\delta}{\delta\gamma_{ab}} + P_\gamma \\ &-\frac{1}{2}\frac{s}{\sqrt{\gamma}}\mathcal{D}^2 - \frac{1}{4}\frac{s}{\sqrt{\gamma}}\left(\frac{s_1}{s} - \frac{3}{2}\frac{U_1}{U}\right)\mathcal{D} + P_\phi.\end{aligned} \tag{6.25}$$

Here, $\mathcal{D}$ defines the combined derivative

$$\mathcal{D} \equiv \frac{\delta}{\delta\phi} - \frac{U_1}{U}\gamma_{ab}\frac{\delta}{\delta\gamma_{ab}}. \tag{6.26}$$

Note, that since the Laplace-Beltrami operator $\Box$ and the potential $\mathcal{P}$ are both proportional to N, the explicit form of $\hat{\mathcal{H}}_\perp$ is independent of the lapse function.

In the quantisation prescription for constrained systems, proposed by Dirac [96], the quantum constraints are implemented by the demand that physical states are annihilated by the quantum constraint operators. The implementation of the momentum constraint operator $\hat{\mathcal{H}}_a$ ensures that the wave functional $|\Psi\rangle$ is invariant under 3-dimensional diffeomorphisms:

$$\hat{\mathcal{H}}_a|\Psi\rangle = 0. \tag{6.27}$$

The configuration space modulo the 3-dimensional diffeomorphisms is called *superspace* [100, 101]. The implementation of the Hamiltonian constraint operator $\hat{\mathcal{H}}_\perp$, which governs the quantum dynamics of the wave functional Ψ, leads to the WDW equation

$$\hat{\mathcal{H}}_\perp|\Psi\rangle = 0. \tag{6.28}$$

At this point it is appropriate to comment on the factors of $\delta^{(3)}(0)$, which have been suppressed in the considerations so far. The spatial 3-dimensional delta function $\delta^{(3)}(x^a, x^b)$ is assumed to be a scalar bi-density with zero weight at the first argument and unit weight at the second argument, where x^a and x^b are the corresponding spatial coordinates. Given the fundamental identity

$$\frac{\delta q^A(x^a)}{\delta q^B(x^b)} = \delta^A_B\,\delta^{(3)}(x^a, x^b),$$

it is readily seen that, since one has to calculate functional derivatives that act on local background quantities at the same point ($x^a = x^b$), the WDW equation contains factors of $\delta^{(3)}(0)$ [94]. The singular factors of undifferentiated delta functions at the same point have to be regularised. It was suggested in reference [94] to adopt a regularisation scheme where field operators at the same point can be freely commuted, which effectively corresponds to $\delta^{(3)}(0) = 0$. This means that the kinetic part of the WDW operator **(6.25)** is reduced to its principal (highest derivative) part, where all functional derivatives only act on the wave functional—not on local background coefficients. At this point no regularisation scheme is adopted. Instead, all factors of $\delta^{(3)}(0)$ are carried through the calculation but, but their explicit occurrence is suppressed for notational reasons. Explicit factors $\delta^{(3)}(0)$ can be restored easily at each step by dimensional considerations.

6.4 WEIGHTING AND TRANSITION BETWEEN FRAMES

It is, in general, difficult to find an exact solution $\Psi[\gamma, \phi]$ to the WDW equation **(6.28)**. Moreover, the quantum theory obtained by the naive definition of the inner product and the Dirac quantisation scheme is not complete [97, 113]. In addition, an exact solution to the WDW equation requires suitable boundary conditions. These boundary conditions are the main subject of quantum cosmology. The most prominent and physically best motivated proposals among the many choices seem to be the no-boundary and tunneling conditions [74, 104–106, 114–117]. One might even extract predictions and consistency equations from quantum cosmology if given a sensible definition of a probability measure, as can be seen, for example, in the references [118–121].

This work does not attempt to find an exact solution to the full WDW equation, but aims to construct the semiclassical branch of the wave functional Ψ, which is limited to a restricted region in configuration space. The approach uses a combined Born-Oppenheimer-WKB-type approximation in order to perform a systematic expansion of the full WDW equation [93, 95, 122–127], which allows the extraction of the first quantum gravitational correction terms [95].

6.4.1 *Born-Oppenheimer approximation*

The Born-Oppenheimer approximation is well known in quantum mechanical multiparticle systems and, and is based on a distinction between heavy degrees of freedom Q with mass m_Q and light degrees of freedom q with mass m_q [128]. Heavy and light are to be understood in terms of the terminology introduced in **chapter 5**. The difference in the characteristic mass scales, expressed in terms of the dimensionless parameter

$$\lambda \equiv \frac{m_q}{m_Q} \ll 1, \tag{6.29}$$

implies that the heavy and light degrees of freedom vary on different characteristic time scales and might therefore be interchangeably referred to as the slow and fast degrees of freedom, respectively. In a more abstract context, λ represents a formal parameter which can be used to implement the distinction between background (slow) and fluctuation (fast) degrees of freedom and which can be set to one after the expansion has been performed. The distinction between slow and fast variables motivates a product ansatz for the total wave function:

$$\Psi(Q,q) = \chi(Q)\,\psi(Q;q).$$

Here $\chi(Q)$ is the wave function for the slow degrees of freedom Q, for which a subsequent WKB approximation is performed. In contrast, the wave function $\psi(Q;q)$ for the fast degrees of freedom q is treated as fully quantum and depends only parametrically on the Q variables.

Practically, the semiclassical expansion can be systematically performed by the following operations. First, the distinction between heavy and light degrees of freedom can be implemented by the introduction of different relative weight factors for the individual terms in the Hamilton operator by the rescaling of each factor of m_Q by a power of λ:

$$H(Q,q) \to H_\lambda(Q,q), \tag{6.30}$$

where the weighted Hamiltonian H_λ has the schematic structure

$$H_\lambda(Q,q) = \tfrac{1}{2}\lambda m_Q^{-1}P_Q^2 + V(Q) + \tfrac{1}{2}m_q^{-1}p_q^2 + W(Q,q).$$

Here P_Q and $V(Q)$ are the momentum and self-interaction potential of the heavy variables Q, while p_q and $W(Q,q)$ are the momentum and potential of the fast variables q. The latter includes the self-interaction among the q's as well as the interaction between the q's and the Q's. Second, in addition to the weighting of the Hamilton operator (6.30), the wave function $\Psi(q,Q)$ in the form of a formal power series in λ:

$$\Psi(Q;q) = \exp\mathrm{i}\left[\lambda^{-1}S_0(Q;q) + S_1(Q,q) + \lambda S_2(Q,q) + \ldots\right]. \tag{6.31}$$

In the context of quantum geometrodynamics, the semiclassical expansion can be obtained by the insertion of the weighted Hamilton operator $\hat{H}_\lambda$ together with the ansatz (6.31) into the WDW equation (6.28). The result can be collected as a polynomial in λ, of which the coefficient of each term must vanish separately. One then obtains a family of equations for S_0, S_1, S_2 and so on. The wave functional Ψ can be reconstructed to the accuracy given by the respective order in λ as these equations are solved iteratively. For the system of a scalar field ϕ that is minimally coupled to gravity in 4 dimensions, the slow and fast degrees of freedom Q and q are then associated with the spatial metric γ_{ab} and the scalar field ϕ, respectively. The weighting procedure then corresponds to the association of the expansion parameter λ with each occurrence of the inverse of half the squared Planck mass $\kappa = \frac{1}{2}M_{\mathrm{P}}$.

At the highest order of the expansion $\mathcal{O}(\lambda^{-2})$, one finds that $S_0(\gamma)$ is a function of γ_{ab} only. This is consistent with the association of the degrees of freedom in γ_{ab} as the slow variable. At the next order $\mathcal{O}(\lambda^{-1})$, one obtains an equation for $S_0(\gamma)$. Since the equation is of the Hamilton-Jacobi type, one recovers in a natural way the notion of a semiclassical time from the timeless WDW equation. At order $\mathcal{O}(\lambda^0)$ one obtains an equation for $S_1(\gamma, \phi)$, which can be formulated as a Schrödinger equation for the light scalar field degree of freedom ϕ, where the time parameter t is identified with the semiclassical time and is effectively provided by the slowly changing background geometry [93,95,122–127]. At order $\mathcal{O}(\lambda)$ one finds an equation for $S_2(\gamma, \phi)$, which incorporates the first quantum gravitational correction terms [95].

This analysis is now extended to the case of a scalar-tensor theory. The action **(6.1)** is rather general, as it involves three arbitrary functions $U(\phi)$, $G(\phi)$ and $V(\phi)$, and covers almost all single field inflationary models in cosmology for different classes of U, G and V. There are several differences compared to the minimally coupled scalar field. First, the non-minimal coupling to gravity U leads to a derivative coupling between the matter and gravitational degrees of freedom, which result in a non-diagonal WDW operator **(6.25)**. Thus, a clear separation of slow and fast degrees of freedom as for the minimally coupled case is no longer available. Second, in contrast to the minimally coupled case, no constant mass scale κ is present *a priori*. This make a straightforward application of the semiclassical expansion scheme difficult, as the Born-Oppenheimer approximation relies on a clear separation of slow and fast variables implemented in the WDW operator **(6.25)** by different powers of λ.

This problem is addressed by the following strategy. It is well known that the scalar-tensor theory **(6.1)** admits a classically equivalent parametrisation in the EF, which resembles the action of a scalar field minimally coupled to gravity. The transition to the EF is achieved by a particular field redefinition $(g, \phi) \to (\tilde{g}, \tilde{\phi})$, which involves a conformal transformation of the 4-dimensional metric field $g_{\mu\nu}$ and a non-linear field redefinition of the scalar field ϕ. This is explained in detail in **chapter 3**. In view of the ADM decomposition, the conformal transformation of the 4-dimensional metric $g_{\mu\nu}$ induces a corresponding conformal transformation of the geometrical fields N, N_a and γ_{ab} in the canonical theory. In terms of the EF variables, the WDW operator **(6.25)** becomes diagonal, which enables a clear weighting procedure. The distinction between heavy and light degrees of freedom can be implemented by the association of a power of λ with each inverse power of κ, owing to the presence natural mass scale κ. Once this weighting procedure has been performed in the EF the WDW operator **(6.25)** can be transformed back to the original JF field variables. The semiclassical expansion can then be performed as outlined before.

6.4.2 *Transition to the Einstein frame*

The reparametrisation of the field variables from the JF to the EF at the level of the 4-dimensional covariant Lagrangian **(3.13)** induces a corresponding transformation of the ADM variables in the canonical formalism

$$\tilde{N} = \sqrt{\frac{U}{\kappa}}N, \qquad \tilde{\gamma}_{ab} = \frac{U}{\kappa}\gamma_{ab}, \qquad \frac{\partial\tilde{\phi}}{\partial\phi} = \left(\frac{Us}{\kappa}\right)^{-\frac{1}{2}}. \tag{6.32}$$

No explicit factors of the shift vector appear in the formalism, as the covariant time derivative (6.7) renders the theory to the manifestly 3-dimensional diffeomorphism invariant. In terms of the abstract multicomponent configuration space variables q^A, the transformations (6.32) between the JF and EF can be described by the Jacobi matrices

$$\left(\frac{\partial q^A}{\partial \tilde{q}^B}\right) = \begin{pmatrix} \frac{U}{\kappa}\delta^{ab}_{cd} & -\frac{U_1}{U}\left(\frac{Us}{\kappa}\right)^{\frac{1}{2}}\gamma_{cd} \\ 0 & \left(\frac{Us}{\kappa}\right)^{\frac{1}{2}} \end{pmatrix}, \tag{6.33}$$

$$\left(\frac{\partial \tilde{q}^B}{\partial q^A}\right) = \begin{pmatrix} \frac{U}{\kappa}\delta^{cd}_{ab} & \frac{U_1}{U}\tilde{\gamma}_{ab} \\ 0 & \left(\frac{Us}{\kappa}\right)^{-\frac{1}{2}} \end{pmatrix}. \tag{6.34}$$

Note, however, that the transformations (6.32) do not simply correspond to a coordinate transformation on configuration space; one would have to transform the lapse function $N \to \tilde{N}$. The lapse function, in contrast to the q^A, is not a dynamical configuration space variable. Nevertheless, the inclusion of the lapse function in the definition of the configuration space metric (6.11) allows the transformation under (6.32) to be written in the standard covariant form (6.33), (6.34). It therefore describes an ordinary coordinate transformation on configuration space $q^A \to \tilde{q}^A$, provided that the lapse function is rescaled according to (6.32). The lapse function N is constant from the viewpoint of a true coordinate transformation $q^A \to \tilde{q}^A$. In contrast, the rescaling of the lapse function has to be taken into account for the transformation between the JF and EF (6.32). In particular, it becomes relevant once derivatives of the configuration space metric $\delta_A \mathcal{M}_{BC}$ are transformed from the JF to the EF. This generates additional terms, which will be necessary for the transformation of the Laplace-Beltrami operator from the JF to the EF parametrisation.

The momenta transform covariantly under (6.32):

$$\tilde{p}_A = \frac{\partial q^B}{\partial \tilde{q}^A} p_B. \tag{6.35}$$

The components of the EF momenta, expressed in terms of the JF momenta, read

$$(\tilde{p}_A) = \begin{pmatrix} \tilde{p}^{ab}_{\tilde{\gamma}} \\ \tilde{p}_{\tilde{\phi}} \end{pmatrix} = \begin{pmatrix} \frac{U}{\kappa} p^{ab}_{\gamma} \\ \left(\frac{Us}{\kappa}\right)^{1/2}\left(p_\phi - \frac{U_1}{U} p_\gamma\right) \end{pmatrix}.$$

Thus, beside multiplicative scaling factors, the EF scalar field momentum $\tilde{p}_{\tilde{\phi}}$ is a combination of the JF scalar field momentum p_ϕ and the trace of the JF metric momentum p_γ. This particular combination is a consequence of the original non-minimal coupling responsible for the derivative mixing between the metric and the scalar field degrees of freedom, which becomes manifest in the combined derivative operator (6.26). The field content is diagonal, in terms of the EF parametrisation $(\tilde{\gamma}_{ab}, \tilde{\phi})$. This can be seen explicitly from the diagonal EF configuration space metric $\tilde{\mathcal{M}}_{AB}$, which, together with its inverse $\tilde{\mathcal{M}}^{AB}$, can be obtained from the JF configuration space metric (6.11) by

$$\tilde{\mathcal{M}}_{AB}(\tilde{q}) = \frac{\partial q^C}{\partial \tilde{q}^A}\frac{\partial q^D}{\partial \tilde{q}^B}\mathcal{M}_{CD}(q), \qquad \tilde{\mathcal{M}}^{AB}(\tilde{q}) = \frac{\partial \tilde{q}^A}{\partial q^C}\frac{\partial \tilde{q}^B}{\partial q^D}\mathcal{M}^{CD}(q). \tag{6.36}$$

Note that $\tilde{\mathcal{M}}_{AB}$ and $\tilde{\mathcal{M}}^{AB}$ transform like ordinary tensors under (6.32). The explicit expressions for the configuration space metric and its inverse in coordinates $\tilde{q}$ read

$$\left(\tilde{\mathcal{M}}_{AB}\right) = \frac{\sqrt{\tilde{\gamma}}}{\tilde{N}}\begin{pmatrix} \frac{1}{2}\kappa\tilde{G}^{abcd} & 0 \\ 0 & 1 \end{pmatrix}, \qquad \left(\tilde{\mathcal{M}}^{AB}\right) = \frac{\tilde{N}}{\sqrt{\tilde{\gamma}}}\begin{pmatrix} 2\kappa^{-1}\tilde{G}_{abcd} & 0 \\ 0 & 1 \end{pmatrix}. \tag{6.37}$$

According to the transformations (6.35) and (6.36), the quadratic form in the kinetic part of the Hamiltonian transforms as a scalar under (6.32):

$$\tilde{p}_A\,\tilde{\mathcal{M}}^{AB}\,\tilde{p}_B = p_A\,\mathcal{M}^{AB}\,p_B.$$

Similarly, according to the general formula (3.9) and the transformation rules (6.32), the spatial Ricci scalar in the EF variables reads

$${}^{(s)}\tilde{R} = (U/\kappa)\left[{}^{(s)}R + 2U^{-1}\Delta U + \tfrac{3}{2}\,\mathrm{D}_a\log U\,\mathrm{D}^a\log U\right]. \tag{6.38}$$

According to (3.11), the EF scalar field potential and the spatial derivatives of the scalar field ϕ transform under (6.32) as

$$\tilde{\mathrm{D}}_a\tilde{\phi} = \frac{\partial\tilde{\phi}}{\partial\phi}\,\mathrm{D}_a\phi = \left(\frac{Us}{\kappa}\right)^{-\frac{1}{2}}\mathrm{D}_a\phi, \qquad \tilde{V} = \left(\frac{U}{\kappa}\right)^{-2}V. \tag{6.39}$$

Inspection of (6.38) and (6.39) leads to the conclusion that the potential transforms as a scalar under (6.32):

$$\tilde{\mathcal{P}} = \tilde{\mathcal{P}}_{\tilde{\gamma}} + \tilde{\mathcal{P}}_{\tilde{\phi}} = \mathcal{P}_\gamma + \mathcal{P}_\phi = \mathcal{P}.$$

Here, the potentials in the EF parametrisation are given by

$$\tilde{\mathcal{P}}_{\tilde{\gamma}} = \tilde{N}\tilde{\gamma}^{\frac{1}{2}}\kappa\,{}^{(s)}\tilde{R}, \qquad \tilde{\mathcal{P}}_{\tilde{\phi}} = \tilde{N}\tilde{\gamma}^{\frac{1}{2}}\left(\tfrac{1}{2}\tilde{\mathrm{D}}_a\tilde{\phi}\tilde{\mathrm{D}}^a\tilde{\phi} + \tilde{V}\right).$$

Finally, provided the wave functional Ψ transforms as a scalar $\tilde{\Psi}(\tilde{q}) = \Psi(q)$, the Laplace-Beltrami operator is seen to transform as a scalar under (6.32) as well:

$$\tilde{\Box}\tilde{\Psi}(\tilde{q}) = \tilde{\mathcal{M}}^{AB}(\tilde{q})\tilde{\nabla}_A\tilde{\nabla}_B\tilde{\Psi}(\tilde{q}) = \mathcal{M}^{AB}(q)\nabla_A\nabla_B\Psi(q) = \Box\Psi(q). \tag{6.40}$$

6.4.3 *Weighting of the Hamiltonian*

In analogy to the discussion of section 5.3.1, it is notationally convenient to decompose the configuration space metric $\mathcal{M}^{AB}$ in terms of submetrics M_{AB} and m_{AB}:

$$\tilde{\mathcal{M}}_{AB} = \tilde{N}^{-1}(\tilde{M}_{AB} + \tilde{m}_{AB}), \tag{6.41.a}$$

where the submetrics $\tilde{M}_{AB}$ and $\tilde{m}_{AB}$ are given by

$$\tilde{M}_{AB} = \kappa\sqrt{\tilde{\gamma}}\begin{pmatrix} \frac{1}{2}\tilde{G}^{abcd} & 0 \\ 0 & 0 \end{pmatrix}, \qquad \tilde{m}_{AB} = \sqrt{\tilde{\gamma}}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{6.41.b}$$

It follows from the discussion above that the metrics NM_{AB} and Nm_{AB} transform as under a frame transformations. Their components in the JF, in addition to the components of their inverses, have the following structure:

$$M_{AB} = U\sqrt{\gamma}\begin{pmatrix} \frac{1}{2}G^{abcd} & -\frac{U_1}{U}\gamma^{ab} \\ -\frac{U_1}{U}\gamma^{cd} & -3\left(\frac{U_1}{U}\right)^2 \end{pmatrix}, \qquad m_{AB} = \frac{\sqrt{\gamma}}{s}\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}. \tag{6.41.c}$$

The quantities M^{AB} and m^{AB} that follow from the decomposition of the inverse $\mathcal{M}^{AB}$ can straightforwardly be seen to have the JF components:

$$M^{AB} = \frac{2}{U\sqrt{\gamma}}\begin{pmatrix} G_{abcd} & 0 \\ 0 & 0 \end{pmatrix}, \quad m^{AB} = \frac{s}{\sqrt{\gamma}}\begin{pmatrix} \left(\frac{U_1}{U}\right)^2\gamma_{ab}\gamma_{cd} & -\frac{U_1}{U}\gamma_{ab} \\ -\frac{U_1}{U}\gamma_{cd} & 1 \end{pmatrix}. \tag{6.41.d}$$

The WDW operator in the EF parametrisation can, in terms of this decomposition, be found by application of the transformations **(6.32)**-**(6.36)** to **(6.24)** to be

$$\hat{\mathcal{H}}_\perp = \sqrt{\frac{U}{\kappa}}\left(-\tfrac{1}{2}\tilde{M}^{AB}\tilde{\nabla}_A\tilde{\nabla}_B + \tilde{P}_{\tilde{\gamma}} - \tfrac{1}{2}\tilde{m}^{AB}\tilde{\nabla}_A\tilde{\nabla}_B + \tilde{P}_{\tilde{\phi}}\right).$$

Note that, while the Hamiltonian **(6.17)** transforms as a scalar under **(6.32)**, the WDW operator $\hat{\mathcal{H}}_\perp$ transforms as a scalar density. The origin of this difference can be traced back to the inverse power of the lapse function in **(6.24)**. In the EF the field content is diagonal and a clear separation between the gravitational degrees of freedom $\tilde{\gamma}_{ab}$ and the scalar degree of freedom $\tilde{\phi}$ is possible. The natural choice is the identification of the gravitational variables $\tilde{\gamma}_{ab}$ with the slow variables and the scalar field $\tilde{\phi}$ with the fast variables, where the terms fast and slow are to be understood of the context of the Born-Oppenheimer approximation scheme. Therefore, in the EF there is a clear weighting scheme by associating with each factor of κ an inverse factor of the dimensionless expansion parameter λ:

$$\hat{\mathcal{H}}^\lambda_\perp = \sqrt{\frac{U}{\kappa}}\left(-\tfrac{1}{2}\lambda\tilde{M}^{AB}\tilde{\nabla}_A\tilde{\nabla}_B + \lambda^{-1}\tilde{P}_{\tilde{\gamma}} - \tfrac{1}{2}m^{AB}\tilde{\nabla}_A\tilde{\nabla}_B + \tilde{P}_{\tilde{\phi}}\right) \tag{6.42}$$

Note that the overall scaling factor is irrelevant for the weighting process, as only the relative weighting of terms in the Hamiltonian is important. The weighted WDW operator **(6.42)** can be transformed back to the JF and reads

$$\hat{\mathcal{H}}^\lambda_\perp = -\tfrac{1}{2}\lambda M^{AB}\nabla_A\nabla_B + \lambda^{-1}P_\gamma + \hat{\mathcal{H}}_s, \tag{6.43}$$

where, in anticipation of what is to come, the Hamiltonian for the light scalar degree of freedom $\hat{\mathcal{H}}_{\mathrm{s}}$ has been defined to be

$$\begin{aligned}\hat{\mathcal{H}}_{\mathrm{s}} &\equiv -\tfrac{1}{2}m^{AB}\nabla_A\nabla_B + P_\phi \\ &= -\frac{1}{2}\frac{s}{\sqrt{\gamma}}\mathcal{D}^2 - \frac{1}{4}\frac{s}{\sqrt{\gamma}}\left(\frac{s_1}{s} - \frac{3}{2}\frac{U_1}{U}\right)\mathcal{D} + P_\phi .\end{aligned} \tag{6.44}$$

Note that, in contrast to the minimally coupled case where the scalar field Hamiltonian is free of any factor ordering ambiguities, the covariant Laplace-Beltrami factor ordering (6.23) induces a dependence on the factor ordering in the scalar Hamiltonian (6.44), reflected by the terms linear in $\mathcal{D}$. These extra terms can be related directly to the presence of the non-minimal coupling and vanish for $U = \kappa$. Only part of the gravitational degrees of freedom (the scale part) in γ_{ab} mix with the scalar degrees of freedom ϕ and, according to the weighting scheme that was introduced in **section 6.4.3**, only these parts are treated as fully quantum. This is in contrast to the remaining gravitational degrees of freedom, which are treated as semiclassical.

6.5 SEMICLASSICAL APPROXIMATION

Substitution of (6.43) together with the semiclassical ansatz for the wave functional (6.31) into the WDW equation (6.28) results in a polynomial in λ. When the coefficient of each term term vanishes independently, one obtains a family of equations for S_0, S_1, S_2 and so on. The wave functional Ψ in (6.30) can be reconstructed within the given accuracy of the approximation by the truncation of (6.31) at a fixed order in λ. The resulting equations, of which the lowest order determines S_0, can be solved consecutively. In the following subsections, the equations are discussed separately at each order.

Order $\mathcal{O}(\lambda^{-2})$ At this order in the semiclassical expansion, one obtains the following equation for S_0:

$$\tfrac{1}{2}m^{AB}\nabla_A S_0 \nabla_B S_0 = 0. \tag{6.45}$$

With (6.26) and (6.41.d) it can be seen that this is equivalent to

$$\mathcal{D}S_0 = \frac{\delta S_0}{\delta\phi} - \gamma_{ab}\frac{U_1}{U}\frac{\delta S_0}{\delta\gamma_{ab}} = 0. \tag{6.46}$$

This implies that $S_0(\gamma, \phi) = S_0(\tilde{\gamma})$ is only a function of the particular combination

$$\tilde{\gamma}_{ab} = \frac{U}{\kappa}\gamma_{ab},$$

which is nothing but the metric in the EF parametrisation (6.32).

One particular consequence of (6.46) is that any contraction of $\nabla_A S_0$ with m^{AB} vanishes. This will be useful later in the semiclassical expansion.

Order $\mathcal{O}(\lambda^{-1})$ At this order in the semiclassical expansion, one obtains, after use of **(6.46)**, the Hamilton-Jacobi equation for S_0:

$$\tfrac{1}{2}M^{AB}\nabla_A S_0 \nabla_B S_0 + P_\gamma = 0. \tag{6.47}$$

From the discussion in **section 5.3** it then follows that **(6.47)** suggests the semiclassical WKB time t_s via

$$\mathrm{D}_{t_\mathrm{s}} \equiv M^{AB}\nabla_A S_0 \nabla_B. \tag{6.48}$$

In terms of this semiclassical time **(6.47)** manifestly acquires the structure of the Hamilton-Jacobi equation for S_0 [129]:

$$\tfrac{1}{2}\,\mathrm{D}_{t_\mathrm{s}} S_0 + P_\gamma = 0. \tag{6.49}$$

The exact WDW equation **(6.36)** is timeless. Therefore, the concept of time only emerges from the semiclassical expansion at the level of the Hamilton-Jacobi equation **(6.49)**, which, together with the momentum constraint $\hat{\mathcal{H}}_a \Psi = 0$, can be shown to be equivalent to the Einstein equations [122].

Thus, within the semiclassical approximation scheme, the flow of time is associated with the slowly changing background geometry S_0, which is adiabatically followed by the quantum states of the matter fields. The wave functional to order λ is simply given by

$$\Psi = \frac{1}{\Delta_0(\tilde{\gamma})},$$

where

$$\Delta_0(\tilde{\gamma}) \equiv \exp\left(-\mathrm{i}\lambda^{-1}S_0\right).$$

Order $\mathcal{O}(\lambda^{0})$ At order $\mathcal{O}(\lambda^0)$ one obtains an equation for S_1, which, when combined with the equation of the previous orders and the definition of semiclassical time **(6.48)**, can be written as

$$\begin{aligned} 0 &= P_\phi + \tfrac{1}{2}m^{AB}(\nabla_A S_1 \nabla_B S_1 - \mathrm{i}\nabla_A \nabla_B S_1) \\ &\quad + M^{AB}(\nabla_A S_0 \nabla_B S_1 - \tfrac{1}{2}\mathrm{i}\nabla_A \nabla_B S_0). \end{aligned} \tag{6.50}$$

The Born-Oppenheimer ansatz suggests that S_1 be split into a part $\sigma_1(\tilde{\gamma})$, which only depends on the background, and a part $\Sigma_1(\gamma, \phi)$, which cannot be reduced further:

$$S_1(\gamma,\phi) \equiv \sigma_1(\tilde{\gamma}) + \Sigma_1(\gamma,\phi). \tag{6.51}$$

There is no loss of generality in such a decomposition. Notice that it follows follows the approach in **section 5.3.1**, where the WDW equation was split into an equation that depends solely on the background metric M^{AB} and a part that depends, in addition, on the perturbations through the metric m^{AB}. One can demand that σ_1 satisfies the equation

$$\mathrm{D}_{t_\mathrm{s}}\sigma = \tfrac{1}{2}\mathrm{i}M^{AB}\nabla_A\nabla_B S_0. \tag{6.52}$$

Insertion of (6.51) and (6.52) into (6.50) results in an equation for Σ_1 alone:

$$\mathrm{D}_{t_\mathrm{s}}\Sigma_1 = \tfrac{1}{2}m^{AB}(\mathrm{i}\nabla_A\nabla_B\Sigma - \nabla_A\Sigma_1\nabla_B\Sigma_1) - P. \tag{6.53}$$

The wave functional ψ_1 up to this order in the expansion is

$$\psi_1 \equiv \exp\left(\mathrm{i}\Sigma_1\right). \tag{6.54}$$

Insertion of (6.54) into (6.53) results in a Schrödinger equation for ψ_1:

$$\mathrm{i}\,\mathrm{D}_{t_\mathrm{s}}\psi_1 = \mathcal{H}_\mathrm{s}\,\psi_1. \tag{6.55}$$

The Hamilton operator $\hat{\mathcal{H}}_\mathrm{s}$ is defined in (6.44). Moreover, σ_1 is related to the Van Vleck determinant Δ_1, which naturally arises in the WKB approximation:

$$\sigma_1 \equiv \mathrm{i}\log\Delta_1(\tilde{\gamma}). \tag{6.56}$$

Equation (6.55) is a Schrödinger equation for the light scalar degree of freedom where the emergent semiclassical time is controlled by the change of the geometry. The wave functional up to this order is given by

$$\Psi = \frac{\psi_1(\gamma,\phi)}{\Delta_0(\tilde{\gamma})\Delta_1(\tilde{\gamma})}.$$

At this level of the semiclassical expansion one can introduce a notion of unitarity for the light quantum degrees of freedom. In this case, unitary evolution of the light degrees of freedom could be defined as the condition

$$\mathrm{D}_{t_\mathrm{s}}\langle\psi_1,\psi_1\rangle_\phi = 0, \tag{6.57}$$

where, in contrast to (6.20), the inner product $\langle\cdot,\cdot\rangle_\phi$ extends only over the light degrees of freedom. Note that such a definition of unitarity can at best be a derived semiclassical one; its very definition (6.57) relies on the notion of a semiclassical time $\mathrm{D}_{t_\mathrm{s}}$ and the derived concept of a Hilbert space of states ψ_1 for the light degrees of freedom. It is be expected that, at the higher orders in the semiclassical expansion, the Klein-Gordon-type inner product has to be used, as discussed in **chapter 5**.

Order $\mathcal{O}(\lambda^1)$ At this order in the semiclassical expansion one finds the equation

$$\mathrm{D}_{t_\mathrm{s}}S_2 = \tfrac{1}{2}M^{AB}(\mathrm{i}\nabla_A\nabla_B S_1 - \nabla_A S_1\nabla_B S_1) + \tfrac{1}{2}m^{AB}(\mathrm{i}\nabla_A\nabla_B S_2 - 2\nabla_A S_1\nabla_B S_2).$$

The function S_2 can be decomposed in the same way as S_1 in (6.51):

$$S_2(\gamma,\phi) \equiv \sigma_2(\tilde{\gamma}) + \Sigma_2(\gamma,\phi).$$

In analogy to (6.56), one can define

$$\sigma_2(\tilde{\gamma}) \equiv \mathrm{i}\log\Delta_2$$

and choose σ_2 to be the solution of the equation

$$\mathrm{D}_{t_\mathrm{s}}\sigma_2 = -\tfrac{1}{2}M^{AB}\left(\frac{\nabla_A\nabla_B\Delta_1}{\Delta_1} - 2\frac{\nabla_A\Delta_1\nabla_B\Delta_1}{\Delta_1^2}\right). \tag{6.58}$$

The functional $\sigma_2(\tilde{\gamma})$ can be interpreted as the second order WKB factor for the heavy degrees of freedom in the Born-Oppenheimer approximation. What remains is an equation for the functional $\Sigma_2(\gamma, \phi)$:

$$\begin{aligned}\mathrm{D}_{t_\mathrm{s}}\Sigma_2 &= \tfrac{1}{2}m^{AB}\left(\mathrm{i}\nabla_A\nabla_B\Sigma_2 - 2\nabla_A\Sigma_1\nabla_B\Sigma_2\right)\\ &\quad + \tfrac{1}{2}M^{AB}\left(\mathrm{i}\nabla_A\nabla_B\Sigma_1 - \nabla_A\Sigma_1\nabla_B\Sigma_1 - 2\nabla_A\sigma_1\nabla_B\Sigma_1\right).\end{aligned} \tag{6.59}$$

In analogy to (6.54), one can define

$$\psi_2 \equiv \exp\left(\mathrm{i}\lambda\Sigma_2\right).$$

The wave functional up to this order of the expansion then reads

$$\Psi = \frac{\psi_1\psi_2}{\Delta_0\Delta_1\Delta_2}. \tag{6.60}$$

It is now possible to derive a Schrödinger equation that contains the first quantum gravitational corrections. The wave functional for the fast degrees of freedom can, up to this order in the semiclassical expansion, be written as

$$\psi = \psi_1\psi_2 = \exp\left(\mathrm{i}\Sigma_1 + \mathrm{i}\lambda\Sigma_2\right). \tag{6.61}$$

It is straightforward to see that, when terms of quadratic order in λ are neglected, the time derivative of the wave functional ψ can be written as

$$\mathrm{i}\,\mathrm{D}_{t_\mathrm{s}}\psi = \mathcal{H}_\mathrm{s}\psi - \tfrac{1}{2}\lambda\psi_1^{-1}M^{AB}\left(\nabla_A\nabla_B\psi_1 - 2\nabla_A\log\Delta_1\nabla_B\psi_1\right)\psi, \tag{6.62}$$

after use has been made of (6.53) and (6.59). This equation is a Schrödinger equation for the functional ψ, which includes correction terms indicated by the overall factor of λ.

A brief pause is in order for the results that were obtained up to now to be summarised. The semiclassical expansion of the WDW equation yielded the Hamiltonian-Jacobi equation (6.49) at order $\mathcal{O}(\lambda^{-1})$, which provides a definition semiclassical time (6.48). A Schrödinger equation (6.55) was found for the wave functional of the light degrees of freedom at order $\mathcal{O}(\lambda^0)$. The first quantum gravitational correction terms were encountered in the corrected Schrödinger equation (6.62) at order $\mathcal{O}(\lambda^1)$. Ultimately, the goal of this chapter is to derive the semiclassical branch (6.31) of the full wave functional, which includes these quantum gravitational corrections. The successive solution of the equations (6.49), (6.52), (6.53), (6.58) and (6.59) allows one to construct this wave functional (6.31) up to the required order. In principle, the analysis can be considered finished at that point.

However, in the following section it will be shown that (6.62) can be reformulated to obtain a clearer interpretation of the structure of the quantum gravitational correction terms. Moreover, since one of the main motivations of this work is to study the impact of the non-minimal coupling on the quantum gravitational correction terms, it is useful to express **(6.62)** in a from in which the corrections can straightforwardly be compared it to the corrections obtained for the minimally coupled case that was analysed in [95].

6.5.1 *Representation of the corrections*

The solutions of equations **(6.55)** and **(6.59)**, together with the solutions for S_0 and S_1, are sufficient to determine the wave functional up to order $\mathcal{O}(\lambda^1)$, as shown in **(6.60)**. The form of the correction terms in **(6.62)** is, however, not very illuminating. In order to write these terms in a more transparent form and to compare them with the result for the minimally coupled scalar field it is convenient to perform an orthogonal decomposition of the correction terms, following the analysis of reference [95].

The correction terms in **(6.62)** can be decomposed into contributions orthogonal and tangential to the surfaces of constant S_0. These surfaces are determined entirely by the heavy degrees of freedom, in the space determined by the submetric M_{AB}:

$$\mathcal{M}^{AB}\nabla_A S_0 \nabla_B S_0 = NM^{AB}\nabla_A S_0 \nabla_B = -2\mathcal{P}_\gamma,$$

where use has been made of **(6.45)**, **(6.48)** and the Hamilton-Jacobi equation **(6.49)** in the last step. A unit normal covector can therefore be defined as

$$n_A \equiv \varepsilon \frac{\nabla_A S_0}{\sqrt{\varepsilon \nabla_B S_0 \nabla^B S_0}},$$

where $\varepsilon = -\operatorname{sgn} \mathcal{P}_\gamma$. The corresponding unit normal vector n^A is then

$$n^A = \varepsilon \frac{\mathcal{M}^{AB}\nabla_B S_0}{\sqrt{\varepsilon \nabla_C S_0 \nabla^C S_0}}. \tag{6.63}$$

Note that this construction breaks down whenever $\mathcal{P}_\gamma = 0$. A unit normal vector cannot be defined in this case. The corrections should then be calculated from the form as they appear in equation **(6.62)**.

With the assumption that a unit normal vector can be defined, one can introduce projection operators normal and tangential to the surface of constant S_0:

$$\mathrm{P}_\perp{}^A{}_B = \varepsilon n^A n_B, \qquad \mathrm{P}_\parallel{}^A{}_B = \delta^A_B - \varepsilon n^A n_B.$$

The gradient of $\nabla_A \psi_1$ can then be decomposed as

$$\nabla_A \psi_1 = \mathrm{P}_\perp{}^B{}_A \nabla_B \psi_1 + \mathrm{P}_\parallel{}^B{}_A \nabla_B \psi_1.$$

The normal projection can be determined explicitly

$$\mathrm{P}_\perp{}^A{}_B \nabla_A \psi_1 = (-2\mathcal{P}_\gamma)^{-1}\mathcal{M}^{CD}\nabla_C S_0 \nabla_D \psi_1 \nabla_B S_0 = \frac{\mathrm{i}\hat{\mathcal{H}}_{\mathrm{s}}\psi_1}{2P_\gamma}\nabla_B S_0. \tag{6.64}$$

where the following identity was used:

$$\mathcal{M}^{AB}\nabla_C S_0 \nabla_D \psi_1 = N M^{AB} \nabla_A S_0 \nabla_B \psi_1 = N\,\mathrm{D}_{t_\mathrm{s}}\,\psi_1.$$

The components of $\nabla_A \psi_1$ that are tangential to the surfaces of constant S_0 can similarly be written down:

$$T_A = \mathrm{P}_{\parallel}{}^{B}{}_{A} \nabla_B \psi_1, \qquad M^{AB} T_A n_B = 0. \tag{6.65}$$

Note that T_{ab} is not specified explicitly the tangential contributions in the correction terms will not be needed [95].

In terms of the orthogonal decomposition the second derivative of ψ_1 can be found to be

$$\nabla_A \nabla_B \psi_1 = \tfrac{1}{2}\mathrm{i} P_\gamma^{-1} \hat{\mathcal{H}}_\mathrm{s} \psi_1 \,\nabla_A \nabla_B S_0 + \left[\tfrac{1}{2}\mathrm{i}\nabla_A \left(P_\gamma^{-1}\hat{\mathcal{H}}_\mathrm{s}\right)\psi_1 - \tfrac{1}{4}\mathrm{i}P_\gamma^{-1}\hat{\mathcal{H}}_\mathrm{s}\nabla_A \psi_1\right]\nabla_B S_0 \\ + \nabla_A T_B.$$

Repeated use of (6.64), (6.65) and contraction with M^{AB} then leads to

$$M^{AB}\nabla_A\nabla_B\psi_1 = \tfrac{1}{2}\mathrm{i}P_\gamma^{-1}\hat{\mathcal{H}}_\mathrm{s}\psi_1 M^{AB}\nabla_A\nabla_B S_0 + \tfrac{1}{2}\mathrm{i}P_\gamma^{-1}\nabla_B S_0 \hat{\mathcal{H}}_\mathrm{s} T_A + \nabla_A T_B \\ + \left[\tfrac{1}{2}\mathrm{i}\nabla_A\left(P_\gamma^{-1}\hat{\mathcal{H}}_\mathrm{s}\right)\psi_1 - \tfrac{1}{4}P_\gamma^{-1}\hat{\mathcal{H}}_\mathrm{s}\left(\nabla_A S_0 P_\gamma^{-1}\hat{\mathcal{H}}_\mathrm{s}\psi_1\right)\right]\nabla_B S_0. \tag{6.66}$$

Substitution of (6.64) and (6.66) into (6.62) allows the correction terms to be split into an orthogonal part B_n, and an tangential part B_t:

$$-\tfrac{1}{2}M^{AB}(\nabla_A\nabla_B\psi_1 - 2\nabla_A \log\Delta_1 \nabla_B\psi_1) = B_n + B_t.$$

The tangential part reads

$$B_t \equiv -\tfrac{1}{2}M^{AB}\left(\nabla_A T^A - 2\nabla_A \log\Delta_1\, T_B + \tfrac{1}{2}\mathrm{i}P_\gamma^{-1}\nabla_A S_0 \hat{\mathcal{H}}_\mathrm{s} T_B\right),$$

while the orthogonal part reads

$$B_n \equiv \tfrac{1}{4}\left[\tfrac{1}{2}M^{AB}\nabla_A S_0 P_\gamma^{-1}\hat{\mathcal{H}}_\mathrm{s}\left(\nabla_B S_0 P_\gamma^{-1}\hat{\mathcal{H}}_\mathrm{s}\psi_1\right) - \mathrm{i}\,\mathrm{D}_{t_\mathrm{s}}(P_\gamma^{-1}\hat{\mathcal{H}}_\mathrm{s})\psi_1\right].$$

Notice that derivatives of ψ_1 can, at this order in the semiclassical expansion, be replaced by derivatives of ψ:

$$\nabla_A \log\psi = \nabla_A \log\psi_1 + \mathcal{O}(\lambda).$$

The correction terms can thus be expressed in the form of a corrected Schrödinger equation for ψ. The normal contributions to the quantum gravitational corrections are determined by the previous orders of the expansion, while the tangential contributions are undetermined and therefore arbitrary. They can therefore be set to zero, as is done in reference [95]. The quantum gravitationally corrected Schrödinger equation can be then be written as

$$\mathrm{i}\,\mathrm{D}_{t_\mathrm{s}}\psi = \hat{\mathcal{H}}_\mathrm{s}\psi + \tfrac{1}{4}\lambda\left[\tfrac{1}{2}M^{AB}\nabla_A S_0 P_\gamma^{-1}\hat{\mathcal{H}}_\mathrm{s}\left(\nabla_B S_0 P_\gamma^{-1}\hat{\mathcal{H}}_\mathrm{s}\right) - \mathrm{i}\,\mathrm{D}_{t_\mathrm{s}}(P_\gamma^{-1}\hat{\mathcal{H}}_\mathrm{s})\right]\psi. \tag{6.67}$$

This result can be compared to the results for a minimally coupled scalar field, performed in reference [95]. There are several differences, which will be discussed in the remaining part of this section.

The last term in **(6.67)** has the same structure as in the minimally coupled case and has been associated with a unitarity violating term in [95]. The appearance of the unitarity violating term in **(6.67)** can be traced back to the use of the uncorrected Schrödinger equation **(6.55)** in the process of reformulating the result **(6.62)** into a form that resembles a Schrödinger equation for ψ. Moreover, the apparent unitarity violation is to be understood here at the semiclassical level in the sense of **(6.57)**. Since this semiclassical concept of unitarity can be at most an effective one, as it emerges from the semiclassical expansion itself, it is expected to break down once quantum gravitational corrections become relevant. Instead, one would have to resort to the Klein-Gordon formulation of unitarity that was introduced in **section 5.3**. If one insists on the effective Schrödinger probability, the unitarity violating terms in **(6.67)** can be dealt with by a formal absorption of these terms in a redefinition of the semiclassical time t_{s}. This is done, for example, in reference [130]. It can be seen that this would correspond to the inclusion of backreaction terms of the light degrees of freedom to the slow degrees of freedom. More precisely, these backreaction terms would modify the background S_0 and therefore the Hamilton-Jacobi equation **(6.49)**, which defines the semiclassical time **(6.48)**. In the context of a reduced minisuperspace model of a minimally coupled scalar field, the authors of reference [126] find that the inclusion of backreaction terms leads to a unitary semiclassical evolution. In the semiclassical expansion of the WDW equation for a minimally coupled scalar field, which was treated in reference [95], backreaction terms where neglected. Since one of the main motivations of this chapter is to determine the influence of the non-minimal coupling in scalar-tensor theories, the unitarity violating terms are similarly neglected.

The relevant quantum gravitational corrections can be found in **(6.67)**, which is the focus of the remaining discussion. This term is not of the same form as in the minimally coupled case in reference [95]. The reason for this is that the Hamiltonian $\hat{\mathcal{H}}_{\mathrm{s}}$ does not commute with $\nabla_A S_0$ and P_γ. This can be seen from **(6.44)**. The correction terms can be written in a form similar to those in reference [95], at the expense of introducing commutators. In particular, if $\nabla_A S_0$ is commuted through $\hat{\mathcal{H}}_{\mathrm{s}}$ one can make use of the Hamilton-Jacobi equation **(6.49)**. This is useful, as it eliminates all occurrences of $\nabla_A S_0$ from the corrected Schrödinger equation. It is straightforward to verify that

$$\begin{aligned}\tfrac{1}{2}M^{AB}\nabla_A S_0\hat{\mathcal{H}}_{\mathrm{s}}\left(\nabla_B S_0 P_\gamma^{-1}\hat{\mathcal{H}}_{\mathrm{s}}\psi\right) = & -P_\gamma\Bigg\{\hat{\mathcal{H}}_{\mathrm{s}}P_\gamma^{-1}\hat{\mathcal{H}}_{\mathrm{s}}\psi - \frac{s}{\sqrt{\gamma}}\frac{U_1}{U}\,\mathcal{D}P_\gamma^{-1}\hat{\mathcal{H}}_{\mathrm{s}}\psi \\ & -\frac{s}{2\sqrt{\gamma}}\left[\frac{(\sqrt{s}U_1)_1}{\sqrt{s}U} - \frac{3}{4}\left(\frac{U_1}{U}\right)^2\right]P_\gamma^{-1}\hat{\mathcal{H}}_{\mathrm{s}}\psi\Bigg\}.\end{aligned}$$

This can be inserted into **(6.67)**, after which the corrected Schrödinger equation for ψ acquires the form

$$\mathrm{i}\,\mathrm{D}_{t_\mathrm{s}}\,\psi = \left(1 - \lambda f(\phi) P_\gamma^{-1}\right)\hat{\mathcal{H}}_\mathrm{s}\psi$$
$$- \tfrac{1}{4}\lambda\left[\hat{\mathcal{H}}_\mathrm{s} P_\gamma^{-1}\hat{\mathcal{H}}_\mathrm{s} - \left[\delta^{(3)}(0)\right]\frac{s}{\sqrt{\gamma}}\frac{U_1}{U}\,\mathcal{D}\,P_\gamma^{-1}\hat{\mathcal{H}}_\mathrm{s} + \mathrm{i}\,\mathrm{D}_{t_\mathrm{s}}\left(P_\gamma^{-1}\hat{\mathcal{H}}_\mathrm{s}\right)\right]\psi. \quad (6.68)$$

Notice that explicit factors of $\delta^{(3)}(0)$ were restored and that the correction terms from the last line in **(6.68)** have been collected in the function $f(\phi)$:

$$f(\phi) = -\frac{\left[\delta^{(3)}(0)\right]^2}{32}\frac{s}{\sqrt{\gamma}}\left[4\frac{U_2}{U} + 2\frac{s_1 U_1}{sU} - 3\left(\frac{U_1}{U}\right)^2\right].$$

Thus, in contrast to the minimally coupled case, singular factors of $\delta^{(3)}(0)$ enter the final result from two different sources: firstly from the Laplace-Beltrami factor ordering **(6.23)** and secondly from the non-commutativity of derivatives in $\hat{\mathcal{H}}_\mathrm{s}$ with background quantities.

The adopting of a regularisation procedure as described in reference [94], where the $\delta^{(3)}(0)$ contributions are regularised to zero, effectively corresponds to the omission of all derivative terms that do not act on the wave functional. In this case, the matter Hamiltonian $\hat{\mathcal{H}}_\mathrm{s}$ can be commuted to the very right, with the result being

$$\mathrm{i}\,\mathrm{D}_{t_\mathrm{s}}\,\psi = \hat{\mathcal{H}}^\mathrm{P}_\mathrm{s}\psi - \tfrac{1}{4}\lambda\left[P_\gamma{-1}\left(\hat{\mathcal{H}}^\mathrm{P}_\mathrm{s}\right)^2 + \mathrm{i}\,\mathrm{D}_{t_\mathrm{s}}\left(P_\gamma^{-1}\hat{\mathcal{H}}^\mathrm{P}_\mathrm{s}\right)\right]\psi. \quad (6.69)$$

The kinetic term of $\hat{\mathcal{H}}_\mathrm{s}$ reduces to its principal part

$$\hat{\mathcal{H}}^\mathrm{P}_\mathrm{s} \equiv -\frac{s}{2\sqrt{\gamma}}\,\mathcal{D}^2 + P_\phi. \quad (6.70)$$

It is understood that the ϕ and γ_{ab} derivatives in $\mathcal{D}$ only act on the wave functional ψ in **(6.69)**. The form of the correction terms in **(6.68)** then features the same structure as in the minimally coupled case. The only difference is that the scalar Hamilton operator **(6.70)** replaces the matter Hamiltonian of the minimally coupled scalar field

$$\hat{\mathcal{H}}_\phi = -\frac{1}{2\sqrt{\gamma}}\frac{\delta^2}{\delta\phi^2} + P_\phi. \quad (6.71)$$

The difference between **(6.70)** and **(6.71)** consists of the generalised derivative operator $\mathcal{D}$ instead of the simple $\delta/\delta\phi$ and the overall factor s in the kinetic part of **(6.70)**.

The effects of these differences can be discussed in the context of the cosmological model of Higgs inflation, where ϕ is associated with the Standard Model Higgs boson and for which the arbitrary functions in the general scalar-tensor theory **(6.1)** acquire the particular form

$$U(\phi) = \tfrac{1}{2}\left(M_\mathrm{P}^2 + \xi\phi^2\right), \qquad G(\phi) = 1, \qquad V(\phi) = \tfrac{1}{4}\lambda\left(\phi^2 - v^2\right)^2. \quad (6.72)$$

Here, ξ is the non-minimal coupling constant, λ the quartic Higgs self-interaction and $v \approx 246\,\mathrm{GeV}$ the electroweak symmetry breaking scale. The form of the non-minimal coupling function $U(\phi)$ in **(6.72)** shows that the relevant parameter is the dimensionless combination

$$x \equiv \frac{\sqrt{\xi}\phi}{M_\mathrm{P}}.$$

The above results can be discussed in two relevant regions in configuration space, which correspond to the asymptotic regimes $x \ll 1$ and $x \gg 1$. For a weak non-minimal coupling and $x \ll 1$ the first term of U in **(6.72)** dominates and one would expect to recover the minimally coupled case. Indeed, for the functions **(6.72)**, the function s, defined in **(6.13)**, can be expressed in terms of x and ξ as

$$s = \frac{1 + x^2}{1 + (1 + 6\xi)x^2}. \tag{6.73}$$

Clearly, for small x, the function s tends to one as **(6.73)** reduces to

$$s = 1 + \mathcal{O}\left(x^2\right). \tag{6.74}$$

The derivative $\mathcal{D}$ becomes

$$\mathcal{D} = \frac{\delta}{\delta\phi} - \frac{x}{1 + x^2}\sqrt{\frac{4\xi}{M_\mathrm{P}}}\gamma_{ab}\frac{\delta}{\delta\gamma_{ab}}.$$

For $x \ll 1$, the γ_{ab} derivative is suppressed by $x\sqrt{\xi}/M_\mathrm{P}$:

$$\mathcal{D} = \frac{\delta}{\delta\phi} + \mathcal{O}(x). \tag{6.75}$$

Thus, in view of **(6.74)** and **(6.75)**, the scalar matter Hamilton operator **(6.70)** reduces to the matter Hamilton operator **(6.71)** for the minimally coupled scalar field in the limit of small x. This relation is in fact a required consistency condition, as one must recover the minimally coupled case in the limit of vanishing ξ.

The case of large x can be considered next. The function s reduces in this case to

$$s = \frac{1}{1 + 6\xi} + \mathcal{O}\left(x^{-2}\right) \approx \frac{1}{6\xi} \ll 1,$$

where a large non-minimal coupling ξ was assumed. Thus, for a strong non-minimal coupling ξ, the function s leads to a strong overall suppression of the kinetic terms in **(6.70)** in the regime of large x. Nevertheless, in this regime the γ_{ab} derivative in $\mathcal{D}$ is still suppressed by a factor of $\sqrt{\xi}/M_\mathrm{P}x$:

$$\mathcal{D} = \frac{\delta}{\delta\phi} + \mathcal{O}\left(x^{-1}\right). \tag{6.76}$$

Therefore, independent of the coupling parameter ξ, the metric derivatives in $\mathcal{D}$ are suppressed in both cases for $x \ll 1$ and $x \gg 1$. This behaviour can be traced back to the function $x/(1+x^2)$, which tends to zero for both small and large values of x and, for positive x, has a global maximum between the two asymptotic regimes at $x = 1$. The effect of the γ_{ab} derivatives is therefore the strongest for $\phi = M_{\mathrm{P}}/\sqrt{\xi}$, which corresponds to $x = 1$. The suppression of the kinetic term in **(6.70)** by the function s for a strong non-minimal coupling might be interpreted as the analogue of the suppression mechanism of the Higgs propagator found in the perturbative covariant approach to Higgs inflation [37,38,41,43,47–49].

6.6 CONCLUSION

This chapter considered the canonical quantisation of a general scalar-tensor theory. The WDW equation was derived, after which a semiclassical expansion was performed. At the lowest orders of this expansion the classical theory was recovered. At the higher orders of the expansion it was found that the semiclassical wave functional satisfies a Schrödinger equation, which includes the first quantum gravitational correction terms. Throughout this chapter the configuration space was treated as a differentiable manifold and in **appendix A** all the associated geometrical tensors, including the scalar Ricci curvature of configuration space, were derived. In particular, it was found that, in contrast to pure gravity, the signature of the configuration space metric for scalar-tensor theories with a non-minimally coupled scalar field depends on the signature of spacetime. This might have interesting consequences regarding the hyperbolicity properties of the WDW operator [131]. As required for consistency, at each step of the calculation the results for a minimally coupled scalar field with a canonically normalised kinetic term obtained were recovered, as they were found in [95]. In contrast, for arbitrary field dependent functions $U(\phi)$ and $G(\phi)$, the canonical quantisation and the subsequent semiclassical expansion lead to essential differences compared to the minimally coupled case—both technical and conceptual.

In particular, the non-minimal coupling $U(\phi)$ leads to a mixing between the gravitational and scalar field momenta. This intertwining of gravitational and scalar field degrees of freedom makes it difficult to separate heavy from light degrees of freedom in the multicomponent configuration space. While this might not pose a problem in principle, it complicates the semiclassical expansion at the level of the exact WDW equation. The semiclassical expansion is based on the Born-Oppenheimer approximation which, in turn, requires a clear separation of heavy and light degrees of freedom. In scalar-tensor theories, where the non-minimal coupling is parametrised by an arbitrary non-minimal coupling function $U(\phi)$, there is no manifest distinction between the different degrees of freedom. This is in contrast to the minimally coupled case, where the Planck mass $\kappa = \frac{1}{2}M_{\mathrm{P}}^2$ serves as a natural indicator for the heavy degrees of freedom. Practically, the semiclassical expansion of the WDW equation requires a relative weighting between individual terms in the WDW operator by different powers of λ. This implements the distinction between heavy and light degrees of freedom. A concrete weighting procedure in case of a non-minimal coupling is therefore difficult, as the WDW operator is non-diagonal. In order to nevertheless obtain a consistent and feasible weighting scheme, a transformation was performed to the EF, in which the WDW operator is diagonal. In the EF, the distinction between gravitational and scalar field degrees of freedom is transparent and a clear weighting can be performed by the association of the EF metric field with the heavy degrees of freedom and the scalar field with the light degrees of freedom. Once the weighting has been implemented, the weighted WDW operator can then be transformed back to the original JF variables and the semiclassical expansion can be carried out.

The justification of this procedure relies on the covariant Laplace-Beltrami ordering. On the basis of covariant perturbative one-loop calculations [39, 132–135], the quantum equivalence between different parametrisations of scalar-tensor theories [134] and the equivalence between $f(R)$-gravity and its reformulation as a scalar-tensor theory has been investigated for the one-loop divergences on a general background manifold in reference [52]. A similar investigation of the equivalence of the effective action in the context of Einstein spaces can be found in reference [136]. There, it has been found that the classical equivalence is broken by off-shell contributions but is restored once the equations of motions have been used. In the geometrical treatment of the configuration space, the quantum equivalence between the JF and EF in the non-perturbative canonical theory can be realised, at least formally, by the covariant Laplace-Beltrami factor ordering in the WDW operator. It would be interesting to investigate whether this quantum equivalence also holds between $f(R)$-gravity and its scalar-tensor formulation in quantum geometrodynamics.

For the minimally coupled scalar field case, the final result for the corrected Schrödinger equation is independent of the factor ordering in the kinetic part of the WDW operator. In contrast, for the general scalar-tensor theory **(6.1)** the factor ordering is determined by the Laplace-Beltrami operator and ultimately enters the corrected Schrödinger equation. The additional terms, which arise in the Laplace-Beltrami factor ordering, correspond to lower order derivative terms in the WDW operator and involve delta functions evaluated at the same point. In addition, compared to the minimally coupled case, extra commutator terms have to be taken into account in the corrected Schrödinger equation, which also carry factors of $\delta^{(3)}(0)$. These singular delta functions need to be regulated. The adoption of a regularisation scheme in which operators at the same point commute [94] leads to a corrected Schrödinger equation that has the same form as for the minimally coupled scalar field, but with the minimally coupled scalar Hamilton operator **(6.71)** replaced by the non-minimal scalar Hamilton operator **(6.70)**. The kinetic term of the latter involves derivatives with respect to the scalar field as well as derivatives with respect to gravitational metric. Moreover, the structure of the quantum gravitational correction terms in the case of non-minimal coupling shows additional interesting differences compared to the minimally coupled case. The nature of these differences have been investigated for the specific model of non-minimal Higgs inflation [37,38,41,43,47–49]. In particular, the kinetic part of the scalar Hamilton operator **(6.44)** was shown to be strongly suppressed in the regime of a strong non-minimal coupling. A similar effect has been found in the model of Higgs inflation, where, in the presence of a strong non-minimal coupling, the Higgs propagator is suppressed at high energies [37,43,48].

It would be interesting to explore the features of the quantum gravitational correction terms and the influence of the non-minimal coupling in the canonical formulation of the homogeneous and isotropic cosmological FLRW background—including cosmological perturbations. In the semiclassical expansion of such a cosmological minisuperspace model the homogeneous scalar field $\phi(t)$ and the scale factor $a(t)$ can be identified as the slow variables and can both be treated on an equal footing. This is in contrast to the weighting scheme adopted in this chapter. The fast degrees of freedom are provided in a natural way by the inhomogeneous cosmological perturbations. Most importantly, such a cosmological application would allow the estimation of the effect of a non-minimal coupling on quantum gravitational contributions to the power spectrum of the cosmic microwave background radiation similar to approaches for the case of a minimally coupled scalar field [109,110,137–140]. This is the subject of the next chapter.

QUANTUM GRAVITATIONAL CORRECTIONS IN INFLATION

7

This chapter applies the formalism of quantum geometrodynamics to the cosmological Friedmann-Lemaître-Robertson-Walker (FLRW) universe. The inflationary parameters are extracted from the semiclassical expansion of the Wheeler-DeWitt (WDW) equation. The quantum gravitational corrections to these parameters are then determined, and their magnitude and observational signatures are discussed. The highly symmetric FLRW universe results in significant simplifications of the WDW equation. It was seen in **chapter 4** that the total number of dynamical degrees of freedom of the metric is equal to one: the homogeneous scale factor. Therefore, together with the scalar field, the infinitely dimensional superspace is reduced to a 2-dimensional configuration space, called *minisuperspace*.

The quantum corrections in these inflationary minisuperspace models can become important. For example, in the model of Higgs inflation, the radiative corrections and the renormalisation group improvement turned out to be crucial for the consistency with particle physics experiments [37, 38, 41, 43, 48–51]. While, in this case, the quantum corrections are dominated by the heavy standard model particles, it is in general interesting to study the effect of quantum gravitational corrections on inflationary predictions. In fact, the strong curvature regime during the inflationary phase make the early universe a natural testing ground for any theory of quantum gravity.

Although the canonical approach to quantum gravity does not come without difficulties, both at the conceptual and technical level, the WDW equation can be considered as a natural starting point for the analysis of quantum gravitational effects, as its semiclassical expansion reproduces the classical theory and the functional Schrödinger equation for quantised matter fields on a curved background at the lowest orders of the expansion [95,141]. Therefore, higher order terms in the expansion can be clearly attributed to the first quantum gravitational corrections. When applied to the inflationary universe, these corrections leave observational signatures in the primordial power spectrum. This has been investigated for a minimally coupled scalar field with a canonical kinetic term [109, 110, 137–140, 142]. The goal of this chapter is to generalise these analyses to a general scalar-tensor theory of a single scalar field with an arbitrary non-minimal coupling to gravity, a non-standard kinetic term, and an arbitrary scalar potential.

The chapter is structured as follows: in **section 7.1**, the model is introduced from which the equations of motion will be derived. The general model will be symmetry reduced to a homogeneous and isotropic FLRW universe and the inflationary dynamics of the background and the cosmic perturbations will be discussed. In **section 7.2** the symmetry reduced classical Hamiltonian constraint is derived. As in **chapter 6**, the Hamiltonian constraint is then quantised in the Dirac formalism, which leads to the minisuperspace the WDW equation for the background and perturbation variables. In **section 7.3** the WDW equation is expanded around a semiclassical solution, based on a combined Born-Oppenheimer and WKB-type approximation. As before, the dynamical background equations and the notion of a semiclassical time are recovered at the lowest order. A Schrödinger equation for the perturbations is obtained at the next order. The subsequent order yields the first quantum gravitational corrections to the Schrödinger equation. In **section 7.4** the connection is made between the results found from the semiclassical expansion of the WDW equation and the inflationary power spectra. In **section 7.5** the impact of the leading quantum gravitational corrections on the inflationary power spectra and their observational consequences are discussed. Finally, the main results and conclude in **section 7.6**. The results for the subleading slow-roll contributions to the quantum gravitational corrections are provided in **appendix B**.

7.1 SCALAR-TENSOR THEORIES OF INFLATION

Almost all models of inflation driven by a single scalar field ϕ can be covered by the general scalar-tensor theory **(3.2)**:

$$S[g,\phi] = \int \left[UR - \tfrac{1}{2}G g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi - V\right]\sqrt{-g}\,\mathrm{d}^4X + \text{ boundary terms.} \tag{7.1}$$

Recall that $U(\phi)$, $G(\phi)$, and $V(\phi)$ are three arbitrary functions of the scalar inflaton field ϕ. They respectively parametrise the non-minimal coupling to gravity, the non-canonical kinetic term, and the scalar field potential. Spacetime is taken to be 4-dimensional and equipped with metric $g_{\mu\nu}(X)$ of mostly plus signature. The scalar curvature is denoted by $R(g)$. Spacetime coordinates are labelled by X^μ, with $\mu = 0,\ldots,3$.

7.1.1 *Field equations and energy-momentum tensor*

The field equations for the metric and the Klein-Gordon equation of the inflaton are obtained by variation of **(7.1)** with respect to $g_{\mu\nu}$ and ϕ, respectively:

$$R_{\mu\nu} - \tfrac{1}{2}Rg_{\mu\nu} = \tfrac{1}{2}U^{-1}T^\phi_{\mu\nu}, \tag{7.2}$$

$$\Box\phi = G^{-1}(V_1 - \tfrac{1}{2}G_1\nabla_\mu\phi\nabla^\mu\phi - U_1R), \tag{7.3}$$

where the effective energy-momentum tensor $T^\phi_{\mu\nu}$ is defined as

$$T^\phi_{\mu\nu} \equiv G\left(\delta^\alpha_\mu\delta^\beta_\nu - \tfrac{1}{2}g_{\mu\nu}g^{\alpha\beta}\right)\nabla_\alpha\phi\nabla_\beta\phi - g_{\mu\nu}V + 2\nabla_\mu\nabla_\nu U - 2g_{\mu\nu}\Box U. \tag{7.4}$$

Here $\Box \equiv g^{\mu\nu}\nabla_\mu\nabla_\nu$ denotes the covariant d'Alembert operator.

7.1.2 *Spacetime foliation*

It is useful to reformulate the action in terms of the Arnowitt-Deser-Misner (ADM) formalism [5], where the 4-dimensional metric $g_{\mu\nu}$ is expressed in terms of the lapse function $N(t,\mathbf{x})$, the spatial shift vector $N^a(t,\mathbf{x})$, and the spatial metric $\gamma_{ab}(t,\mathbf{x})$:

$$\mathrm{d}s^2 = g_{\mu\nu}\mathrm{d}X^\mu\mathrm{d}X^\nu = -N^2\mathrm{d}t^2 + 2N_a\mathrm{d}t\mathrm{d}x^a + \gamma_{ab}\mathrm{d}x^a\mathrm{d}x^b. \tag{7.5}$$

Here, the spatial coordinates x^a, with $a = 1,\dots,3$, are denoted by small letters. In terms of the ADM variables, the action (7.1) can be represented as

$$S[\gamma,\phi] = \int \mathrm{d}t\mathrm{d}^3x\left[\tfrac{1}{2}\mathcal{M}_{AB}\,\mathrm{D}_t\,\mathcal{Q}^A\,\mathrm{D}_t\,\mathcal{Q}^B - \mathcal{P}\right],\quad \mathcal{Q}^A = \begin{pmatrix}\gamma_{ab}\\ \phi\end{pmatrix}, \tag{7.6}$$

with the dynamical configuration space variables collectively denoted by $\mathcal{Q}^A$ and the reparametrisation invariant covariant time derivative $\mathrm{D}_t \equiv (\partial_0 - \mathcal{L}_{\mathbf{N}})/N$, with the Lie derivative $\mathcal{L}_{\mathbf{N}}$ along the spatial shift vector $\mathbf{N} = N^a\partial_a$. The bilinear form $\mathcal{M}^{AB}$ corresponds to the inverse of the configuration space metric derived in chapter 6:

$$\mathcal{M}_{AB} = -N\gamma^{1/2}\begin{pmatrix}-\frac{1}{2}UG^{abcd} & U_1\gamma^{ab}\\ U_1\gamma^{cd} & -G\end{pmatrix}. \tag{7.7}$$

The DeWitt metric $G^{abcd} \equiv \gamma^{a(c}\gamma^{d)b} - \gamma^{ab}\gamma^{cd}$ is as defined in (6.10). The potential $\mathcal{P}$, which includes the spatial gradient terms of the scalar field, is defined as

Note that the configuration space metric (7.7) was defined with additional inverse factors of the lapse function in (6.11).

$$\mathcal{P} \equiv N\gamma^{1/2}\left[\tfrac{1}{2}s^{-1}\,\mathrm{D}_a\,\phi\,\mathrm{D}^a\,\phi + V - U{}^{(s)}R - 2\Delta U - \tfrac{3}{2}U^{-1}\,\mathrm{D}_a\,U\,\mathrm{D}^a\,U\right].$$

Here, $\Delta \equiv -\gamma^{ab}\,\mathrm{D}_a\,\mathrm{D}_b$ is the positive definite spatial Laplacian, D_a the spatial covariant derivative compatible with γ_{ab} and ${}^{(s)}R$ is the three-dimensional spatial curvature. In addition, the suppression function s (3.3) is defined to be

$$s \equiv \frac{U}{GU + 3U_1^2}, \tag{7.8}$$

where the subscript is a shorthand for a derivative of the function with respect to the argument.

7.1.3 *Cosmological background evolution*

The background spacetime is described by the spatially flat FLRW line element

$$\mathrm{d}s^2 = -N^2\,\mathrm{d}t^2 + a^2\delta_{ab}\,\mathrm{d}x^a\mathrm{d}x^b. \tag{7.9}$$

Comparison with the ADM line element (7.5) one sees that spatial flatness, homogeneity and isotropy imply $\gamma_{ab} = a^2\delta_{ab}$ and $N_a = 0$, where the lapse function $N = N(t)$ and the scale factor $a = a(t)$ are functions of time t only. Similarly, homogeneity implies that the scalar field is a function of time only $\phi = \phi(t)$. Moreover, for the isotropic line element (7.9) the reparametrisation invariant time derivative D_t reduces to $\mathrm{D}_t = N^{-1}\partial_0$.

From this point on it is convenient to introduce the conformal time τ, related to the coordinate time t by $N\mathrm{d}t = a\mathrm{d}\tau$. This choice corresponds to the gauge $N = a$. In terms of τ, the FLRW metric (7.9) acquires the manifestly conformally flat structure $g_{\mu\nu}(\tau) = a^2(\tau)\eta_{\mu\nu}$, and the reparametrisation invariant time derivative reduces to

$\eta_{\mu\nu}$ is understood to be the flat Minkowski metric.

a partial derivative with respect to conformal time $D_t = a^{-1}\partial_\tau$. It is then convenient to introduce the conformal Hubble parameter $\mathcal{H}(\tau)$, which is defined as

$$\mathcal{H} \equiv \frac{a'}{a}.$$

The prime denotes a derivative with respect to conformal time τ. In the FLRW universe $T^{\phi}_{\mu\nu}$ takes on the form of the energy-momentum tensor of a perfect fluid:

$$T^{\phi}_{\mu\nu} = (\rho_\phi + p_\phi)u_\mu u_\nu + p_\phi g_{\mu\nu}.$$

Here, u_μ is the fluid's four-velocity with norm $u_\mu u^\mu = -1$, ρ_ϕ is its energy density, and p_ϕ is its pressure. Comparison with **(7.4)** leads to the identifications

$$\rho_\phi = \tfrac{1}{2}Ga^{-2}(\phi')^2 + V - 6U'a^{-2}\mathcal{H}, \tag{7.10}$$

$$p_\phi = \tfrac{1}{2}Ga^{-2}(\phi')^2 - V + 2U''a^{-2} + 2U'a^{-2}\mathcal{H}. \tag{7.11}$$

The symmetry reduced FLRW action expressed in terms of the compact notation has a form similar to **(7.6)** with $Q^A = (a, \phi)$ and

$$\mathcal{M}_{AB} = -\begin{pmatrix} 12U & 6U_1 a \\ 6U_1 a & -Ga^2 \end{pmatrix}, \qquad \mathcal{D} = a^4 V. \tag{7.12}$$

The explicit expression for the background action is given by

$$S^{\mathrm{bg}}[a,\phi] = \int \mathcal{L}^{\mathrm{bg}}\left(a, a', \phi, \phi'\right) \mathrm{d}\tau \mathrm{d}^3x, \tag{7.13}$$

$$\mathcal{L}^{\mathrm{bg}}(a, a', \phi, \phi') \equiv a^4\left[-6\frac{U}{a^2}\left(\frac{a'}{a}\right)^2 - 6\frac{U_1}{a}\frac{\phi' a'}{a^2} + \frac{G}{2}\left(\frac{\phi'}{a}\right)^2 - V\right]. \tag{7.14}$$

In particular, the derivative coupling between the gravitational and scalar field degrees of freedom induced by the non-minimal coupling becomes manifest. The Friedmann equations and the Klein-Gordon equation are obtained from the variation of **(7.13)** with respect to N, a, and ϕ, or directly from symmetry reducing the equations of motion **(7.2)** and **(7.3)**:

$$\mathcal{H}^2 = \tfrac{1}{6}a^2\, U^{-1}\rho_\phi, \tag{7.15}$$

$$\mathcal{H}' = -\tfrac{1}{12}a^2\, U^{-1}\left(\rho_\phi + 3p_\phi\right), \tag{7.16}$$

$$\phi'' + 2\mathcal{H}\phi' + \tfrac{1}{2}(\log(U/s))\phi' + a^2 s U^2 W_1 = 0. \tag{7.17}$$

The dimensionless ratio W is related to the EF potential and is defined as in **(3.4)**:

$$W \equiv \frac{V}{U^2}.$$

The spatial integral in the action **(7.13)** is formally divergent in a flat spatially homogeneous FLRW universe. This corresponds to an infinite spatial volume V_0. In order to regularise the spatial integral, a large but finite reference length scale ℓ_0 such that $\ell_0^3 = V_0$. The reference volume V_0 can then be removed from the formalism by through a redefinition of the time variable and the scale factor, such that the action **(7.13)** is independent of V_0 [110,140]:

$$\tau \to \ell_0^{-1}\tau, \qquad a \to \ell_0 a. \tag{7.18}$$

While, in this way, any dependence on the reference scale ℓ_0 has been eliminated form the formalism, the restriction of the spatial volume to a compact subregion nevertheless has observational consequences, which are discussed in **section 7.5**.

7.1.4 *Inflationary background dynamics in the slow-roll approximation*

During the inflationary epoch the universe underwent a quasi-De Sitter stage in which the energy density of the universe is approximately constant and effectively dominated by the potential of the slowly rolling inflaton field. The slow-roll conditions for minimally coupled theories can be generalised in scalar-tensor theories for any function f of the inflaton field ϕ [77]. In terms of conformal time the slow-roll conditions are taken to be

$$f''(\phi) \ll \mathcal{H}f'(\phi) \ll \mathcal{H}^2 f(\phi). \tag{7.19}$$

Notice that the slow-roll condition is slightly different from how it was defined in **chapter 4**. The difference is second order in the slow-roll approximation, and does not influence the main results.

In particular, for the scalar-tensor theory **(7.1)**, this encompasses the generalised potentials $f = \{U, G, V\}$. With use of **(7.10)** and **(7.11)**, within the slow-roll regime **(7.19)** the background equations **(7.15)**-**(7.17)** lead to

$$\frac{\mathcal{H}^2}{Ua^2} \approx \frac{W}{6}, \tag{7.20}$$

$$3\frac{\mathcal{H}\phi'}{U^2a^2} \approx -sW_1. \tag{7.21}$$

The slow-roll conditions **(7.19)** motivate the definitions of the following four slow-roll parameters [143], which quantify small deviations from De Sitter space:

$$\varepsilon_{1,\mathcal{H}} \equiv 1 - \frac{\mathcal{H}'}{\mathcal{H}^2}, \qquad \varepsilon_{2,\mathcal{H}} \equiv 1 - \frac{\phi''}{\mathcal{H}\phi'},$$
$$\varepsilon_{3,\mathcal{H}} \equiv \frac{1}{2}\frac{U'}{\mathcal{H}U}, \qquad \varepsilon_{4,\mathcal{H}} \equiv \frac{1}{2}\frac{s'}{\mathcal{H}s}. \tag{7.22}$$

The slow-roll parameters $\varepsilon_{1,\mathcal{H}}$ and $\varepsilon_{2,\mathcal{H}}$ are the same as for a minimally coupled canonical scalar field, while the slow-roll parameters $\varepsilon_{3,\mathcal{H}}$, $\varepsilon_{4,\mathcal{H}}$ contain information about the non-minimal coupling U and the generalised kinetic term G via the function s defined in **(7.8)**. It is assumed that the slow-roll approximation holds. From **section 4.5.4** it then follows that the slow-roll parameters can be treated as constant. In addition to **(7.22)**, the *potential slow-roll parameters* $\varepsilon_{i,W}$ can be defined as in **(4.21)**:

$$\varepsilon_{1,W} \equiv \frac{(UW)_1}{UW}\frac{sUW_1}{W}, \qquad \varepsilon_{2,W} \equiv 2\left(\frac{sUW_1}{W}\right)_1 + \frac{sUW_1}{W}\frac{(UW)_1}{UW},$$
$$\varepsilon_{3,W} \equiv -\frac{sU_1W_1}{W}, \qquad \varepsilon_{4,W} \equiv -\frac{sUW_1}{W}\frac{s_1}{s}. \tag{7.23}$$

The potential slow-roll parameters allow inflation to be quantified directly in terms of the generalised potentials U, G and V and their derivatives. Within the slow-roll approximation $\varepsilon_{i,W} \approx \varepsilon_{i,\mathcal{H}}$. In what follows the slow-roll parameters will simply be denoted as ε_i for both sets of slow-roll parameters **(7.22)** and **(7.23)**. During the slow-roll regime a sufficiently long quasi-De Sitter phase of inflation is realised for $|\varepsilon_i| \ll 1$, as discussed in **section 4.5.4**.

7.1.5 *Cosmological perturbations*

Small deviations from the perfect FLRW universe justify a decomposition of the field $g_{\mu\nu}$ and ϕ into a background $\bar{g}_{\mu\nu}$, $\bar{\phi}$ and perturbation $\delta g_{\mu\nu}$ and $\delta\phi$,

$$g_{\mu\nu}(\mathbf{x}) \equiv \bar{g}_{\mu\nu}(\tau) + \delta g_{\mu\nu}(\tau,\mathbf{x}), \qquad \phi(\tau,\mathbf{x}) \equiv \bar{\phi}(\tau) + \delta\phi(\tau,\mathbf{x}). \tag{7.24}$$

It was argued in **section 4.6.2** that the perturbations are in general not invariant under infinitesimal coordinate transformations. In order to unambiguously quantify the perturbations one resorts to combinations that of $\delta g_{\mu\nu}$ and $\delta\phi$ that are invariant under coordinate transformations. The fluctuations that are relevant to inflation are the gauge invariant transverse traceless part h^{TT}_{ab} of the metric perturbation, and the comoving curvature perturbation ζ [144]:

$$\zeta \equiv h - \frac{\mathcal{H}}{\phi'}\delta\phi,$$

where h is defined in **(4.27)**. The transverse traceless metric perturbation can be associated with primordial gravitational waves:

$$h^{\mathrm{TT}}_{ab} \equiv \sum_{I=+,\times} e^I_{ab} h^{\mathrm{TT}}_I,$$

where e^I_{ab} denotes the polarisation tensor. The vector perturbations decay during inflation.

The action for the perturbations takes on its canonical form with the introduction of the scalar and tensor MS variables:

$$v \equiv a\, z_{\mathrm{S}}\, \delta\phi_g, \qquad u_I \equiv a\, z_{\mathrm{T}}\, \mathrm{h}^{\mathrm{TT}}_{\mathrm{I}}. \tag{7.25}$$

The corresponding factors z_{S} and z_{T} defined as [143]:

$$z_{\mathrm{S}}^2 \equiv s^{-1}\left(1 + \tfrac{1}{2}\mathcal{H}^{-1}U^{-1}U'\right)^{-2}\left(\frac{\phi'}{\mathcal{H}}\right)^2, \qquad z_{\mathrm{T}}^2 \equiv \tfrac{1}{2}U. \tag{7.26}$$

The action quadratic in the perturbations v and u_I reads [143]:

$$S^{\text{pert}}[v,u] = \int \left[\mathcal{L}^S(v,v') + \mathcal{L}^T(u,u')\right] \mathrm{d}\tau \mathrm{d}^3x, \tag{7.27.a}$$

$$\mathcal{L}^{\text{S}}(v,v') \equiv \frac{1}{2}\left[(v')^2 + \delta^{ij}\partial_i v \partial_j v + \frac{(az_{\text{S}})''}{(az_{\text{S}})} v^2\right], \tag{7.27.b}$$

$$\mathcal{L}^{\text{T}}(u,u') \equiv \frac{1}{2}\sum_{I=+,\times}\left[(u_I')^2 + \delta^{ij}\partial_i u_I \partial_j u_I + \frac{(az_{\text{T}})''}{(az_{\text{T}})}(u_I)^2\right]. \tag{7.27.c}$$

In the derivation of **(7.27.b)** and **(7.27.c)**, total derivative terms are neglected and it is assumed that the background fields satisfy their equations of motion **(7.15)-(7.17)**. Since the linear perturbations are assumed to be linear, the expansion stops at second order and the total combined action of background plus perturbations reads

$$S^{\text{tot}}[a,\phi,v,u] \equiv \int \mathcal{L}^{\text{tot}}(a,a',\phi,\phi',v,v',u,u')\, \mathrm{d}\tau \mathrm{d}^3x \tag{7.28}$$

$$= \int \left[\mathcal{L}^{\text{bg}}(a,a',\phi,\phi') + \mathcal{L}^{\text{S}}(v,v') + \mathcal{L}^{\text{T}}(u,u')\right] \mathrm{d}\tau \mathrm{d}^3x. \tag{7.29}$$

The inhomogeneous perturbations can be expanded in terms of their Fourier components:

$$v(\tau,\mathbf{x}) = V_0^{-1}\sum_k \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} v_k(\tau), \qquad u(\tau,\mathbf{x}) = V_0^{-1}\sum_k \mathrm{e}^{\mathrm{i}\mathbf{k}\cdot\mathbf{r}} u_k(\tau). \tag{7.30}$$

The Fourier components satisfy $v_{\mathbf{k}}^* = v_{-\mathbf{k}}$ and $u_{\mathbf{k},I}^* = u_{-\mathbf{k},I}$, since the position space perturbations are real. The restriction of the spatial volume to a compact subregion makes it necessary to perform the discrete Fourier transform **(7.30)** with the volume factor $V_0 = \ell_0^3$, introduced to regularise the spatial integral in **(7.13)**. Moreover, due to the isotropy of the FLRW background, the mode components can only depend on the magnitude $k \equiv |\mathbf{k}|$ of the wave vector $\mathbf{k}$, rather than its direction. The Fourier transformed action **(7.28)** then acquires the form of a sum of harmonic oscillators

$$S^{\text{pert}}[\{v_k\},\{u_k\}] = \frac{1}{V_0}\sum_k \int \left[\mathcal{L}_k^S(v_k,v_k') + \mathcal{L}_k^T(u_k,u_k')\right] \mathrm{d}\tau, \tag{7.31}$$

$$\mathcal{L}_k^{\text{S}}(v_k,v_k') = \tfrac{1}{2}\left[v_k'(v^*)_k' - \omega_{\text{S}}^2 v_k v_k^*\right], \tag{7.32}$$

$$\mathcal{L}_k^{\text{T}}(u_k,u_k') = \tfrac{1}{2}\sum_{I=+,\times}\left[u_{k,I}'(u^*)_{k,I}' - \omega_{\text{T}}^2 u_{k,I} u_{k,I}^*\right], \tag{7.33}$$

with time-dependent frequencies

$$\omega_{\text{S}}^2(\tau;k) \equiv k^2 - \frac{(az_{\text{S}})''}{az_{\text{S}}}, \qquad \omega_{\text{T}}^2(\tau;k) \equiv k^2 - \frac{(az_{\text{T}})''}{az_{\text{T}}}. \tag{7.34}$$

In a similar fashion as in **(7.18)**, it is possible to eliminate any explicit occurrence of the reference volume in the Fourier transformed action **(7.31)** by the rescaling of the wave number and Fourier components [110,140]:

$$k \to \ell_0 k, \qquad v_k \to \ell_0^{-2} v_k, \qquad u_{k,I} \to \ell_0^{-2} u_{k,I}. \tag{7.35}$$

The Fourier transformed version of the total action **(7.29)** is the starting point for the Hamiltonian formulation carried out in the next section.

7.2 QUANTUM GEOMETRODYNAMICS

7.2.1 *Hamiltonian formalism*

The canonical quantisation of gravity is based on its Hamiltonian formulation. We perform a Legendre transformation of $\mathcal{L}_{\rm tot}$ with the generalised momenta

$$\pi_a \equiv \frac{\partial \mathcal{L}_{\rm tot}}{\partial(a')}, \qquad \pi_\phi \equiv \frac{\partial \mathcal{L}_{\rm tot}}{\partial(\phi')}, \qquad \pi_{v,k} \equiv \frac{\partial \mathcal{L}_{\rm tot}}{\partial(v_k^*)'}, \qquad \pi_{u,k}^I \equiv \frac{\partial \mathcal{L}_{\rm tot}}{\partial(u_{k,I}^*)'}, \tag{7.36}$$

which leads to the Hamiltonian constraint

$$\mathcal{H}^{\rm tot} \equiv \mathcal{H}^{\rm bg} + \mathcal{H}^{\rm pert} = 0. \tag{7.37}$$

The individual Hamiltonians of the background and perturbation variables read

$$\mathcal{H}^{\rm bg}(a,\phi) \equiv -\frac{s}{24Ua^2}\left(Ga^2\pi_a^2 + 12U_1 a\pi_a\pi_\phi - 12U\pi_\phi^2\right) + a^4 V,$$

$$\mathcal{H}^{\rm pert}(v_k, u_k^I, a, \phi) \equiv \sum_k \mathcal{H}_k^{\rm pert} = \sum_k \left(\mathcal{H}_k^{\rm S} + \mathcal{H}_k^{\rm T}\right), \tag{7.38}$$

$$\mathcal{H}_k^{\rm S}(v_k, a, \phi) \equiv \tfrac{1}{2}|\pi_{v,k}|^2 + \tfrac{1}{2}\omega_{\rm S}^2 |v_k|^2, \tag{7.39}$$

$$\mathcal{H}_k^{\rm T}(u_k^I, a, \phi) \equiv \tfrac{1}{2}\sum_{I=+,\times}\left(|\pi_{u,k}^I|^2 + \omega_{\rm T}^2 |u_{I,k}|^2\right). \tag{7.40}$$

7.2.2 *Quantum Geometrodynamics and Wheeler-DeWitt equation*

In the canonical quantisation procedure, the configuration space variables a, ϕ, v_k, $u_{k,I}$, and momenta π_a, π_ϕ, $\pi_{v,k}$, $\pi_{u,k}^I$, are promoted to operators that act on states Ψ and obey the canonical commutation relations:

$$\begin{aligned} \left[\hat{a}, \hat{\pi}_a\right] &= \mathrm{i}, & \left[\hat{\phi}, \hat{\pi}_\phi\right] &= \mathrm{i}, \\ \left[\hat{v}_k, \hat{\pi}_{v,k'}\right] &= \mathrm{i}\,\delta_{k,k'}, & \left[\hat{u}_{k,I}, \hat{\pi}_{u,k'}^J\right] &= \mathrm{i}\,\delta_{k,k'}\delta_I^J, \end{aligned}$$

with all other commutators equal to zero. Formally, the configuration space variables associated with the perturbations should be doubled for a consistent quantisation. This is done by the decomposition of the complex Fourier modes v_k and $u_{k,I}$ into real and imaginary parts [145]. The resulting Hamiltonians are form equivalent to **(7.39)** and **(7.40)**, and therefore there is no loss in generality if one considers the variables to be real. In the Schrödinger representation, the position space operators act multiplicatively and the momentum space operators act as differential operators with the explicit form

$$\pi_a = -\mathrm{i}\frac{\partial}{\partial a}, \qquad \pi_\phi = -\mathrm{i}\frac{\partial}{\partial \phi}, \qquad \pi_{v,k} = -\mathrm{i}\frac{\partial}{\partial v_k}, \qquad \pi^I_{u,k} = -\mathrm{i}\frac{\partial}{\partial u_{k,I}}. \tag{7.41}$$

The WDW equation is obtained by the promotion of **(7.37)** to an operator equation, which acts on the wave function $\Psi(a, \phi, v_k, u_{k,I})$. In accordance with the prescription for the quantisation of constrained systems described in **chapter 5**, the implementation of the classical constraint equation **(7.37)** at the quantum level corresponds to the selection of only those states Ψ which are annihilated by $\hat{\mathcal{H}}_{\text{tot}}$:

$$\hat{\mathcal{H}}^{\text{tot}}\Psi = 0. \tag{7.42}$$

The WDW equation **(7.42)** is defined only up to operator ordering. The results for the semiclassical expansion performed in the subsequent sections are—as reasoned in **chapter 5**—independent of the factor ordering [95].

7.3 SEMICLASSICAL EXPANSION OF THE WHEELER-DEWITT EQUATION

For almost all cases, the full WDW equation cannot be solved exactly. However, a consistent semiclassical expansion in analogy to **section 5.3.1** yields approximate results that allow one to determine the first quantum gravitational corrections. This semiclassical expansion is based on the combined use of a Born-Oppenheimer and WKB-type approximation scheme. The former relies on a clear distinction between the heavy and light degrees of freedom. In the original Born-Oppenheimer approach to molecular physics, this distinction is based on the presence of a mass hierarchy between different degrees of freedom. For a scalar field ϕ minimally coupled to Einstein gravity, such a mass hierarchy could be related to the ratio $\lambda \equiv m_\phi^2/M_{\text{P}}^2 \ll 1$, with the effective scalar field mass m_ϕ. In this context, the gravitational degrees of freedom are the heavy ones, while the scalar field degrees of freedom are the light ones [93, 95, 122–126, 139, 140]. Such a scenario would correspond to a slowly varying background geometry on which the quantum matter (in this case the scalar field) degrees of freedom propagate. For a scalar field non-minimally coupled to gravity, the identification of light and heavy degrees of freedom becomes more subtle, as became clear in **chapter 6**. In the case of the Hamiltonian **(7.37)**, the heavy degrees of freedom are identified with the homogeneous background variables a and ϕ, while the light degrees of freedom are associated with the infinitely many degrees of freedom of the Fourier components of the inhomogeneous perturbations v and u^I. In the cosmological framework, this distinction follows naturally from the observed temperature anisotropies $\Delta T/T \approx 10^{-5}$ in the CMB.

7.3.1 *Implementation of the semiclassical expansion*

In the following it is more convenient to temporarily return to the condensed notation of **chapter 5** and collectively denote the heavy degrees of freedom by $Q^A \equiv (a, \phi)$

and the light degrees of freedom by $q_n \equiv (v_k, u_k^+, u_k^\times)$. The index n labels both the Fourier modes k as well as the different types of perturbations. At a technical level, the distinction between heavy and light degrees of freedom can be implemented by the introduction of a formal weighting parameter λ in the Hamiltonian for the heavy degrees of freedom, which can be set to one after the expansion

$$\hat{\mathcal{H}}^{\rm bg}(\hat{Q}, \hat{\pi}_Q) \to \hat{\mathcal{H}}^{\rm bg}_\lambda(\hat{Q}, \hat{\pi}_Q) = -\frac{\lambda}{2}\mathcal{M}^{AB}(\hat{Q})\frac{\partial^2}{\partial Q^A \partial Q^B} + \lambda^{-1}\mathcal{P}(\hat{Q}). \tag{7.43}$$

Here, in correspondence with the notation in **(7.6)**, $\hat{Q}^A$ collectively denotes the operators $\hat{a}$ and $\hat{\phi}$, and $\mathcal{M}^{AB}(\hat{Q})$ and $\mathcal{P}(\hat{Q})$ denote the operator versions of **(7.12)**. The combination of **(7.37)** with **(7.41)** and the weighting of the background Hamiltonian **(7.43)**, the WDW equation has the form

$$\left(\hat{\mathcal{H}}^{\rm bg}_\lambda + \sum_n \hat{\mathcal{H}}^{\rm pert}_n\right)\Psi = 0. \tag{7.44}$$

The Hamiltonian of the perturbation q_n has the explicit form

$$\hat{\mathcal{H}}^{\rm pert}_n = \tfrac{1}{2}\left(-\frac{\partial^2}{\partial q_n^2} + \omega_n^2 \hat{q}_n^2\right). \tag{7.45}$$

The background variables a, ϕ in the frequencies $\omega_{\rm S}^2(\tau; k)$ and $\omega_{\rm T}^2(\tau; k)$ are explicitly treated as classical. Within the semiclassical expansion, this means that the variables Q^A enter the frequencies only parametrically via τ. This procedure might be justified ***a posteriori***, by showing that a full quantum treatment of these variables would only affect terms at higher order in the semiclassical expansion [109].

In what follows the hats on operators are suppressed and the following abbreviations are introduced:

$$\partial_A \equiv \frac{\partial}{\partial Q^A}, \qquad \partial_A \partial^A \equiv \mathcal{M}^{AB}\partial_A\partial_B.$$

The additive structure of the WDW equation **(7.44)** suggests the product ansatz

$$\begin{aligned}\Psi(Q, \{q_n\}) &\equiv \Psi_{\rm bg}(Q)\Psi_{\rm pert}(Q; \{q_{\rm n}\}),\\ \Psi_{\rm pert}(Q; \{q_n\}) &\equiv \prod_n \Psi_n(Q; q_n).\end{aligned} \tag{7.46}$$

The ansatz **(7.46)** can be inserted into the WDW equation **(7.44)**. After the result has been divided by Ψ, the terms that only depend on the background variables Q can be separated from those that depend additionally on the perturbations q. One then obtains a family of separate equations [108,126]:

$$-\tfrac{1}{2}\lambda\partial_A\partial^A\Psi_{\rm bg}+\lambda^{-1}\mathcal{P}(\hat{Q})\Psi_{\rm bg}=f(Q)\Psi_{\rm bg}, \tag{7.47}$$

$$\sum_n\left[-\frac{1}{2}\frac{\partial_A\partial^A\Psi_n}{\Psi_n}-\frac{\partial_A\Psi_{\rm bg}\partial^A\Psi_{\rm n}}{\Psi_{\rm bg}\Psi_{\rm n}}\right.$$
$$\left.+\lambda^{-1}\frac{\mathcal{H}_n\Psi_n}{\Psi_n}-\frac{1}{2}\sum_{m\neq n}\frac{\partial_A\Psi_n\partial^A\Psi_m}{\Psi_n\Psi_m}\right]=-f(Q). \tag{7.48}$$

Here, $f(Q)$ is an arbitrary function that corresponds to the backreaction of the perturbations on the background. In addition, it is assumed that the random phase approximation holds:

$$\sum_{n\neq m}\frac{\partial_A\Psi_n\partial^A\Psi_m}{\Psi_n\Psi_m}\approx 0.$$

Under these assumptions, $f(Q)\equiv\sum_n f_n(Q)$ and **(7.48)** decomposes into a family of separate equations for each n. In the following the backreaction is neglected by the choice $f_n(Q)=0$, such that the background wave function $\Psi_{\rm bg}(Q)$ satisfies the background part of the WDW equation and **(7.44)** reduces to the following family of equations:

$$-\tfrac{1}{2}\lambda\partial_A\partial^A\Psi_{\rm bg}+\lambda^{-1}\mathcal{P}(\hat{Q})\Psi_{\rm bg}=0, \tag{7.49}$$

$$-\tfrac{1}{2}\lambda\partial_A\partial^A\Psi_n-\lambda\partial_A\log\Psi_{\rm bg}\partial^A\Psi_{\rm n}+\mathcal{H}_{\rm n}\Psi_{\rm n}=0. \tag{7.50}$$

In order to get approximate solutions to these equations one can perform a WKB-type approximation and assume that the Ψ_n depend only adiabatically on the background variables $\Psi_n(Q,q_n)=\Psi_n(Q;q_n)$. Concretely, a change of the background variables Q causes the Ψ_n to change much slower than $\Psi_{\rm bg}$:

$$\left|\frac{\partial_A\Psi_{\rm bg}}{\Psi_{\rm bg}}\right|\gg\left|\frac{\partial_A\Psi_n}{\Psi_n}\right|. \tag{7.51}$$

This motivates the following ansatz for $\Psi_{\rm bg}$ and Ψ_n, where the expansion in Ψ_n starts at order $\mathcal{O}(\lambda^0)$ rather than $\mathcal{O}(\lambda^{-1})$:

$$\Psi_{\rm bg}(Q)=\exp\left\{\mathrm{i}\left[\lambda^{-1}S^{(0)}(Q)+S^{(1)}(Q)+\lambda S^{(2)}(Q)+\dots\right]\right\}, \tag{7.52}$$

$$\Psi_n(Q;q_n)=\exp\left\{\mathrm{i}\left[I_n^{(1)}(Q;q_n)+\lambda I_n^{(2)}(Q;q_n)+\dots\right]\right\}. \tag{7.53}$$

Substitution of **(7.52)** and **(7.53)** into **(7.49)** and **(7.50)** leads to two polynomials in λ. If the coefficients in each of those polynomials are separately set to zero, two families of equations are obtained: one for the background functions $S^{(j)}$ and one for the perturbation functions $I_n^{(j)}$. The resulting equations are then solved order by order. It must be noted that the equations for the background must be solved first, as their solutions enter the equations for the perturbations. In order to extract the first quantum gravitational corrections, it is sufficient to consider the expansions **(7.52)** and **(7.53)** up to $\mathcal{O}(\lambda)$.

7.3.2 *Hierarchy of background equations*

The wave function Ψ_{bg} can be reconstructed up to $\mathcal{O}(\lambda)$ by an iterative procedure, where the functions $S^{(0)}$, $S^{(1)}$ and $S^{(2)}$ are solved order by order in the semiclassical expansion.

$\mathcal{O}(\lambda^{-1})$ At this order one obtains a Hamilton-Jacobi equation for $S^{(0)}$:

$$\frac{1}{2}\frac{\partial S^{(0)}}{\partial t_{\text{s}}} + \mathcal{P} = 0. \tag{7.54}$$

The semiclassical time t_{s} arises from the expansion of the timeless WDW equation (7.42) as the projection along the gradient of the background geometry $S^{(0)}(Q)$:

$$\frac{\partial}{\partial t_{\text{s}}} \equiv \partial_A S^{(0)} \partial^A. \tag{7.55}$$

The consistency of the semiclassical expansion requires that the classical theory is recovered at the lowest order. Indeed, by the identification of the semiclassical time (7.55) with the conformal time τ and the gradient of $S^{(0)}$ with the background momenta

$$\pi_A = \frac{\partial S^{(0)}}{\partial Q^A}, \tag{7.56}$$

the Hamilton-Jacobi equation (7.54) yields the equations of motion (7.15)-(7.17) after one uses (7.56). The Hamilton-Jacobi equation therefore implies the classical equations of motion for the background variables $\mathcal{Q} = (a, \phi)$.

$\mathcal{O}(\lambda^{0})$ Equipped with the semiclassical notion of time (7.55), at the next order of the semiclassical expansion one obtains

$$\frac{\partial S^{(1)}}{\partial \tau} = \tfrac{1}{2}\mathrm{i}\partial_A \partial^A S^{(0)}. \tag{7.57}$$

With the definition of the semiclassical time (7.55), the solution can be seen to be

$$S^{(1)} = -\tfrac{1}{2}\mathrm{i}\log\Delta. \tag{7.58}$$

Here, Δ is a function that satisfies the transport equation

$$\partial_A\left(\Delta \partial^A S^{(0)}\right) = 0.$$

This is consistent with the first order corrections to the WKB prefactor, where Δ is associated with the Van Vleck determinant.

$\mathcal{O}(\lambda^{1})$ The next order in the expansion yields

$$\frac{\partial S^{(2)}}{\partial \tau} = \tfrac{1}{2}\left(\mathrm{i}\,\partial_A \partial^A S^{(1)} - \partial_A S^{(1)} \partial^A S^{(1)}\right). \tag{7.59}$$

After substitution of **(7.58)** into **(7.59)** it is found that $S^{(2)}$ satisfies the differential equation

$$\frac{\partial S^{(2)}}{\partial \tau} = \tfrac{1}{4}\left[\partial_A \partial^A \log \Delta - \tfrac{1}{2}\partial_A \log \Delta \partial^A \log \Delta\right].$$

This shows that $S^{(2)}$ corresponds to the second order correction to the WKB prefactor.

7.3.3 *Hierarchy of perturbation equations*

With the approximate solution of **(7.49)** it is now possible to expand **(7.50)**. The equations **(7.54)**-**(7.59)** for the background can be used to reconstruct the Ψ_n up to first order in the expansion parameter λ.

$\mathcal{O}(\lambda^0)$ The equation obtained at this order in the expansion can, with the help of **(7.55)**, be written as

$$-\frac{\partial I_n^{(1)}}{\partial \tau} = -\frac{\mathrm{i}}{2}\frac{\partial^2 I_n^{(1)}}{\partial q_n^2} + \frac{1}{2}\frac{\partial I_n^{(1)}}{\partial q_n}\frac{\partial I_n^{(1)}}{\partial q_n} + \tfrac{1}{2}\omega_n^2 q_n^2. \tag{7.60}$$

It can be verified by substitution that equation is equivalent to the Schrödinger equation for the states $\Psi_n^{(1)} \equiv \exp(\mathrm{i}I_n^{(1)})$:

$$\mathrm{i}\frac{\partial \Psi_n^{(1)}}{\partial \tau} = \hat{\mathcal{H}}_n^{\text{pert}}\Psi_n^{(1)}. \tag{7.61}$$

$\mathcal{O}(\lambda^1)$ The first quantum gravitational corrections arise from the semiclassical expansion at order $\mathcal{O}(\lambda^1)$. With the use of **(7.55)**, the equation obtained at this order in the expansion can be written as

$$\begin{aligned}-\frac{\partial I_n^{(2)}}{\partial \tau} &= \partial_A S^{(1)}\partial^A I_n^{(1)} + \tfrac{1}{2}\partial_A I_n^{(1)}\partial^A I_n^{(1)} - \tfrac{1}{2}\mathrm{i}\partial_A\partial^A I_n^{(1)} \\ &\quad - \frac{\mathrm{i}}{2}\frac{\partial^2 I_n^{(2)}}{\partial q_n^2} + \frac{\partial I_n^{(1)}}{\partial q_n}\frac{\partial I_n^{(2)}}{\partial q_n}.\end{aligned} \tag{7.62}$$

This equation can be written in the form of a corrected Schrödinger equation for the state $\Psi_n^{(2)} \equiv \Psi_n^{(1)}\exp(\mathrm{i}\lambda I_n^{(2)})$, where $\Psi_n^{(1)}$ satisfies **(7.61)**. The equation for $\Psi_n^{(2)}$ is then

Terms of order λ^2 are neglected in the conversion from $I^{(2)}$ to $\Psi^{(2)}$ in the transition from **(7.62)** to **(7.63)**,

$$\mathrm{i}\frac{\partial \Psi_n^{(2)}}{\partial \tau} = \hat{\mathcal{H}}_n^{\text{pert}}\Psi_n^{(2)} - \lambda\Psi_n^{(2)}\left(\mathrm{i}\frac{\partial_A S^{(1)}\partial^A\Psi_n^{(1)}}{\Psi_n^{(1)}} + \frac{1}{2}\frac{\partial_A\partial^A\Psi_n^{(1)}}{\Psi_n^{(1)}}\right). \tag{7.63}$$

The terms proportional to λ are identified as the first quantum gravitational corrections. In accordance with the strategy of reference [95] these terms can be projected along the direction normal to the hypersurfaces of constant $S^{(0)}$. With **(7.54)**, **(7.57)** and **(7.61)**, the quantum gravitational correction terms can be represented in the form

$$\mathcal{V}_n^{\mathrm{QG}} \equiv -\tfrac{1}{4}\lambda\ \mathrm{Re}\left[\frac{1}{\Psi_n^{(1)}\,\mathcal{D}}\left(\hat{\mathcal{H}}_n^{\mathrm{pert}}\right)^2\Psi_n^{(1)} + \mathrm{i}\frac{1}{\Psi_n^{(1)}}\left(\frac{\partial}{\partial\tau}\frac{\hat{\mathcal{H}}_n^{\mathrm{pert}}}{\mathcal{D}}\right)\Psi_n^{(1)}\right]. \tag{7.64}$$

In accordance with the treatment of reference [109, 110, 137] one can consider only take the real part of the corrections **(7.64)** in order to preserve unitarity defined with respect to the Schrödinger inner product on the Hilbert space of the perturbations. The question of unitarity in the context of the canonical approach to quantum gravity and the semiclassical expansion is controversially discussed and an interesting topic on its own. It has been studied extensively in the literature [95, 97, 113, 126, 127, 130, 146], and has also been discussed in **chapter 5** and **chapter 6**. The term $\mathcal{V}_n^{\mathrm{QG}}$ in **(7.64)** might be viewed as a contribution to the effective potential

$$\mathrm{i}\frac{\partial\Psi_n^{(2)}}{\partial\tau} = -\frac{1}{2}\frac{\partial^2\Psi_n^{(2)}}{\partial q_n^2} + \mathcal{V}_n^{\mathrm{eff}}\Psi_n^{(2)}, \qquad \mathcal{V}_n^{\mathrm{eff}} \equiv \tfrac{1}{2}\omega_n^2 q_n^2 + \mathcal{V}_n^{\mathrm{QG}}. \tag{7.65}$$

The iterative scheme of the semiclassical expansion implies that equations obtained at lower orders in λ are used to derive equations that arise at higher orders in the expansion. In order to solve the corrected Schrödinger equation **(7.65)** for $\Psi_n^{(2)}$, one is required to have knowledge about the solution $\Psi_n^{(1)}$ of the uncorrected Schrödinger equation **(7.61)**.

7.4 COSMOLOGICAL POWER SPECTRA IN THE SCHRÖDINGER PICTURE

In order to extract physical information from the semiclassical expansion it is necessary to relate observations to the WKB states $\Psi_n^{(j)}$, where the (j) indicates the order of the semiclassical expansion. The inflationary perturbations are assumed to be Gaussian, which means that they are determined by the two-point correlation function. The main observable in inflationary cosmology is the inflationary power spectrum, which results from a Fourier transform of the two-point correlation function. Since observational data do not show any evidence for non-Gaussian features, higher n-point correlation functions can be neglected. This is consistent with the truncation of **(7.27.a)** to quadratic order in the perturbations, since investigations of, for example, the bispectrum would imply interaction terms cubic in the perturbations. In the Schrödinger picture, this suggests that the $\Psi_n^{(j)}$ obtained from the semiclassical expansion can be assumed to be normalised Gaussian states:

$$\Psi_n^{(j)}(\tau;q_n) = N_n^{(j)}(\tau)\exp\left(-\tfrac{1}{2}\Omega_n^{(j)}(\tau)q_n^2\right), \quad \mathrm{Re}(\Omega_n) > 0. \tag{7.66}$$

They are fully characterised by the complex Gaussian width $\Omega_n(\tau)$ which depends parametrically on the semiclassical time τ. For Gaussian states **(7.66)**, the quantum average in the WKB state $\Psi_{\mathrm{pert}}^{(j)} = \prod_n \Psi_n^{(j)}$ can be evaluated explicitly. The two-point correlation function in the Schrödinger picture is a simple function of the Gaussian width [145]:

$$\left\langle \Psi^{(j)}_{\text{pert}} | v_{\mathbf{k}} v^*_{\mathbf{k}'} | \Psi^{(j)}_{\text{pert}} \right\rangle = \frac{2\pi^2}{k^3} P^{(j)}_v \delta(\mathbf{k} - \mathbf{k}'), \tag{7.67.a}$$

$$P^{(j)}_v(\tau; k) \equiv \frac{k^3}{4\pi^2} \text{Re} \left[\Omega^{(j)}_v(k;\tau) \right]^{-1}. \tag{7.67.b}$$

Similar relations hold for the tensor power spectrum. Thus, the width $\Omega^{(j)}_n$ fully determines the power spectrum $P^{(j)}_v$ up to order $\mathcal{O}(\lambda^j)$ of the semiclassical expansion. Since the canonical field variables for the scalar and tensor perturbations v and u^I are related to the original perturbations $\delta\phi_{\text{gi}}$ and h^I via **(7.25)**, the corresponding power spectra are related to **(7.67)** by

$$P_{\text{S}}(k) \equiv \frac{1}{a^2 z_{\text{S}}^2} P_v(k), \qquad P_{\text{T}}(k) \equiv \frac{2}{a^2 z_{\text{T}}^2} P_u(k). \tag{7.68}$$

The extra factor of 2 in $P_{\text{T}}(k)$ accounts for the two polarisations. The index (j) is suppressed for notational simplicity. The power spectra can be parametrised by the power law forms

$$P_{\text{S}}(k) = A_{\text{S}}(k_*) \left(\frac{k}{k_*} \right)^{n_{\text{S}}(k_*) - 1 + \dots}, \qquad P_{\text{T}}(k) = A_{\text{T}}(k_*) \left(\frac{k}{k_*} \right)^{n_{\text{T}}(k_*) + \dots}. \tag{7.69}$$

The pivot scale k_* is chosen to correspond to a mode within the experimentally accessible window of scales that re-entered the horizon N_{e} e-folds after the end of inflation. In terms of the parametrisation **(7.69)**, the power spectra are characterised by their amplitudes $A_{\text{S/T}}$ which measure the heights, and their spectral indices $n_{\text{S/T}}$, which measure the tilts

$$n_{\text{S}} \equiv 1 + \left. \frac{\mathrm{d} \log P_{\text{S}}}{\mathrm{d} \log k} \right|_{k=k_*}, \qquad n_{\text{T}} \equiv \left. \frac{\mathrm{d} \log P_{\text{T}}}{\mathrm{d} \log k} \right|_{k=k_*}. \tag{7.70}$$

Since the primordial tensor modes have not yet been measured there only exists an upper bound on A_{T} and it is convenient to introduce the tensor-to-scalar ratio

$$r = \frac{P_{\text{T}}(k_*)}{P_{\text{S}}(k_*)}. \tag{7.71}$$

For single field models of inflation, there is a consistency equation that relates r to n_{T}:

$$r = -8 n_{\text{T}}. \tag{7.72}$$

7.4.1 *Power spectra without quantum gravitational corrections*

At order $\mathcal{O}(\lambda^0)$ of the semiclassical expansion the Schrödinger equation **(7.61)** for the states $\Psi^{(1)}_n(\tau, q_n)$ was obtained. According to **(7.66)**, the Gaussian ansatz

$$\Psi^{(1)}_n = N^{(1)}_n \exp\left(-\tfrac{1}{2} \Omega^{(1)}_n q_n^2 \right) \tag{7.73}$$

Equation (7.74) can be transformed into the conventional second order equation via an auxiliary variable. This is explained in detail in appendix C.

can be substituted into (7.61), after which terms of equal order in the q_n can be collected. This leads to the two separate equations

$$\mathrm{i}\frac{\mathrm{d}N^{(1)}}{\mathrm{d}\tau} = \tfrac{1}{2}\Omega^{(1)}N^{(1)}, \qquad \mathrm{i}\frac{\mathrm{d}\Omega^{(1)}}{\mathrm{d}\tau} = \left(\Omega^{(1)}\right)^2 - \omega^2. \tag{7.74}$$

Here and in what follows the subindex n will be suppressed. The frequency $\omega = \omega_{\mathrm{S/T}}$ is given by (7.34) for the scalar and tensor modes, respectively. The equation for $N_k^{(1)}$ just reproduces the usual normalisation condition for the Gaussian. The equation for $\Omega_k^{(1)}$ is a first order non-linear differential equation. To linear order in the slow-roll approximation one finds, in accordance with reference [147], the following expressions for the frequencies of the scalar and tensor modes

$$\omega_{\mathrm{S}}^2(\tau,k) = \omega_{\mathrm{DS}}^2 - 3\frac{\mathcal{E}_{\mathrm{S}}}{\tau^2}, \qquad \omega_{\mathrm{T}}^2(\tau,k) = \omega_{\mathrm{DS}}^2(\tau,k) - 3\frac{\mathcal{E}_{\mathrm{T}}}{\tau^2}, \tag{7.75}$$

where the abbreviations $\mathcal{E}_{\mathrm{S}}$ and $\mathcal{E}_{\mathrm{T}}$ collect the contributions from the slow-roll parameters:

$$\mathcal{E}_{\mathrm{S}} \equiv 2\varepsilon_1 - \varepsilon_2 - \varepsilon_3 + \varepsilon_4, \qquad \mathcal{E}_{\mathrm{T}} \equiv \varepsilon_1 + \varepsilon_3, \tag{7.76}$$

and the universal time-dependent De Sitter frequency

$$\omega_{\mathrm{DS}}^2(k,\tau) \equiv k^2 - \frac{2}{\tau^2}. \tag{7.77}$$

The equation (7.74) for $\Omega^{(1)}$ can be solved analytically. The result can be expressed in terms of Bessel functions, which in turn can be expanded in powers of the slow-roll parameters. However, since so far the analysis is restricted to the first order in the slow-roll approximation, it is simpler to obtain the correction solutions with the

$$\Omega^{(1)} \equiv k(\Omega_{\mathrm{DS}} + \mathcal{E}\Omega_{\mathcal{E}}), \tag{7.78}$$

with $\mathcal{E} = \mathcal{E}_{\mathrm{S/T}}$, for scalar and tensor modes respectively. Insertion of (7.78), (7.75) and (7.76) into (7.74) leads to a linear function in $\mathcal{E}$. The coefficients of this function can be set to vanish separately, which results in two separate equations. These equations can be written in terms of variable $x \equiv -k\tau$ as

$$\frac{\mathrm{d}\Omega_{\mathrm{DS}}}{\mathrm{d}x} = \mathrm{i}\Omega_{\mathrm{DS}}^2 - \mathrm{i}x^{-2}(x^2-2), \tag{7.79}$$

$$\frac{\mathrm{d}\Omega_{\mathcal{E}}}{\mathrm{d}x} = 2\mathrm{i}\Omega_{\mathrm{DS}}\Omega_{\mathcal{E}} + 3\mathrm{i}x^{-2}. \tag{7.80}$$

The system (7.79) and (7.80) can be solved successively. The equation for Ω_{DS} must be solved first, since its solution is needed to find $\Omega_{\mathcal{E}}$. In order to solve these differential equations the physical Bunch-Davies boundary condition is imposed. Specifically, the solution $\Omega^{(1)}$ of (7.74) is required to match the solution $\Omega_\infty^{(1)}$ of the equation obtained form (7.74) in the early time limit $\tau \to -\infty$ (corresponding to $x \to \infty$).

Since the frequencies $\omega_{\mathrm{S/T}}$ become time-independent for early times, the asymptotic limit of (7.74) reads

$$\mathrm{i}\frac{\mathrm{d}\Omega_\infty^{(1)}}{\mathrm{d}\tau} = \left(\Omega_\infty^{(1)}\right)^2 - k^2. \tag{7.81}$$

An obvious solution to (7.81) is the time-independent Gaussian width $\Omega_\infty^{(1)} = k$. Therefore, in the limit $\tau \to -\infty$, the WKB wave function $\Psi^{(1)}$, satisfies a stationary Schrödinger equation $\hat{\mathcal{H}}^{\mathrm{pert}}\Psi^{(1)} = 0$ with the Hamiltonian $\hat{\mathcal{H}}^{\mathrm{pert}}$ of a harmonic oscillator with time-independent frequency $\omega_\infty = k$. In view of the ansatz (7.78), the early-time asymptotic Bunch-Davies boundary condition

$$\lim_{\tau\to-\infty} \Omega^{(1)}(\tau) \equiv \Omega_\infty^{(1)} = k \tag{7.82}$$

implies the asymptotic boundary conditions

$$\lim_{x\to\infty} \Omega_{\mathrm{DS}}(x) = 1, \qquad \lim_{x\to\infty} \Omega_{\mathcal{E}}(x) = 0. \tag{7.83}$$

With the boundary conditions (7.83), the solutions to (7.79) and (7.80) read

$$\Omega_{\mathrm{DS}} = \frac{x^2 - \mathrm{i}x^{-1}}{x^2+1}, \tag{7.84}$$

$$\Omega_{\mathcal{E}} = \mathrm{i}\frac{1 + (2\mathrm{i}+x)x - 2\mathrm{e}^{2\mathrm{i}x}x^3[\pi - \mathrm{i}\,\mathrm{Ei}(-2\mathrm{i}x)]}{x(x-\mathrm{i})^2}. \tag{7.85}$$

Here $\mathrm{Ei}(z)$ is the exponential integral function of complex argument z (defined on the complex plane with a branch cut along the negative z-axis), which is most conveniently defined in terms of the exponential integral function $E_1(z)$, which, in turn, is defined explicitly by its integral representation for $\mathrm{Re}(z) > 0$:

$$\mathrm{Ei}(z) \equiv -\mathrm{E}_1(-z) + \tfrac{1}{2}\left[\log(z) - \log\left(\frac{1}{z}\right)\right] - \log(-z),$$

$$\mathrm{E}_1(z) \equiv \int_1^\infty \mathrm{d}t\,\frac{\mathrm{e}^{-zt}}{t}, \qquad \mathrm{Re}(z) > 0.$$

For small (large) arguments $|z| \ll 1$ ($|z| \gg 1$), the exponential integral $\mathrm{Ei}(z)$ has the (asymptotic) expansions

$$\mathrm{Ei}(z) = \gamma_{\mathrm{E}} + \tfrac{1}{2}\left[\log(z) - \log\left(\frac{1}{z}\right)\right] + \sum_{k=1}^\infty \frac{1}{k!}\frac{z^k}{k}, \qquad |z| \ll 1, \tag{7.86}$$

$$\mathrm{Ei}(z) = \tfrac{1}{2}\left[\log(z) - \log\left(\frac{1}{z}\right)\right] - \log(-z) + z^{-1}\mathrm{e}^{-z}\left[1 + \mathcal{O}\left(z^{-1}\right)\right], \qquad |z| \gg 1.$$

The logarithms in the conversion between Ei and E_1 arise because Ei is multivalued. The combination of (7.84) and (7.85) with the ansatz (7.78) yields the solution for the Gaussian width $\Omega^{(1)}$ to first order in the slow-roll approximation. According to (7.67) and (7.68), the width then fully determines the inflationary scalar and tensor power spectra. From (7.86) it can then be deduced that, at superhorizon scales $k^{-1} \gg \mathcal{H}^{-1}$ (or, equivalently, $x \ll 1$), the leading behaviour of the Gaussian width is

$$\mathrm{Re}(\Omega^{(1)}) \approx kx^2 \left[1 - 2c_\gamma \mathcal{E} + 2\mathcal{E} \log x\right]. \tag{7.87}$$

Here, $c_\gamma \equiv 2 - \gamma_\mathrm{E} - \log 2 \approx 0.7296$ is a numerical constant that involves the Euler-Mascheroni constant $\gamma_\mathrm{E} \approx 0.5772$. From (7.67) and (7.68), one then obtains the scalar and tensor power spectra

$$P_\mathrm{S}^{(1)}(k) \approx \frac{1}{(2\pi z_\mathrm{S})^2} \left(\frac{k}{ax}\right)^2 \left[1 + 2c_\gamma \mathcal{E}_\mathrm{S} - 2\mathcal{E}_\mathrm{S} \log x\right], \tag{7.88}$$

$$P_\mathrm{T}^{(1)}(k) \approx \frac{2}{(2\pi z_\mathrm{T})^2} \left(\frac{k}{ax}\right)^2 \left[1 + 2c_\gamma \mathcal{E}_\mathrm{T} - 2\mathcal{E}_\mathrm{T} \log x\right]. \tag{7.89}$$

With the explicit expressions (7.76) for $\mathcal{E}_{\mathrm{S}/\mathrm{T}}$ and the first order slow-roll relation between the conformal time and the conformal Hubble parameter

$$\tau \approx -\frac{1 + \varepsilon_1}{\mathcal{H}}, \tag{7.90}$$

together with explicit expressions for (7.26) in terms of slow-roll parameters

$$z_\mathrm{S}^2 = 4U \frac{\varepsilon_1 + \varepsilon_3}{(1 + \varepsilon_3)^2} \qquad \text{and} \qquad z_\mathrm{T}^2 = \tfrac{1}{2} U,$$

the inflationary power spectra to first order in the slow-roll approximation are found to be

$$P_\mathrm{S}^{(1)}(k) \approx \frac{W}{96\pi^2(\varepsilon_1 + \varepsilon_3)} \left[1 - \tfrac{1}{3}\left(5\varepsilon_1 - 6c_\gamma \mathcal{E}_\mathrm{S}\right) - 2\mathcal{E}_\mathrm{S} \log\left(k/\mathcal{H}\right)\right], \tag{7.91}$$

$$P_\mathrm{T}^{(1)}(k) \approx \frac{W}{6\pi^2} \left[1 - \tfrac{1}{3}\left(5\varepsilon_1 + 6\varepsilon_3 - 6c_\gamma \mathcal{E}_\mathrm{T}\right) - 2\mathcal{E}_\mathrm{T} \log(k/H)\right]. \tag{7.92}$$

Note that, instead of (7.21), use has been made of the conformal time version of the relation (4.19):

$$\frac{\mathcal{H}^2}{Ua^2} \approx \frac{W}{6} \left(1 + \tfrac{1}{3}\varepsilon_1 - 2\varepsilon_3\right), \tag{7.93}$$

in order to express factors of $\mathcal{H}^2/Ua^2$ in (7.91) and (7.92) in terms of W. Equation (7.93) follows from (7.15) and (7.16) in the slow-roll approximation. Finally, the spectral observables can be calculated by substitution of (7.91) and (7.92) into (7.70) and (7.71):

$$n_{\mathrm{S}}^{(1)} = 1 - 2\mathcal{E}_{\mathrm{S}} = 1 - 2\left(2\varepsilon_1 - \varepsilon_2 - \varepsilon_3 + \varepsilon_4\right),$$
$$n_{\mathrm{T}}^{(1)} = -2\mathcal{E}_{\mathrm{T}} = -2\left(\varepsilon_1 + \varepsilon_3\right),$$
$$r^{(1)} = \frac{A_{\mathrm{T}}^{(1)}}{A_{\mathrm{S}}^{(1)}} = 16\left(\varepsilon_1 + \varepsilon_3\right) = -8n_{\mathrm{T}}^{(1)}. \tag{7.94}$$

These expressions for the inflationary observables coincide with the expressions found in reference [147] within the standard Heisenberg quantisation of the fluctuations propagating in the classical time-dependent FLRW background. In contrast, here the observables (7.94)-(7.94) were derived in the Schrödinger picture by performing the semiclassical expansion of the WDW equation. This result provides an important consistency check for the method, as it shows that the semiclassical expansion of the WDW equation recovers the classical background equations (7.15)-(7.17) as well as the Schrödinger equation for the fluctuations propagating on this classical background (7.61) order-by-order, which leads to the correct inflationary slow-roll observables (7.94). Finally, the next order in the semiclassical expansion yields the first quantum gravitational corrections (7.65), of which the impact on the inflationary observables will be derived in the next section.

Note that ε_1 and ε_2 in reference [147] are defined with signs opposite to the definitions (7.22) and (7.23) adopted here.

7.4.2 *Power spectra including quantum gravitational corrections*

In this subsection the contribution of the first quantum gravitational corrections to the inflationary power spectra of the general scalar-tensor theory (7.1), which is obtained at order $\mathcal{O}\left(\lambda^1\right)$ of the semiclassical expansion of the WDW equation (7.44), will be calculated. The WKB state $\Psi_n^{(2)}$ at order $\mathcal{O}\left(\lambda^1\right)$ is defined by the corrected Schrödinger equation (7.65). The quantum gravitational correction term (7.64) requires the WKB states $\Psi_n^{(1)}$ obtained at order $\mathcal{O}\left(\lambda^0\right)$ as the solution of the uncorrected Schrödinger equation (7.61). Under the assumption that the $\Psi_n^{(1)}$ have the Gaussian form (7.73), the quantum gravitational corrections (7.64) can be expressed as a function of the background potential $\mathcal{P}$, as well as the frequency ω_n^2 and the Gaussian width $\Omega_n^{(1)}$ of the scalar and tensor modes, respectively. The subindex that labels different species and modes will again be suppressed.

The quantum gravitational corrections (7.64) read

$$\mathcal{V}^{\mathrm{QG}} = \mathrm{Re}\left\{-\frac{\Omega^{(1)}\left[\Omega^{(1)} - 2\mathrm{i}\partial_\tau(\log\mathcal{P})\right] + 2\left[\left(\Omega^{(1)}\right)^2 - \omega^2\right]}{16\mathcal{P}}\right.$$
$$-\frac{\left[\mathrm{i}\partial_\tau(\log\mathcal{P}) - 3\Omega^{(1)}\right]\left[\left(\Omega^{(1)}\right)^2 - \omega^2\right] + 2\mathrm{i}\omega\partial_\tau\omega}{8\mathcal{P}}q^2$$
$$\left.-\frac{\left[(\Omega^{(1)})^2 - \omega^2\right]^2}{16\mathcal{P}}q^4\right\}. \tag{7.95}$$

Insertion of (7.95) into the corrected Schrödinger equation (7.65) with the Gaussian ansatz

$$\Psi^{(2)} = N^{(2)} \exp\left(-\tfrac{1}{2}\Omega^{(2)} q^2\right)$$

The terms independent of q only enter the equation for the normalisation factor.

yields an equation for $\Omega^{(2)}$, after all terms of equal power in q are collected:

$$\mathrm{i}\frac{\mathrm{d}\Omega^{(2)}}{\mathrm{d}\tau} = \left(\Omega^{(2)}\right)^2 - \left[\omega^2 - (\omega^{\mathrm{QG}})^2\right]. \tag{7.96}$$

Note that, since it was assumed that at each order the WKB wave function is of the Gaussian form (7.66), the terms proportional to q^4 in (7.95) are neglected. This is consistent, as the Hamiltonian (7.37) was quantised only up to quadratic terms in the perturbations, which implies that interactions among the perturbations were neglected. This truncation might also be justified on a phenomenological basis, as so far there is no observational evidence for primordial non-Gaussianities [148,149]. Therefore, the assumption that the perturbations are in their vacuum (that is, Gaussian) state seems to be a reasonable one.

The frequency ω^{QG}, which includes the quantum gravitational corrections, is defined as

$$(\omega^{\mathrm{QG}})^2 \equiv \frac{\mathrm{Re}\left(\Omega^{(1)}\right)}{4\mathcal{P}}\Big\{\mathrm{Im}\left(\Omega^{(1)}\right)\left[3\,\mathrm{Im}\left(\Omega^{(1)}\right) - 2\partial_\tau\left(\log\mathcal{P}\right)\right] - 3\left[\mathrm{Re}\left(\Omega^{(1)}\right)^2 - \mathrm{Im}\left(\Omega^{(1)}\right)^2 - \omega^2\right]\Big\}. \tag{7.97}$$

The inhomogeneous non-linear ordinary first order differential equation (7.96) is difficult to solve analytically. Since it is expected that the quantum gravitational contributions to $\Omega^{(2)}$ are small, (7.96) can be linearised around $\Omega^{(1)}$:

$$\delta\Omega \equiv \Omega^{(2)} - \Omega^{(1)}, \qquad \delta\Omega/\Omega^{(1)} \ll 1. \tag{7.98}$$

The linearisation of (7.96), together with (7.74), leads to

$$\mathrm{i}\frac{\mathrm{d}\delta\Omega}{\mathrm{d}\tau} = 2\Omega^{(1)}\delta\Omega + (\omega^{\mathrm{QG}})^2. \tag{7.99}$$

In addition, in the following it is assumed that the quantum gravitational contributions are small compared to the slow-roll contributions proportional to $\mathcal{E}\Omega_{\mathcal{E}}$ in the uncorrected Gaussian width $\Omega^{(1)}$. That is, it is assumed that $\delta\Omega/\Omega^{(1)} \ll |\varepsilon_i| \ll 1$. This implies that only the dominant De Sitter contribution is kept, which corresponds to terms $\mathcal{O}\left(\varepsilon_i^0\right)$ in the quantum gravitational frequency $(\omega^{\mathrm{QG}})^2$ and the De Sitter part Ω_{DS} in the solution $\Omega^{(1)}$. For completeness, the observational consequences that follow from the inclusion of the first order slow-roll contributions $\mathcal{O}\left(\varepsilon_i^1\right)$ to the quantum gravitational corrections have been worked out in **appendix B**. The dominant De Sitter contribution to $(\omega_{\mathrm{QG}})^2$ is

$$(\omega^{\mathrm{QG}})^2 \approx \frac{\mathrm{Re}\,(\Omega_{\mathrm{DS}})}{4\mathcal{P}}\Big\{\mathrm{Im}\,(\Omega_{\mathrm{DS}})\,[3\,\mathrm{Im}\,(\Omega_{\mathrm{DS}}) - 2\partial_\tau\,(\log\mathcal{P})]$$
$$- 3\left[\mathrm{Re}\,(\Omega_{\mathrm{DS}})^2 - \mathrm{Im}\,(\Omega_{\mathrm{DS}})^2 - \omega_{\mathrm{DS}}^2\right]\Big\}. \tag{7.100}$$

In order to proceed, derivatives with respect to conformal time τ can be expressed in terms of the variable x by $\partial_\tau = -k\partial_x$. Equation (7.100) can be simplified through the use of (7.77), (7.78), (7.84) and the background equations of motion in the slow-roll approximation (7.20). The potential $\mathcal{P}$ can then be expressed in terms of the constant conformal Hubble parameter in De Sitter space $\mathcal{H}$ and the non-minimal coupling U. The dominant De Sitter contribution to the quantum gravitational frequency is then found to be

$$(\omega_{\mathrm{DS}}^{\mathrm{QG}})^2 \approx -\frac{W}{144k}\frac{x^4(x^2-11)}{(x^2+1)^3}. \tag{7.101}$$

In order to solve (7.99) with the source (7.101) one can again impose the asymptotic Bunch-Davis boundary condition for $\Omega^{(2)}$. That is, the Gaussian width $\Omega^{(2)}$ is again required to be time-independent in the limit $\tau \to -\infty$ ($x \to \infty$). In this case, (7.96) reduces to the algebraic condition

$$\left(\Omega_\infty^{(2)}\right)^2 = k^2 + \frac{W}{144k},$$

where it has been used that

$$\lim_{x\to\infty}(\omega_{\mathrm{DS}}^{\mathrm{QG}})^2 = -\frac{W}{144k}.$$

With (7.82), this implies the initial condition $\delta\Omega_\infty$ for $\delta\Omega_\infty$:

$$\delta\Omega_\infty = \frac{\Omega_\infty^{(1)}}{2}\left[\left(\frac{\Omega_\infty^{(2)}}{\Omega_\infty^{(1)}}\right)^2 - 1\right] = \frac{W}{288k^2}. \tag{7.102}$$

The linearised differential equation for the De Sitter part of the quantum gravitational corrections then reads

$$\frac{\mathrm{d}}{\mathrm{d}x}\delta\Omega_{\mathrm{DS}} = 2\mathrm{i}\left(\frac{x^2-\mathrm{i}x^{-1}}{x^2+1}\right)\delta\Omega_{\mathrm{DS}} - \mathrm{i}\frac{W}{144k^2}\frac{x^4(x^2-11)}{(x^2+1)^3}. \tag{7.103}$$

With the boundary condition (7.102), the solution reads

$$\delta\Omega_{\mathrm{DS}} = \frac{W}{288k^2}\frac{\mathrm{e}^{2\mathrm{i}x}x^2}{(x-\mathrm{i})^2}\Big[3\mathrm{i}\pi\frac{3+\mathrm{e}^4}{\mathrm{e}^2} + \mathrm{e}^{-2\mathrm{i}x}\frac{1+x\,(x-6\mathrm{i})}{(x+\mathrm{i})^2}$$
$$+ 3\mathrm{e}^2\,\mathrm{Ei}(-2-2\mathrm{i}x) + 9\mathrm{e}^{-2}\,\mathrm{Ei}(2-2\mathrm{i}x)\Big]. \tag{7.104}$$

The power spectrum (7.67), obtained from the Gaussian width $\Omega^{(2)}$, is found to be

$$P^{(2)} = \frac{k^3}{2\pi^2} \frac{1}{2\mathrm{Re}\left[\Omega^{(1)} + \delta\Omega_{\mathrm{DS}}\right]} \approx P^{(1)} \left[1 - \frac{\mathrm{Re}\left(\delta\Omega_{\mathrm{DS}}\right)}{\mathrm{Re}\left(\Omega^{(1)}\right)}\right]. \tag{7.105}$$

The superhorizon limit ($x \ll 1$) of the real part of the solution (7.104) then yields

$$\mathrm{Re}\left(\delta\Omega_{\mathrm{DS}}\right) = \frac{\beta_0 W}{144} \frac{x^2}{k^2} + \mathcal{O}\left(x^4\right) \approx -\frac{W}{72} \frac{x^2}{k^2}, \tag{7.106}$$

with the numerical constant $\beta_0 \equiv [1 - 3\mathrm{e}^2\,\mathrm{Ei}(-2) - 9\mathrm{e}^{-2}\,\mathrm{Ei}(2)]/2 \approx -2$. The De Sitter contribution to $\Omega^{(1)}$ in the superhorizon limit reads

$$\mathrm{Re}\left(\Omega^{(1)}\right) \approx \mathrm{Re}\left(\Omega_{\mathrm{DS}}\right) \approx kx^2. \tag{7.107}$$

The ratio $\mathrm{Re}\left(\delta\Omega_{\mathrm{DS}}\right)/\mathrm{Re}\left(\Omega^{(1)}\right)$ can, with (7.106) with (7.107), be found to be

$$P^{(2)}(k) = P^{(1)}(k)\left[1 + \delta_{\mathrm{DS}}^{\mathrm{QG}}(k/k_0)\right], \qquad \delta_{\mathrm{DS}}^{\mathrm{QG}}(k/k_0) \equiv \frac{W}{72}\left(\frac{k_0}{k}\right)^3. \tag{7.108}$$

The reference wavelength $k_0 = l_0^{-1}$, which originated from reversing the rescalings (7.18) and (7.35), has been reintroduced in the last step. Note that, in particular, the uncorrected part of the power spectrum $P^{(1)}$ is invariant under this rescaling due to its logarithmic dependence on the invariant ratio $k/\mathcal{H}$—only the quantum gravitational corrections are affected.

Finally, the corrected scalar and tensor spectra can, with the definitions of the scalar and tensor power spectra (7.68) and the results obtained in the previous order of the expansion (7.88) and (7.89), be found to be

$$P_{\mathrm{S}}^{(2)}(k) \approx \frac{W}{96\pi^2(\varepsilon_1 + \varepsilon_3)}\Big[1 - \tfrac{1}{3}\left(5\varepsilon_1 - 6c_\gamma \mathcal{E}_{\mathrm{S}}\right) - 2\mathcal{E}_{\mathrm{S}} \log\left(k/\mathcal{H}\right) + \delta_{\mathrm{DS}}^{\mathrm{QG}}(k/k_0)\Big], \tag{7.109}$$

$$P_{\mathrm{T}}^{(2)}(k) \approx \frac{W}{6\pi^2}\Big[1 - \tfrac{1}{3}(5\varepsilon_1 + 6\varepsilon_3 - 6c_\gamma \mathcal{E}_{\mathrm{T}}) - 2\mathcal{E}_{\mathrm{T}} \log(k/H) + \delta_{\mathrm{DS}}^{\mathrm{QG}}(k/k_0)\Big]. \tag{7.110}$$

Note that (7.109) and (7.110) only consider the dominant De Sitter contribution (7.108) of the quantum gravitational corrections. The particular features and observable consequences of the corrected power spectra (7.109) and (7.110) are discussed in more detail in the next section.

7.5 OBSERVATIONAL SIGNATURES OF THE CORRECTIONS

In this section several features of the quantum gravitational corrections to the power spectra, such as their observable signatures, their magnitude and the impact of the non-minimal coupling, will be discussed.

a. Just as the uncorrected power spectra **(7.91)**, **(7.92)**, the quantum gravitationally corrected power spectra **(7.109)** and **(7.110)** become time-independent at superhorizon scales. This is important, as it allows one to calculate the power spectrum in the superhorizon limit at horizon crossing.

b. The dominant De Sitter part of the quantum gravitational corrections $\delta^{\mathrm{QG}}_{\mathrm{DS}}$ is universal, in the sense that it equally contributes to the scalar and tensor power spectrum. In particular, this implies that these corrections will drop out in the tensor-to-scalar ratio, as found in the minimally coupled case [109]. This degeneracy between the scalar and tensorial spectra can be broken by the inclusion of slow-roll contributions to the quantum gravitational corrections. Although these slow-roll contributions are additionally suppressed by powers of the slow-roll parameters and therefore even less relevant than the already minuscule dominant De Sitter part of the quantum gravitational corrections, the analysis is of theoretical interest and for completeness included in **appendix B**.

c. The quantum gravitational effects lead to an enhancement of the power spectra.

d. The quantum gravitational corrections have a characteristic $1/k^3$-dependence, which would—at least in principle—allow one to observationally distinguish between the quantum gravitational contributions and the uncorrected constant De Sitter and logarithmic k-dependent slow-roll parts of the power spectrum.

e. The quantum gravitational corrections are heavily suppressed relative to the uncorrected part of the power spectra. From the $(k_0/k_*)^3$ dependence it is clear that the quantum gravitational corrections are large on the largest scales (smallest values of k_*) and also depend on the infrared regularising reference scale $k_0 \sim \ell_0^{-1}$. The latter is undetermined ***a priori***. Note that the uncorrected part of the power spectra is independent of the reference scale k_0.

   More precise statements about the magnitude of the quantum gravitational corrections can be made by adopting the standard power law parametrisation of the power spectra **(7.69)**, in which the quantum gravitational corrected power spectra **(7.109)** and **(7.110)** are characterised by their amplitudes and spectral indices:

$$
\begin{aligned}
A_{\mathrm{S}}^{(2)} &\approx \frac{W_*}{96\pi^2(\varepsilon_1+\varepsilon_3)}\Big[1-\tfrac{1}{3}\left(5\varepsilon_1-6c_\gamma\mathcal{E}_{\mathrm{S}}\right)+\delta^{\mathrm{QG}}_{\mathrm{DS}}(k_*/k_0)\Big], \qquad (7.111)\\
A_{\mathrm{T}}^{(2)} &\approx \frac{W_*}{6\pi^2}\left[1-\tfrac{1}{3}\left(5\varepsilon_1+6\varepsilon_3-6c_\gamma\mathcal{E}_{\mathrm{T}}\right)+\delta^{\mathrm{QG}}_{\mathrm{DS}}(k_*/k_0)\right],\\
n_{\mathrm{S}}^{(2)} &\approx 1-2\left(2\varepsilon_1-\varepsilon_2-\varepsilon_3+\varepsilon_4\right)-3\delta^{\mathrm{QG}}_{\mathrm{DS}}(k_*/k_0),\\
n_{\mathrm{T}}^{(2)} &\approx -2\left(\varepsilon_1+\varepsilon_3\right)-3\delta^{\mathrm{QG}}_{\mathrm{DS}}(k_*/k_0),\\
r^{(2)} &\approx \frac{A_{\mathrm{T}}^{(2)}}{A_{\mathrm{S}}^{(2)}} \approx r^{(1)} = 16(\varepsilon_1+\varepsilon_3) \neq -8n_{\mathrm{T}}^{(2)}. \qquad (7.112)
\end{aligned}
$$

W_* denotes the value of W at the moment $k_* = \mathcal{H}_*$, when the pivot mode k_* first crosses the horizon. Note that, according to **(7.112)**, the quantum gravitational corrections lead to a tiny violation of the consistency condition **(7.72)**. The power-law ansatz **(7.69)** is usually justified by the weak logarithmic scale dependence of the power spectra. While this is true for the uncorrected part, in view of the $1/k^3$ dependence of the quantum gravitational corrections, it seems questionable whether such a parametrisation is adequate. The pivot scale is chosen within the window of scales observable in the CMB [84]:

$$k_*^{\text{min}} < k_* < k_*^{\text{max}}, \quad k_*^{\text{min}} = 10^{-4}\,\text{Mpc}^{-1}, \quad k_*^{\text{max}} = 10^{-1}\,\text{Mpc}^{-1}. \tag{7.113}$$

Measurements of the CMB constrain A_{S} and n_{S} and give an upper bound on the tensor-to-scalar ratio r, here quoted for $k_* = 5 \times 10^{-2}\,\text{Mpc}^{-1}$ [84]:

$$A_{\text{S},*}^{\text{obs}} = (2.099 \pm 0.014) \times 10^{-9}, \quad 68\,\% \;\; \text{CL} \tag{7.114}$$

$$n_{\text{S},*}^{\text{obs}} = 0.9649 \pm 0.0042, \quad 68\,\% \;\; \text{CL}, \tag{7.115}$$

$$r_*^{\text{obs}} < 0.11, \quad 95\,\% \;\; \text{CL}. \tag{7.116}$$

The upper bound on r_*^{obs} is connected to an upper bound on the energy scale during inflation. This corresponds to an upper bound on the energy density, given by the potential $\tilde{V} = \frac{1}{4}M_{\text{P}}^4 W$ in the EF **(3.11)**. In terms of the inflationary observables, $\tilde{V}_*$ reads

$$\tilde{V}_* = \left(\frac{M_{\text{P}}^4 W_*}{4}\right) \approx \tfrac{3}{2}\pi^2 M_{\text{P}}^4 A_{\text{T},*}^{(2)} \approx \tfrac{3}{2}\pi^2 M_{\text{P}}^4 A_{\text{S},*}^{(2)}\, r_*^{(2)}. \tag{7.117}$$

With the identification $A^{(2)} \approx A^{(1)} \approx A^{\text{obs}}$ and $r^{(2)} \approx r^{(1)} \approx r^{\text{obs}}$, **(7.114)** and **(7.116)** imply the upper bound

It would be interesting to compare the energy scale of inflation, including quantum gravitational corrections, with the energy scale derived from quantum cosmology [118, 120, 121].

$$\frac{W_*}{72} \approx \frac{\left(6\pi^2 A_{\text{S},*}^{\text{obs}}\, r_*^{\text{obs}}\right)}{72} \lesssim 10^{-10}. \tag{7.118}$$

The estimate **(7.118)** shows that, independently of the choices for U, G, and V in **(7.1)**, for a viable inflationary model the quantum gravitational corrections to the power spectrum are suppressed relative to the uncorrected part by

$$\delta_{\text{DS}}^{\text{QG}}(k_*/k_0) \lesssim 10^{-10}\left(\frac{k_0}{k_*}\right)^3.$$

Therefore, the only way to enhance the impact of the quantum gravitational corrections is to increase the ratio k_0/k_*. The measurements **(7.114)-(7.116)** are obtained at a fixed pivot point $k_* = 5 \times 10^{-2}\,\text{Mpc}^{-1}$. However, this reference scale is in principle arbitrary and only constrained to lie with the interval **(7.113)**, which is accessible to observations of the CMB. Since the leading quantum gravitational correction $\delta_{\text{DS}}^{\text{QG}}$ is the largest for the smallest value of k_*, experiments should be most sensitive to a detection of a potential quantum gravitational effect at k_*^{min}.

Although the infrared scale k_0 is unspecified, an upper bound on k_0 might be derived along the lines of the discussion in reference [110]. On the one hand, the observed scalar amplitude of the perturbations has a measured value $A_{S,*}^{obs}$ with experimental uncertainty $\delta_{exp} \approx 0.014$. On the other hand, the dominant quantum gravitational correction to the amplitude reads $A_S^{(2)} = A_S^{(1)}\left(1 + \delta_{DS}^{QG}\right)$. The quantum gravitational effects must satisfy $\delta_{DS}^{QG} > \delta_{exp}$ in order to be detectable. Conversely, the absence of a quantum gravitational effect implies an upper bound on k_0 from $\delta_{DS}^{QG} \leq \delta_{exp}$. This condition can be expressed in terms of observable quantities with the use of **(7.108)** and **(7.117)**. This results in the inequality

$$k_0 \leq \left[\frac{72}{6\pi^2 A_{S,*}^{obs}} \frac{\delta_{exp}}{r_*^{obs}}\right]^{1/3} k_*. \tag{7.119}$$

An upper bound k_0^{max} for k_0 is obtained for if **(7.119)** is an equality.

Equation **(7.119)** largely depends on the ratio δ_{exp}/r_*^{obs}. It can be seen that an increase in experimental precision will lower δ_{exp} and therefore lower k_0^{max}. Conversely, since the tensor-to-scalar ratio r_*^{obs} is only bounded from above, a measurement of r_*^{obs} smaller than the upper bound **(7.116)** would increase k_0^{max}. The values from **(7.114)-(7.116)** can be inserted into **(7.119)** in order to get a rough estimate of the order of magnitude. With $\delta_{exp} = 0.014$ and $r_*^{obs} = 0.11$ at $k_* = 0.05$, one obtains

$$k_0^{max} \approx 21\,\mathrm{Mpc}^{-1}. \tag{7.120}$$

Note that the higher bound $\delta_{exp} = 1$ leads to the value $k_0^{max} \approx 87\,\mathrm{Mpc}^{-1}$ for a pivot point $k_0 = 0.05\,\mathrm{Mpc}^{-1}$, which is in agreement with the estimate obtained in reference [110].

The length scale that corresponds to **(7.120)** is $l_0^{max} = 1/k_0^{max} \approx 0.048\,\mathrm{Mpc}^{-1}$. The quantum gravitational effects would be resolvable within the assumed precision for reference scales $l_0 < l_0^{max}$. This can be compared to the natural choice for the infrared cutoff, the radius of the observable universe. This corresponds to a scale $k_0 \approx k_*^{min}$. Since the maximum value for the quantum gravitational effects is obtained for $k_* = k_*^{min}$, the ratio is $(k_0/k_*)^3 = 1$. If one assumes that $A_{S,*}^{obs}$ and r_*^{obs} do not change drastically under a change of the pivot point from $k_* = 0.05\,\mathrm{Mpc}^{-1}$ to $k_* = k_*^{min}$, the dominant quantum gravitational corrections to the power spectra is of order

$$\delta_{DS}^{QG} \approx 10^{-10}.$$

f. Finally, one should recover the results obtained in reference [109] for the minimally coupled case as a consistency check of the results obtained here. Indeed, after substitution of the constant non-minimal coupling function $U(\phi) = M_P^2/2$ and $V = \mathcal{V}$ into $W = V/U^2$, the results **(7.109)** and **(7.110)** reduce to the corresponding expressions obtained in reference [109]. Therefore, the main impact of the non-minimal coupling U on the quantum gravitational corrections corresponds to a replacement of the constant Planck mass by the effective field-dependent Planck mass $M_{\mathrm{P}}^2 \to M_{\mathrm{P}}^2(\phi) = 2U(\phi)$—a result that might have been expected naively.

However, despite the arbitrariness in the field dependent non-minimal coupling U, the ratio $W = V/U^2$ is constrained by observations for any viable scalar-tensor theory of inflation and therefore sets an upper bound on the magnitude of the dominant quantum gravitational corrections. Thus, for any scalar-tensor theory of the form **(7.1)** which is consistent with observational data, the non-minimal coupling U does not lead to an enhancement of the quantum gravitational corrections. Nevertheless, the impact of the generalised potentials U, G and V enters in the subleading slow-roll contribution to the quantum gravitational corrections, discussed in **appendix B**.

7.6 CONCLUSION

In this chapter the first quantum gravitational corrections to the inflationary power spectra for a general scalar-tensor theory were calculated from a semi-classical expansion of the WDW equation.

The general scalar-tensor action was expanded around a flat FLRW background up to quadratic order in the perturbations. The inhomogeneous perturbations were expanded in terms of their Fourier modes, after which the combined background and perturbation variables were quantised in a canonical way. The quantisation of the Hamilton constraint lead to the WDW equation which describes the exact quantum dynamics. The WDW equation was then expanded around a semiclassical solution based on a combined Born-Oppenheimer and WKB-type approximation. The Born-Oppenheimer approximation relies on the division of the configuration space variables into heavy and light degrees of freedom. In the cosmological context the background variables are naturally identified with the heavy degrees of freedom, while the Fourier modes of the perturbations are identified with the light degrees of freedom. At the lowest order in the semiclassical expansion, the classical homogeneous background equations of motion and the notion of a background-dependent semiclassical time were recovered from the timeless WDW equation. At the next order, the Schrödinger equation for the perturbations was obtained. The perturbations were found to evolve with respect to this semiclassical time. Finally, at the subsequent order in the expansion, the first quantum gravitational contributions were derived. These could be represented in the form of a corrected Schrödinger equation. In order to extract the observational consequences, the inflationary power spectra were calculated with the assumption that at each order of the expansion the semiclassical wavefunction has a Gaussian form and satisfies the asymptotic Bunch-Davies boundary condition at early times. For such Gaussian states, the inflationary power spectra are fully determined by the real part of the Gaussian width. The assumption about the Gaussian nature of the wave function is natural if the system is in the ground state. A recent extension to excited states can be found in reference [142].

Under these assumptions, the standard power spectra were recovered at the level of the uncorrected Schrödinger equation. The power spectra are standard in the sense that they coincide with the results that are usually obtained in the Heisenberg quantisation of gauge invariant perturbations that propagate on a classical background. In contrast, in the approach presented here these results follow directly from the semiclassical expansion of the WDW equation. This shows that the semiclassical expansion not only correctly reproduces the classical background equations, but also the inflationary power spectra. Therefore, this provides an important consistency check for the semiclassical treatment to the geometrodynamical approach to quantum gravity. Finally, the first quantum gravitational corrections to the inflationary power spectra were derived from the corrected Schrödinger equation. Since these quantum gravitational corrections are highly suppressed, the analysis in this chapter was restricted to the dominant De Sitter contribution. The first order slow-roll contributions to the quantum gravitational corrections are derived separately in **appendix B**.

It was found that the dominant quantum gravitational corrections for a general scalar-tensor theory lead to an increase in the amplitude of the inflationary scalar and tensor power spectra. This increase is universal in the sense that it affects both power spectra in the same way. However, even the dominant quantum gravitational corrections are strongly suppressed compared to the uncorrected part of the power spectra. This is in agreement with previous results obtained for a minimally coupled scalar field [109,110,137]. Although the non-minimal coupling U enters the quantum gravitational corrections to the power spectra, it only enters in the dimensionless combination $W = V/U^2$. Since the uncorrected power spectra are proportional to W, observations put strong constraints on the value $W_* \lesssim 10^{-9}$ at horizon crossing and therefore on the magnitude of the quantum gravitational corrections. This implies that, independently of the concrete choice for the generalised potentials U, G and V present in the general scalar-tensor theory **(7.1)**, the dominant quantum gravitational corrections are strongly suppressed as long as the observational constraints for a successful phase of inflation are satisfied. In particular, this shows that it is not possible to enhance the quantum gravitational corrections to the inflationary power spectra due to the presence of a non-minimal coupling. The impact of the generalised potentials on the quantum gravitational corrections only affects the subleading slow-roll contributions.

The quantum gravitational corrections feature a characteristic scale dependence proportional to k^{-3}, independently of the strong suppression factor. This has been found in similar approaches [109,110,127,137,139,150]. This scale dependence is not only a prediction of quantum gravity but also provides an observational signature. In fact, the scale dependence of the quantum gravitational corrections enters the power spectra in the form $(k_0/k_*)^3$, where k_* is the pivot point and k_0 the infrared regulating scale which arises in the flat FLRW universe. Although the quantum gravitational corrections are suppressed by $W_*/72 \approx 10^{-10}$, depending on the values for k_0 and k_*, the scaling factor $(k_0/k_*)^3$ might increase the magnitude of the quantum gravitational corrections. While the value of k_* is constrained to lie within the observable window, the value of k_0 is *a priori* undetermined. Quantum gravitational effects are at their strongest for a pivot point at the lower end of the allowed interval $k_* = k_*^{\mathrm{min}} \approx 10^{-4}\,\mathrm{Mpc}^{-1}$. A natural choice for the infrared regulating scale l_0 is the size of the observable universe, which corresponds to a scale $k_0 = k_*^{\mathrm{min}}$. For these choices of k_*

and k_0, the ratio $(k_0/k_*)^3$ is of order unity. In this case, quantum gravitational effects are suppressed by a factor of 10^{-10}. They are therefore at present unobservable. Conversely, a value $k_0 \approx 10^{-1}\,\mathrm{Mpc}^{-1}$ would be required for quantum gravitational effects to come into observational reach. This, in turn, would single out a preferred astrophysical length scale $l_0 \approx 10\,\mathrm{Mpc}$. However, since the reference scale l_0 was introduced to regularise the infinite spatial volume integral arising in a homogeneous and isotropic flat FLRW universe, such a value seems to be inconsistent as it is well below the smoothing scale of approximately 100 Mpc. The conclusion is therefore that, within the available precision of the current observations, quantum gravitational from the semiclassical expansion of the WDW equation are unobservable—even for general scalar-tensor theories.

8 DISCUSSION AND OUTLOOK

IN THIS WORK the formalism of quantum geometrodynamics was extended to a general single-field scalar-tensor theory. Since the formalism of quantum geometrodynamics is complicated on both technical and conceptual grounds, the analysis described in this work is necessarily formal. However, it is reasonable to expect that the results obtained have physical relevance, as they were derived in a systematic semiclassical expansion of the Wheeler-DeWitt equation. This expansion satisfies the correspondence with classical physics and quantum physics, as these are reproduced at the lowest levels.

In **chapter 5** the canonical formalism was presented for a general class of scalar-tensor theories. The diffeomorphism invariance of the theory gives rise to constraints. The quantisation of these constraints resulted in a non-perturbative theory of quantum gravity, where the quantum dynamics is dictated by the Wheeler-DeWitt equation. It was then shown that, if some of the degrees of freedom can be considered to be semiclassical in certain regions of configuration space, the wave function of the system could be given a probabilistic interpretation, as a notion of unitarity could be defined in the semiclassical approximation with respect to the semiclassical time t_s that emerged from the Hamilton-Jacobi equation for the background variables. It is still an open question whether unitarity exists at all orders in the semiclassical expansion, and is an active topic of discussion in the literature. This is a question that is left for further work. The interpretation of the wave function in quantum geometrodynamics is not well understood outside of semiclassical domains of superspace. It is not clear whether a notion of unitarity can be defined if not at least one degree of freedom in superspace is semiclassical, as one would have to resolve negative probabilities. In the semiclassical domains of superspace negative probabilities can be resolved by means of a foliation in superspace in terms of appropriate hypersurfaces. These hypersurfaces are provided by the solution to the Hamilton-Jacobi functional of the semiclassical degrees of freedom. It is unclear whether such a function exists if all degrees of freedom are treated as fully quantised. This, too, would be an interesting topic to pursue in a further study.

In **chapter 6** it was found that the semiclassical expansion of the WDW equation for general scalar-tensor theories is complicated by the absence of an energy scale hierarchy, in contrast to the semiclassical expansion for minimally coupled theories. Although a distinction between the heavy and light degrees of freedom can be made by appealing to the classical frame equivalence, the resulting Schrödinger equation depends explicitly on the factor ordering, again in contrast to the results for minimally coupled theories. This reflects the fact that, in contrast to a minimally coupled scalar field in the formalism of the theory of general relativity, the scalar field in scalar-tensor theories is a part of the gravitational interaction, and cannot be straightforwardly isolated from the metric degrees of freedom. The factor ordering that was chosen was the Laplace-Beltrami factor ordering, which preserves the equivalence between the Jordan and Einstein frame after quantisation. In addition, it was found that the quantum gravitational corrections to the effective Schrödinger equation are suppressed by the non-minimal coupling function. The differences between minimally coupled theories and non-minimally coupled theories were briefly discussed in the specific model of Higgs inflation. Since, in this case, the non-minimal coupling is required to be strong, it can be expected that the corrections are too small to be measured by current detectors. However, an interesting result was that the kinetic part of the Hamiltonian of the light degree of freedom are suppressed at large field values. This coincides with studies of the perturbative covariant approach of Higgs inflation, where a similar suppression mechanism was found for the Higgs propagator. However, it is unclear what this means from the perspective of the semiclassical expansion of the WDW equation. This question is left for further study.

Finally, in **chapter 7** the implications for a non-minimal coupling in a general class of scalar-tensor theories on the quantum gravitational corrections to the inflationary power spectra of gauge-invariant cosmological perturbations were investigated. The general Hamiltonian that was derived in **chapter 6** was evaluated for a homogeneous and isotropic FLRW universe, on which propagate small inhomogeneous perturbations. The Wheeler-DeWitt equation was split into two equations—an equation that determines the evolution of the FLRW background geometry and an equation that determines the cosmic perturbations. Both equations were solved using an iterative semiclassical expansion, which was truncated at the level of the first quantum gravitational corrections. The semiclassical expansion of the background equation resulted in the classical Friedmann equations, which gave rise to a semiclassical concept of time from the timeless Wheeler-DeWitt equation. The semiclassical expansion of the equation for the perturbations gave rise to an effective Schrödinger equation, which includes small corrections due to the quantised nature of the FLRW background. The effective Schrödinger equation was used to derive the mode equations for the inflationary power spectra. The quantum gravitational corrections from the Schrödinger equation gave rise to small perturbations in the mode equations. The dominant perturbations were solved analytically, while the subdominant perturbations could be solved numerically, for a general scalar-tensor theory. It was found that the corrections are proportional to the conformally invariant potential $W = V/U^2$. The corrections are therefore too small to be measured by current experiments, as observations from the *Planck* experiment result in the upper bound $W \leq 10^{-9}$.

The results obtained in this work correctly reproduce the results obtained in the semiclassical expansion of the Wheeler-DeWitt equation for gravity with a minimally coupled scalar field. The dominant difference of the quantum gravitational corrections for a general non-minimal coupling is the formal replacement of the Planck mass with the non-minimal coupling function U. Subdominant differences arise in the slow-roll parameters, which carry information about the derivatives of the functions U, G and V of the model. The homogeneity of the FLRW framework requires the introduction of a regulating cutoff scale. It is not yet clear what this cutoff scale represents physically. In the main text it is assumed to be within the observable window of experiments, such as *Planck*, although there seems to be no *a priori* reason why this should be so. The resolution of this apparent ambiguity could provide key insights of the applicability of the FLRW framework in the very early universe, and is left for further work.

In this appendix the configuration space is formally considered as a (formally infinite-dimensional) differentiable manifold with line element

$$\mathrm{d}s^2 = \int \mathcal{M}_{AB}\, \mathrm{d}q^A \mathrm{d}q^B \mathrm{d}^d x.$$

Singular delta functions that arise from functional differentiation at the same point are suppressed. Spacetime is assumed to be of dimensionality d and signature ε. The starting point is the action

$$S[g,\phi] = \int_{\mathcal{M}} \left(U R + \tfrac{1}{2} G \nabla_\mu \phi \nabla^\mu \phi - V \right) \sqrt{\varepsilon g}\, \mathrm{d}^d X,$$

which can be foliated in terms of spatial hypersurfaces Σ_t. The components of the configuration space metric can be read off in the same way as in **chapter 6**:

$$\mathcal{M}_{AB} = \frac{\varepsilon \sqrt{\gamma}}{N} \begin{pmatrix} -\frac{1}{2} U G^{abcd} & U_1 \gamma^{ab} \\ U_1 \gamma^{cd} & -G \end{pmatrix}.$$

A.1 Christoffel symbols

The Christoffel symbol constructed from the configuration space metric $\mathcal{M}_{AB}$ reads

$$\Gamma^C_{AB} = \frac{1}{2} \mathcal{M}^{CD} \left(\delta_A \mathcal{M}_{DB} + \delta_B \mathcal{M}_{AD} - \delta_D \mathcal{M}_{AB} \right).$$

For the explicit components of the Christoffel symbol, one finds

$$\Gamma^{\phi}{}_{\phi\phi} = -\frac{1}{2} \left(\frac{s_1}{s} + \frac{2d}{(d-1)^2} \frac{s U_1^3}{U^2} - \frac{d}{d-1} \frac{U_1}{U} \right), \tag{A.1}$$

$$\Gamma^{\phi ab}{}_{\phi} = \frac{1}{4} \left(1 - \frac{2}{d-1} \frac{s U_1^2}{U} \right) \gamma^{ab}, \tag{A.2}$$

$$\Gamma^{\phi abcd} = \frac{1}{4} \frac{s U_1}{d-1} G^{abcd}, \tag{A.3}$$

$$\Gamma_{ab\phi\phi} = \frac{1}{2(d-1)} \left[\frac{1}{sU} - \frac{4d}{d-1} \left(\frac{U_1}{U} \right)^2 + 2 \frac{s_1}{s} \frac{U_1}{U} + 4 \frac{U_2}{U} + \frac{4d}{(d-1)^2} \frac{s U_1^4}{U^3} \right] \gamma_{ab}, \tag{A.4}$$

$$\Gamma_{ab\phi}{}^{cd} = \frac{1}{2} \frac{U_1}{U} \left[\delta^{cd}_{ab} - \frac{1}{d-1} \left(1 - \frac{2}{d-1} \frac{s U_1^2}{U} \right) \gamma_{ab} \gamma^{cd} \right], \tag{A.5}$$

$$\Gamma_{ab}{}^{efcd} = \frac{1}{2}\delta_{ab}^{((cd}\gamma^{ef))} - \delta_{ab}^{(c(e}\gamma^{f)d)} + \frac{1}{4(d-1)}\left(1 - \frac{2}{d-1}\frac{sU_1^2}{U}\right)\gamma_{ab}G^{cdef}. \tag{A.6}$$

It can be verified by construction and direct substitution of **(A.1)**-**(A.6)** that the Christoffel symbol Γ^C_{AB} satisfies the metric compatibility condition

$$\nabla_A \mathcal{M}_{BC} = 0.$$

A.2 METRIC DETERMINANT

Since the DeWitt metric $\gamma^{\frac{1}{2}}G^{abcd}$ is invertible for non-singular metrics, the determinant of the configuration space metric **(6.11)** can be calculated by

$$\mathcal{M} \equiv \det\left(\mathcal{M}_{AB}\right) = \det\left(\mathcal{M}_{\gamma\gamma}\right)\det\left(\mathcal{M}_{\phi\phi} - \mathcal{M}_{\phi\gamma}\mathcal{M}^{-1}_{\gamma\gamma}\mathcal{M}_{\gamma\phi}\right). \tag{A.7}$$

Using **(6.11)** and **(6.15)**, the second determinant gives

$$\det\left(\mathcal{M}_{\phi\phi} - \mathcal{M}_{\phi\gamma}\mathcal{M}^{-1}_{\gamma\gamma}\mathcal{M}_{\gamma\phi}\right) = -\frac{\varepsilon\gamma^{\frac{1}{2}}}{Ns}. \tag{A.8}$$

The remaining determinant $\det\left(\mathcal{M}_{\gamma\gamma}\right)$ can be expressed in terms of the determinant of the DeWitt metric:

$$\det\left(\mathcal{M}_{\gamma\gamma}\right) = \det\left(-\frac{\varepsilon U}{2N}\gamma^{1/2}G^{abcd}\right) = \left(-\frac{\varepsilon U}{2N}\right)^{\frac{d(d+1)}{2}}\det\left(\gamma^{1/2}G^{abcd}\right), \tag{A.9}$$

where in the last step it was used that the space of symmetric rank two tensor fields is $\frac{1}{2}d(d+1)$-dimensional. The trace of the unit matrix in configuration space gives

$$\operatorname{tr}\delta^A_B = \tfrac{1}{2}d(d+1) + 1.$$

The determinant of the DeWitt metric $\gamma^{1/2}G^{abcd}$ is well known [94], and can be obtained by variation of the DeWitt metric $\gamma^{1/2}G^{abcd}$ with respect to γ_{ab}. The result reads

$$\det\left(\gamma^{1/2}G^{abcd}\right) = -\alpha\left(\gamma^{1/2}\right)^{\frac{(d+1)(d-4)}{2}}. \tag{A.10}$$

Here α is some positive constant [94]. The explicit value of this constant is irrelevant for the results presented in this work as it cancels in all relevant expressions. Combining **(A.8)** with **(A.9)** and **(A.10)** one obtains the determinant of the configuration space metric (for $d \geq 3$):

$$\mathcal{M} = \frac{\alpha\varepsilon}{Ns}\left(\frac{\gamma^{1/2}U}{2N}\right)^{\frac{d(d+1)}{2}}\left(\gamma^{1/2}\right)^{-(2d+1)}, \tag{A.11}$$

where $\varepsilon^2 = 1$ was used in the last equality. As a consistency check $\mathcal{M}$ can be calculated from the metric (6.11) and the Christoffel symbols (A.1)-(A.6) as the solution to the differential equation

$$\delta_A \ln \mathcal{M}^{1/2} = \Gamma^B_{AB}.$$

This reproduces the same expression (A.9). In this approach α arises as constant of integration. Note that in the purely gravitational case, the signature of the configuration space metric (DeWitt metric) is indefinite, independent of the signature of spacetime. In contrast, (A.11) implies that the signature of the configuration space metric $\mathcal{M}_{AB}$ does depend on the signature ε of the metric $g_{\mu\nu}$ in the $(d+1)$-dimensional ambient space $\mathcal{M}$, due to the extra scalar degree of freedom ϕ in configuration space.

A.3 RIEMANN TENSOR CONFIGURATION SPACE

Given the expressions for the Christoffel symbols (A.1)-(A.6), it is straightforward to calculate the non-vanishing components of the configuration space Riemann tensor

$$\mathcal{R}^A{}_{BCD} = \delta_C \Gamma^A_{BD} - \delta_D \Gamma^A_{BC} + \Gamma^A_{CE}\Gamma^E_{BD} - \Gamma^A_{DE}\Gamma^E_{BC}.$$

It is convenient to express the components of the Riemann tensor with two indices raised $\mathcal{R}^{AB}{}_{CD} = \mathcal{M}^{BF}\mathcal{R}^A{}_{FCD}$. There are four independent non-vanishing components:

$$\begin{aligned}
\mathcal{R}_{abcd}{}^{efgh} = &-\frac{\varepsilon N\gamma^{-\frac{1}{2}}}{4(d-1)}\left[\frac{d}{U} - \frac{2}{d-1}s\left(\frac{U_1}{U}\right)^2\right]\delta^{[[ef}_{[[ab}\delta^{gh]]}_{cd]]} \\
&+\frac{\varepsilon N\gamma^{-\frac{1}{2}}}{2(d-1)}\left[\frac{1}{U} - \frac{2}{d-1}s\left(\frac{U_1}{U}\right)^2\right]\delta^{[[ef}_{[[ab}\gamma_{cd]]}\gamma^{gh]]} \\
&+\frac{2\varepsilon\gamma^{-\frac{1}{2}}}{U}\delta^{[[(e}_{(a}\gamma^{f)(g}\delta^{h)]]}_{(c}\gamma_{d)b)},
\end{aligned} \tag{A.12}$$

$$\mathcal{R}_{abcd}{}^{ef\phi} = \frac{2\varepsilon N\gamma^{-\frac{1}{2}}}{(d-1)^2}s\frac{U_1}{U}\left[\frac{d-1}{4sU} + \frac{s_1}{2s}\frac{U_1}{U} - \left(\frac{U_1}{U}\right)^2 + \frac{U_2}{U}\right]\delta^{ef}_{[[ab}\gamma_{cd]]}, \tag{A.13}$$

$$\mathcal{R}_{ab}{}^{\phi efgh} = \frac{\varepsilon N\gamma^{-\frac{1}{2}}}{4(d-1)}s\frac{U_1}{U}\delta^{[[ef}_{ab}\gamma^{gh]]}, \tag{A.14}$$

$$\begin{aligned}
\mathcal{R}_{ab}{}^{\phi ef}{}_{\phi} = &-\frac{\varepsilon N\gamma^{-\frac{1}{2}}}{2(d-1)}s\left[\frac{d}{4sU} + \frac{s_1}{2s}\frac{U_1}{U} - \frac{2d-1}{2(d-1)}\left(\frac{U_1}{U}\right)^2 + \frac{U_2}{U}\right]\delta^{ef}_{ab} \\
&+\frac{\varepsilon N\gamma^{-\frac{1}{2}}}{4(d-1)}s\left[\frac{1}{2sU} - \frac{1}{d-1}\left(\frac{U_1}{U}\right)^2\right]\gamma_{ab}\gamma^{ef}.
\end{aligned} \tag{A.15}$$

Note that, in the above expressions, a compact notation for the (anti)symmetrisation of a symmetric index pair was introduced:

$$A_{((ab}B_{cd))} \equiv \tfrac{1}{2}\left(A_{ab}B_{cd} + A_{cd}B_{ab}\right), \quad A_{[[ab}B_{cd]]} \equiv \tfrac{1}{2}\left(A_{ab}B_{cd} - A_{cd}B_{ab}\right).$$

A.4 RICCI TENSOR CONFIGURATION SPACE

The configuration Ricci tensor can then be found via contraction:

$$\mathcal{R}^{A}{}_{B} = \mathcal{R}^{CA}{}_{CB}. \tag{A.16}$$

The explicit components are straightforwardly found to be

$$\begin{aligned}
\mathcal{R}_{ab}{}^{cd} = &-\frac{\varepsilon N\gamma^{-\frac{1}{2}}s}{2(d-1)}\left[\frac{d(d^2-1)}{8sU} + \frac{s_1}{2s}\frac{U_1}{U} - \frac{(d+4)}{4}\left(\frac{U_1}{U}\right)^2 + \frac{U_2}{U}\right]\delta^{cd}_{ab} \\
&+ \frac{\varepsilon N\gamma^{-\frac{1}{2}}s}{8}\left[\frac{d-6}{2sU} - \frac{d+2}{d-1}\left(\frac{U_1}{U}\right)^2\right]\gamma_{ab}\gamma^{cd}, \\
\mathcal{R}_{ab\phi} = &\frac{\varepsilon(d+2)N\gamma^{-\frac{1}{2}}s}{2(d-1)}\frac{U_1}{U}\left[\frac{d-1}{4sU} + \frac{s_1}{2s}\frac{U_1}{U} - \left(\frac{U_1}{U}\right)^2 + \frac{U_2}{U}\right]\gamma_{ab}, \\
\mathcal{R}^{\phi ab} = &\frac{\varepsilon(d+2)N\gamma^{-\frac{1}{2}}s}{16}\frac{U_1}{U}\gamma^{ab}, \\
\mathcal{R}^{\phi}{}_{\phi} = &-\frac{\varepsilon dN\gamma^{-\frac{1}{2}}s}{4}\left[\frac{d+2}{4sU} + \frac{d+1}{2(d-1)}\frac{s_1}{s}\frac{U_1}{U}\right. \\
&\left. -\frac{2d+3}{2(d-1)}\left(\frac{U_1}{U}\right)^2 + \frac{d+1}{d-1}\frac{U_2}{U}\right].
\end{aligned}$$

A.5 RICCI SCALAR CONFIGURATION SPACE

The configuration space Ricci scalar is obtained by tracing (A.16). The result is

$$\begin{aligned}
\mathcal{R} = &-\varepsilon\frac{d(d+2)(d^2-7d+8)}{32}\frac{N\gamma^{-\frac{1}{2}}}{U} \\
&-\varepsilon\frac{d(d+1)}{4(d-1)}N\gamma^{-\frac{1}{2}}s\left[\frac{s_1}{s}\frac{U_1}{U} - \frac{(d+6)}{4}\left(\frac{U_1}{U}\right)^2 + 2\frac{U_2}{U}\right].
\end{aligned} \tag{A.17}$$

The purely gravitational contribution to the Ricci scalar with $U = U_0$ is given by

$$\mathcal{R}_{\text{grav}} = -\varepsilon N\gamma^{-\frac{1}{2}}\frac{d(d+2)(d^2-7d+4)}{32\,U_0}.$$

For $\varepsilon = -1$, $d = 3$, $N = 1$ and $U_0 = \frac{1}{2}$, this result for $\mathcal{R}_{\text{grav}}$ can be compared with the expression obtained in reference [94]. The result obtained here reduces to

$$\mathcal{R}_{\text{grav}} = -\gamma^{-\frac{1}{2}} \frac{15}{4}. \tag{A.18}$$

This disagrees with reference [94], where $\mathcal{R}_{\text{grav}}$ was found to be three times the value of **(A.18)**.

QUANTUM GRAVITATIONAL CORRECTIONS IN THE SLOW-ROLL APPROXIMATION

B

THE QUANTUM-GRAVITATIONAL CORRECTIONS to the inflationary power spectra were truncated in the slow-roll expansion at the contribution of a perfect De Sitter space in **chapter 7**. The subleading quantum-gravitational corrections at linear order in the slow-roll parameters are presented in this appendix. There are two reasons to include this analysis: first, the universal character of the dominant De Sitter contribution to the quantum-gravitational corrections affects both scalar and tensor modes in the same way. It is interesting to see whether this degeneracy is lifted after the inclusion of the slow-roll contributions. It may, in addition, allow the impact of the quantum-gravitational corrections to the tensor-to-scalar ratio to be investigated. The tensor-to-scalar ratio remains unaffected by the dominant universal De Sitter contribution. Second, the slow-roll contributions to the quantum-gravitational corrections also carry information about the generalised potentials U, G and V, which characterise the general scalar-tensor theory **(7.1)**. Therefore, these contributions allow one to determine the dependence of the quantum-gravitational corrections $\delta^{\mathrm{QG}}(k/k_0)$ on the parameters of the theory. Of course, the uncorrected part of the power spectrum is sensitive to the generalised potentials via the dependence on the generalised slow-roll parameters.

While the following analysis is based on the assumption that the relative quantum-gravitational corrections $\delta\Omega/\Omega^{(1)}$ are small and that the slow-roll approximation $|\varepsilon_i \ll 1|$ is valid, it is nevertheless useful to keep mixed terms of the form $\delta\Omega\varepsilon_i$, while neglecting terms like $\delta\Omega^2$ and ε_i^2. A complete treatment would require the higher order slow-roll contributions to the uncorrected power spectra to be included up to the order where they compete with the quantum-gravitational slow-roll corrections. These terms are neglected here. Under these assumptions, one has to solve the linearised equation **(7.99)** with the terms linear in the slow-roll parameters included:

$$\frac{\mathrm{d}}{\mathrm{d}x}\delta\Omega = 2\mathrm{i}(\Omega_{\mathrm{DS}} + \mathcal{E}\Omega_{\mathcal{E}})\delta\Omega - \mathrm{i}\frac{W}{144k^2}(\omega_0^2 + \omega_1^2\varepsilon_1 + \omega_3^2\varepsilon_3 + \mathcal{E}\omega_{\mathcal{E}}^2). \tag{B.1}$$

The functions ω_i parametrise the different slow-roll contributions to the quantum-gravitational corrections, and are defined by

The bar denotes complex conjugation.

$$\omega_0^2(x) \equiv \frac{144k^2}{W}(\omega_{\mathrm{DS}}^{\mathrm{QG}})^2 = x^4\frac{x^2-11}{(1+x^2)^3},$$

$$\omega_1^2(x) \equiv -\tfrac{2}{3}x^4\frac{11x^2-49}{(1+x^2)^3},$$

$$\omega_3^2(x) \equiv -8x^4\frac{x^2-5}{(1+x^2)^3},$$

$$\omega_{\mathcal{E}}^2(x) \equiv -\frac{x^4}{(1+x^2)^4}\left[P(x) + Q(x) + \overline{Q}(x)\right],$$

with the polynomials

$$P(x) = 7x^6 - 21x^4 + 89x^2 - 27,$$
$$Q(x) = e^{2ix}(i + x)^4 \left[6x^4 - 34x^2 + 11 - i(20x^3 - 22x)\right] [i\pi + \mathrm{Ei}(-2ix)] .$$

One can make a similar ansatz for the perturbation of the Gaussian width:

$$\delta\Omega = \frac{W}{144k^2} (\delta\Omega_0 + \varepsilon_1\delta\Omega_1 + \varepsilon_3\delta\Omega_3 + \mathcal{E}\delta\Omega_{\mathcal{E}}) . \tag{B.2}$$

Equation **(B.1)** can then be written as a system of linear equations:

$$\frac{d\delta\Omega_i}{dx}(x) = M_i^{\,j}(x)\delta\Omega_j(x) - X_i(x), \tag{B.3}$$

with the matrix M and the vector X defined as

$$M(x) \equiv 2i\begin{pmatrix} \Omega_{DS} & 0 & 0 & 0 \\ i/x & \Omega_{DS} & 0 & 0 \\ i/x & 0 & \Omega_{DS} & 0 \\ \Omega_{\mathcal{E}} & 0 & 0 & \Omega_{DS} \end{pmatrix}, \qquad X \equiv i\begin{pmatrix} \omega_0^2 \\ \omega_1^2 \\ \omega_3^2 \\ \omega_{\mathcal{E}}^2 \end{pmatrix}. \tag{B.4}$$

The natural choice for the initial conditions is the asymptotic Bunch-Davis boundary condition. Up to linear order in the slow-roll parameters **(7.96)** reduces to

$$(\Omega_\infty^{(2)})^2 = k^2 + \frac{W}{144k}\left[1 - \tfrac{22}{3}\varepsilon_1 - 8\varepsilon_3 + \tfrac{3}{2}\mathcal{E}\right].$$

in the limit $x \to \infty$. Equations **(7.82)** and **(B.2)** together imply the asymptotic values

$$\delta\Omega_i^\infty = \left(\tfrac{1}{2}, -\tfrac{11}{3}, -4, \tfrac{3}{4}\right) . \tag{B.5}$$

The solution to **(B.3)** that satisfies this asymptotic condition can then formally be written as

$$\delta\Omega_i(x) = \mathfrak{M}_i^{\,j}(x)\left\{\delta\Omega_j^\infty - \int_\infty^x \left[\mathfrak{M}^{-1}\right]_j^{\,k}(z)X_k(z)\,dz\right\}, \tag{B.6}$$
$$\mathfrak{M}(x) \equiv \left[\exp\left(\int_\infty^x M(y)\,dy\right)\right]_{ij},$$

The only solution that can be explicitly calculated is $\delta\Omega_0$. The non-diagonal elements of the matrix exponential prevent an analytic calculation for the remaining functions. For the numerical evaluation of **(B.3)** a finite cutoff x_0 for the lower integration bound has to be chosen. For sufficiently large values of x_0, the final results for the numerical solutions $\delta\Omega_i^N$ do not depend on this choice as they quickly asymptote their constant values **(B.5)**. The value has been fixed at $x_0 = 10^6$ in order to compare results derived here with those of [110]. The real parts of the $\delta\Omega_i^N$ are plotted in **figure B.1**. Recall that it is the real part of the width that appears in the power spectra.

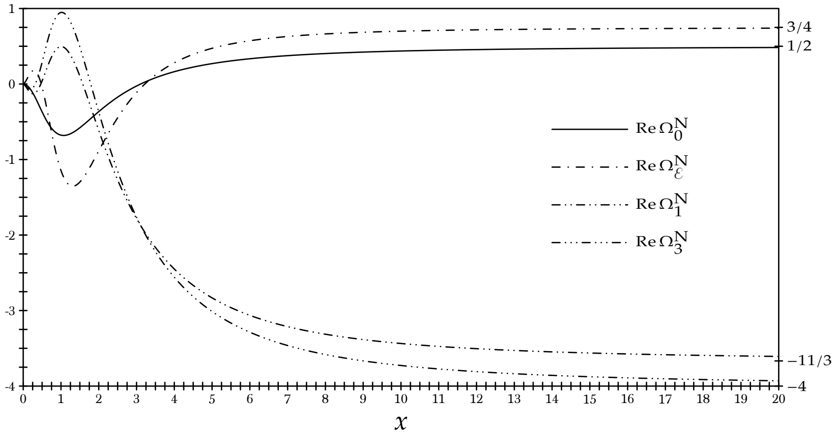

Figure B.1 Numerical solutions to (B.3).

The inflationary power spectra are obtained in the superhorizon limit $x \ll 1$. This limit can be obtained directly from the numerical solutions $\delta\Omega_i^{\mathrm{N}}$, but it is convenient to use a hybrid analytic-numerical approach to extract the analytic x-dependence of the power spectra in the superhorizon limit. In this regime the system of linear equations (B.3) reduces to the simple set of equations

$$\frac{\mathrm{d}\delta\Omega_0^{\mathrm{SH}}}{\mathrm{d}x} = 2x^{-1}\delta\Omega_0^{\mathrm{SH}}, \qquad \frac{\mathrm{d}\delta\Omega_{\mathcal{E}}^{\mathrm{SH}}}{\mathrm{d}x} = 2x^{-1}\left(\delta\Omega_{\mathcal{E}}^{\mathrm{SH}} + \delta\Omega_0^{\mathrm{SH}}\right),$$
$$\frac{\mathrm{d}\delta\Omega_1^{\mathrm{SH}}}{\mathrm{d}x} = 2x^{-1}\left(\delta\Omega_1^{\mathrm{SH}} - \delta\Omega_0^{\mathrm{SH}}\right), \qquad \frac{\mathrm{d}\delta\Omega_3^{\mathrm{SH}}}{\mathrm{d}x} = 2x^{-1}\left(\delta\Omega_3^{\mathrm{SH}} - \delta\Omega_0^{\mathrm{SH}}\right),$$

with the real part of the solutions determined up to integration constants β_i:

$$\mathrm{Re}\left(\delta\Omega_0^{\mathrm{SH}}\right) = \beta_0\, x^2, \qquad \mathrm{Re}\left(\delta\Omega_{\mathcal{E}}^{\mathrm{SH}}\right) = (\beta_{\mathcal{E}} + 2\beta_0 \log x)x^2,$$
$$\mathrm{Re}\left(\delta\Omega_1^{\mathrm{SH}}\right) = (\beta_1 - 2\beta_0 \log x)x^2, \qquad \mathrm{Re}\left(\delta\Omega_3^{\mathrm{SH}}\right) = (\beta_3 - 2\beta_0 \log x)x^2. \tag{B.7}$$

The superhorizon solutions $\delta\Omega_i^{\mathrm{SH}}$ in (B.7) are obtained from (B.3) in the limit $x \ll 1$. Therefore, the integration constants β_0 and β_i, $i = 1, 3, \mathcal{E}$ cannot be determined by the asymptotic Bunch-Davies boundary conditions (B.5), as they are imposed at $x \gg 1$. Instead, they have to be determined by fitting the analytic superhorizon solutions (B.7) to the numerical solutions $\delta\Omega_i^{\mathrm{N}}$ at $x \ll 1$:

$$\beta_0 \approx -1.98, \qquad \beta_1 \approx 3.30, \qquad \beta_3 \approx 4.62, \qquad \beta_{\mathcal{E}} \approx -2.24. \tag{B.8}$$

Here β_0 consistently reproduces the constant in the analysis of the dominant De Sitter contribution in (7.106). The $\delta\Omega_i^{\mathrm{SH}}$ are compared to their numerical counterpart in figure B.2. Using the analytical solutions (B.7) with the fits (B.8) for the coefficients β_i, the power spectra including the slow-roll contributions to the quantum-gravitational corrections are compactly written as

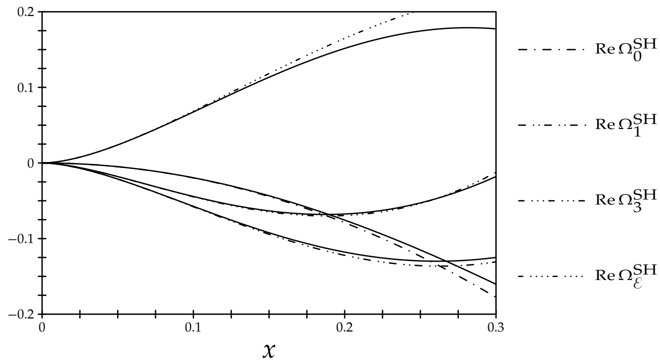

Figure B.2 Comparison of the numerical solution (solid lines) **(B.3)** to the approximate analytical solution (dotted lines) **(B.7)**.

$$P^{(2)}_{\mathrm{S/T}} \approx P^{(1)}_{\mathrm{S/T}} \left(1 + \delta^{\mathrm{QG}}_{\mathrm{S/T}}\right). \tag{B.9}$$

The quantum-gravitational corrections $\delta^{\mathrm{QG}}_{\mathrm{S/T}}$ for the scalar and tensor perturbations which include the slow-roll contributions are

$$\delta^{\mathrm{QG}}_{\mathrm{S/T}} \approx \delta^{\mathrm{QG}}_{\mathrm{DS}} \Big[1 + \beta_0^{-1}\beta_1\varepsilon_1 + \beta_0^{-1}\beta_3\varepsilon_3 + \left(\beta_0^{-1}\beta_{\mathcal{E}} + 2c_\gamma\right)\mathcal{E}_{\mathrm{S/T}} - 2(\varepsilon_1 + \varepsilon_3)\log x\Big], \tag{B.10}$$

where the correction $\delta^{\mathrm{QG}}_{\mathrm{DS}}$ is given by

$$\delta^{\mathrm{QG}}_{\mathrm{DS}}(k/k_0) = -\frac{\beta_0 W}{144}\left(\frac{k_0}{k}\right)^3 \approx \frac{W}{72}\left(\frac{k_0}{k}\right)^3 .$$

From **(B.10)** it can be inferred that, in contrast to the dominant universal De Sitter contribution of the quantum-gravitational corrections, the subleading slow-roll contributions are different for scalar and tensor perturbations. The spectral observables can be obtained straightforwardly by the insertion of the corrected power spectra **(B.9)** into **(7.70)**. In addition it is found that the tensor-to-scalar ratio **(7.71)** is no longer unaffected by quantum-gravitational corrections.

The results obtained here can be compared to the analysis performed in [110] for a single minimally coupled scalar field with a canonically normalised kinetic term and an arbitrary scalar potential $\mathcal{V}$. It is found that the dominant De Sitter contribution to the quantum-gravitational corrections agree with the above results for the identifications $\mathcal{V} = M_{\mathrm{P}}^4 W/4$ and $U = \frac{1}{2}M_{\mathrm{P}}^2$. However, differences can be found in the subleading slow-roll contributions to the quantum-gravitational corrections considered in [110]. First, note that here the final result for the corrected power spectra is parametrised in terms of W. The results in [110] were expressed in terms of H^2/M_{P}^2. The conversion from H^2/M_{P}^2 to W induces additional terms linear in the slow-roll parameters, as can be seen in **(7.93)**. Second, there is a true difference between both results. This difference originates from the treatment of the ansatz **(B.2)** in the differential equation **(B.1)**. In deriving **(B.3)** derivatives of W lead to the off-diagonal terms i/x

Furthermore, note the difference in notation $\mathcal{V} \to V$.

in the matrix (B.4), while the approach of [110] was to expand W around horizon crossing W_* in (B.1) and to identify the additional terms linear in slow-roll to the source term in (B.1). Consequently, one assumes a constant value W_* in the ansatz (B.2). Both procedures ultimately lead to different results. In particular, the implementation of boundary data becomes more complicated in the procedure followed [110], as the corrections to the Gaussian width $\delta\Omega_i$ do not asymptote to constant values in the early time limit—in contrast to the approach presented here, where this arises naturally (as seen in (B.5)).

THE PRIMORDIAL POWER SPECTRA can be calculated in the Heisenberg picture, as in **chapter 4**, or in the Schrödinger picture, as in **chapter 7**. The question of which picture is used in the calculated of the power spectra is purely one of convenience. The results should be independent of such a choice. This was shown in the main text. A brief proof will be given in this appendix that both approaches can be smoothly transformed into each other, and shows that the identification of **(7.82)** with the Bunch-Davies vacuum is justified.

In the Schrödinger picture the state is described by the wavefunction ψ, which satisfies the Schrödinger equation

$$\mathrm{i}\frac{\partial\psi}{\partial t_\mathrm{s}} = \mathcal{H}\psi, \tag{C.1}$$

where the Hamiltonian $\mathcal{H}$ is given by **(7.45)**. In order to illustrate the identification of **(7.82)** with the Bunch-Davies boundary condition the quantum-gravitational corrections are neglected here, as they are absent in cosmological perturbation theory.

Using the decomposition **(7.73)** one obtains the differential equation for the Gaussian width Ω:

$$\mathrm{i}\frac{\partial\Omega}{\partial t_\mathrm{s}} = \Omega^2 - \omega^2.$$

The boundary condition imposed on ψ is that it describes the ground state of a harmonic oscillator in the very early universe $-kt_\mathrm{s} \to \infty$. The differential equation **(C.1)** is first order in time. Defining an auxiliary variable μ through

$$\Omega \equiv -\mathrm{i}\partial_{t_\mathrm{s}} \log\mu, \tag{C.2}$$

one obtains the second order differential equation

$$\partial_{t_\mathrm{s}}^2\mu + \omega^2\mu = 0. \tag{C.3}$$

It can straightforwardly be seen that this is **(4.39)**. Substitution of **(C.2)** into **(7.74)** therefore yields the MS expression for the power spectrum:

$$P(k) = \frac{k^3}{2\pi^2}\frac{|\mu|^2}{W(k)}.$$

The function $W(k)$ is the Wronskian of the auxiliary variable μ:

$$W(k) = \mathrm{i}(\mu\partial_{t_\mathrm{s}}\mu^* - \mu^*\partial_{t_\mathrm{s}}\mu).$$

It follows from **(C.3)** that $W(k)$ is a constant in time. Notice that it is always possible for $W(k)$ to be normalised to be unity, since **(C.3)** is a linear second order differential equation. It is therefore assumed from this point on that this choice has been made. The asymptotic behaviour for $|\mu|$ can be obtained by considering (C.2) in the asymptotic limit:

This condition is in fact required by the canonical commutation relation in the Hamiltonian picture.

$$\lim_{-kt_\mathrm{s}\to\infty} |\mu|^2 = \tfrac{1}{2}k^{-1}.$$

The asymptotic behaviour for μ can be found by integrating (C.2). The result is

$$\lim_{-kt_\mathrm{s}\to\infty} \mu = (2k)^{-1/2}\exp(\mathrm{i}\,kt_\mathrm{s}),$$

which, up to an unphysical sign, is precisely (4.40). This justifies the identification of the initial condition (7.82) for Ω with the Bunch-Davies vacuum.

ACKNOWLEDGEMENTS

Progress is made accumulatively. That means that human knowledge is expanded incrementally, by directly building on past accomplishments. It is because of this interrelation between past and present work that, for a given result, an innumerable amount of people have worked to bring that result about. It therefore stands to reason that the number of people involved in the making of this dissertation goes beyond counting. However, it is not too difficult to find a limited set of these people, who have had a sizable impact on me as I progressed during my research. I would like to thank them here.

First and foremost, I am indebted to Jochum van der Bij. During my time in Freiburg, he provided me with an environment which allowed me to focus on my work, and with his dry humour was more than willing to have a lively discussion. The breadth and depth of his knowledge, both within physics and without, continue to make an impression on me.

Secondly, I am grateful for the discussions I had with Harald Ita, who was my second supervisor during my time in the *Graduiertenkolleg*. Although not directly active in the field of my work, his experience and perspectives often proved helpful for me when I was stuck.

Thirdly, I would like to thank Christian Steinwachs for the opportunity to collaborate with him and to have discussions on various topics in quantum field theory, general relativity and cosmology. His comments and guidance greatly improved the quality of my work.

In a broader sense, I would like to thank the people in the *Graduiertenkolleg* (Research Training Group) 2044 for providing a well-organised platform for both theoretical and experimental physicists to collaborate and exchange ideas, and additionally the means to participate with the academic world at large. As a special thanks, I would like to warmly thank Christina Skorek for going above and beyond the call of duty during the first months of my PhD. Without her help, I definitely would be lost.

This project would not have been possible without the funding provided by the Deutsche Forschungsgemeinschaft (DFG) during the first three years of my PhD, and the *Abschlussstipendium* kindly provided by the University of Freiburg during the last year. I heartily acknowledge their support.

Additionally, I would like to thank the many PhDs and postdocs in the Dittmaier and Ita groups for providing a comfortable environment to work in. In particular I wish to extend my thanks to Samuel de Abreu, Jerry Dormans, Matthieu Jacquier, Lokesh Mishra, Ben Page, Michael Ruf, Timo Schmidt and Christopher Schwan for the interesting discussions and the regular table football and bouldering sessions.

Finally, I would like to thank my family and friends for their love and support, their eagerness to learn more about physics in general and my work in particular, and their willingness to be a sounding board for me to formulate ideas.

На крају, желео бих се захвалити мојој малој мишици која ме је, чак и кад је била заузета, непрестано подржавала.

[1] C.F. Steinwachs and M.L. van der Wild, Quantum gravitational corrections from the Wheeler–DeWitt equation for scalar–tensor theories, *Class. Quant. Grav.* **35**(13), 135010 (2018).

[2] C.F. Steinwachs and M.L. van der Wild, Quantum gravitational corrections to the inflationary power spectra in scalar-tensor theories, (2019). (Accepted for publication in *Class. Quant. Grav.*)

[3] B.P. Abbott *et al.*, Observation of Gravitational Waves from a Binary Black Hole Merger, *Phys. Rev. Lett.* **116**(6), 061102 (2016).

[4] K. Akiyama *et al.*, First M87 Event Horizon Telescope Results. V. Physical Origin of the Asymmetric Ring, *Astrophys. J.* **875**(1), L5 (2019).

[5] R.L. Arnowitt, S. Deser, and C.W. Misner, Canonical variables for general relativity, *Phys. Rev.* **117** 1595-1602 (1960).

[6] C.W. Misner, K.S. Thorne, and J.A. Wheeler, *Gravitation* (W. H. Freeman, San Francisco, 1973).

[7] K. Kuchař, Geometry of Hyperspace. 1., *J. Math. Phys.* **17** 777-791 (1976).

[8] K. Kuchař, Kinematics of Tensor Fields in Hyperspace. 2., *J. Math. Phys.* **17** 792-800 (1976).

[9] K. Kuchař, Dynamics of Tensor Fields in Hyperspace. 3., *J. Math. Phys.* **17** 801-820 (1976).

[10] L.D. Landau and E.M. Lifshitz, *The Classical Theory of Fields* (Pergamon Press, Oxford, 1975).

[11] R.M. Wald, *General Relativity* (Chicago Univ. Pr., Chicago, USA, 1984).

[12] S.M. Carroll, *Spacetime and geometry: An introduction to general relativity* (San Francisco, USA: Addison-Wesley, 2004).

[13] E. Gourgoulhon, *3+1 formalism in general relativity: bases of numerical relativity* (Springer Berlin Heidelberg, 2012).

[14] S. Weinberg, *Cosmology* (Oxford Univ. Pr., 2008).

[15] E. Poisson, *A Relativist's Toolkit: The Mathematics of Black-Hole Mechanics* (Cambridge University Press, 2009).

[16] L.W. Tu, *An Introduction to Manifolds* (Springer New York, 2010).

[17] J.W. York Jr., Role of conformal three geometry in the dynamics of gravitation, *Phys. Rev. Lett.* **28** 1082-1085 (1972).

[18] G.W. Gibbons and S.W. Hawking, Action Integrals and Partition Functions in Quantum Gravity, *Phys. Rev. D* **15** 2752-2756 (1977).

[19] E. Dyer and K. Hinterbichler, Boundary Terms, Variational Principles and Higher Derivative Modified Gravity, *Phys. Rev. D* **79** 024028 (2009).

[20] D. Lovelock, The Einstein tensor and its generalizations, *J. Math. Phys.* **12** 498-501 (1971).

[21] G.W. Horndeski, Second-Order Scalar-Tensor Field Equations in a Four Dimensional Space, *Int. J. Theor. Phys.* **10** 363-384 (1974).

[22] R.H. Dicke, Mach's principle and invariance under transformation of units, *Phys. Rev.* **125** 2163-2167 (1962).

[23] T.P. Sotiriou, $f(R)$ Gravity and scalar-tensor theory, *Class. Quant. Grav.* **23** 5117-5128 (2006).

[24] A. De Felice and S. Tsujikawa, $f(R)$ Theories, *Living Rev. Rel.* **13** 3 (2010).

[25] V. Faraoni and S. Capozziello, *Beyond Einstein Gravity*, Vol. 170 (Springer, Dordrecht, 2011).

[26] S. Nojiri and S.D. Odintsov, Unified cosmic history in modified gravity: from $f(R)$ theory to Lorentz non-invariant models, *Phys. Rept.* **505** 59-144 (2011).

[27] T. Clifton, P.G. Ferreira, A. Padilla, and C. Skordis, Modified Gravity and Cosmology, *Phys. Rept.* **513** 1-189 (2012).

[28] S. Nojiri, S.D. Odintsov, and V.K. Oikonomou, Modified Gravity Theories on a Nutshell: Inflation, Bounce and Late-time Evolution, *Phys. Rept.* **692** 1-104 (2017).

[29] H. Bondi and J. Samuel, The Lense-Thirring effect and Mach's principle, *Phys. Lett.* **A228** 121 (1997).

[30] C. Brans and R.H. Dicke, Mach's principle and a relativistic theory of gravitation, *Phys. Rev.* **124** 925-935 (1961).

[31] C.H. Brans, Mach's Principle and a Relativistic Theory of Gravitation. II, *Phys. Rev.* **125** 2194-2201 (1962).

[32] T. Damour and A.M. Polyakov, String theory and gravity, *Gen. Rel. Grav.* **26** 1171-1176 (1994).

[33] T. Damour, F. Piazza, and G. Veneziano, Runaway dilaton and equivalence principle violations, *Phys. Rev. Lett.* **89** 081601 (2002).

[34] B.L. Spokoiny, Inflation and generation of perturbations in broken symmetric theory of gravity, *Phys. Lett. B* **147** 39-43 (1984).

[35] D.S. Salopek, J.R. Bond, and J.M. Bardeen, Designing Density Fluctuation Spectra in Inflation, *Phys. Rev. D* **40** 1753 (1989).

[36] R. Fakir and W.G. Unruh, Improvement on cosmological chaotic inflation through nonminimal coupling, *Phys. Rev. D* **41** 1783-1791 (1990).

[37] A. De Simone, M.P. Hertzberg, and F. Wilczek, Running Inflation in the Standard Model, *Phys. Lett. B* **678** 1-8 (2009).

[38] A.O. Barvinsky, A. Yu. Kamenshchik, C. Kiefer, A.A. Starobinsky, and C. Steinwachs, Asymptotic freedom in inflationary cosmology with a nonminimally coupled Higgs field, *JCAP* **0912** 003 (2009).

[39] C.F. Steinwachs and A. Yu. Kamenshchik, One-loop divergences for gravity non-minimally coupled to a multiplet of scalar fields: calculation in the Jordan frame. I. The main results, *Phys. Rev. D* **84** 024026 (2011).

[40] E.E. Flanagan, The Conformal frame freedom in theories of gravitation, *Class. Quant. Grav.* **21** 3817 (2004).

[41] A.O. Barvinsky, A. Yu. Kamenshchik, and A.A. Starobinsky, Inflation scenario via the Standard Model Higgs boson and LHC, *JCAP* **11** 021 (2008).

[42] F.L. Bezrukov and D.S. Gorbunov, Distinguishing between R^2-inflation and Higgs-inflation, *Phys. Lett. B* **713** 365-368 (2012).

[43] A.O. Barvinsky, A. Yu. Kamenshchik, C. Kiefer, A.A. Starobinsky, and C.F. Steinwachs, Higgs boson, renormalization group, and naturalness in cosmology, *Eur. Phys. J.* **C72** 2219 (2012).

[44] A. Kehagias, A. Moradinezhad Dizgah, and A. Riotto, Remarks on the Starobinsky model of inflation and its descendants, *Phys. Rev. D* **89**(4), 043527 (2014).

[45] V. Mukhanov, Quantum Cosmological Perturbations: Predictions and Observations, *Eur. Phys. J. C* **73** 2486 (2013).

[46] R. Kallosh, A. Linde, and D. Roest, Universal Attractor for Inflation at Strong Coupling, *Phys. Rev. Lett.* **112**(1), 011303 (2014).

[47] F.L. Bezrukov and M. Shaposhnikov, The Standard Model Higgs boson as the inflaton, *Phys. Lett. B* **659** 703-706 (2008).

[48] F.L. Bezrukov, A. Magnin, and M. Shaposhnikov, Standard Model Higgs boson mass from inflation, *Phys. Lett. B* **675** 88-92 (2009).

[49] F. Bezrukov, A. Magnin, M. Shaposhnikov, and S. Sibiryakov, Higgs inflation: consistency and generalisations, *JHEP* **01** 016 (2011).

[50] F. Bezrukov and M. Shaposhnikov, Standard Model Higgs boson mass from inflation: Two loop analysis, *JHEP* **07** 089 (2009).

[51] K. Allison, Higgs ξ-inflation for the 125-126 GeV Higgs: a two-loop analysis, *JHEP* **02** 040 (2014).

[52] M.S. Ruf and C.F. Steinwachs, Quantum equivalence of $f(R)$ gravity and scalar-tensor theories, *Phys. Rev. D* **97**(4), 044050 (2018).

[53] H.A. Buchdahl, Non-linear Lagrangians and cosmological theory, *Mon. Not. R. Astron. Soc.* **150**(1), 1–8 (1970).

[54] A.A. Starobinsky, A New Type of Isotropic Cosmological Models Without Singularity, *Phys. Lett. B* **91** 99-102 (1980).

[55] K.S. Stelle, Classical gravity with higher derivatives, *Gen. Relativ. Gravit.* **9** 353–371 (1978).

[56] S. Dodelson, *Modern Cosmology* (Academic Press, Amsterdam, 2003).

[57] V. Mukhanov, *Physical Foundations of Cosmology* (Cambridge University Press, Oxford, 2005).

[58] Baumann, Daniel, Inflation, In Physics of the large and the small, TASI 09, proceedings of the Theoretical Advanced Study Institute in Elementary Particle Physics, Boulder, Colorado, USA, 1-26 June 2009 (2011).

[59] D.H. Lyth and A.R. Liddle, *The primordial density perturbation: Cosmology, inflation and the origin of structure* (2009).

[60] P. Peter and J.P. Uzan, *Primordial Cosmology* (Oxford University Press, 2013).

[61] A.G. Riess *et al.*, Observational evidence from supernovae for an accelerating universe and a cosmological constant, *Astron. J.* **116** 1009-1038 (1998).

[62] S. Perlmutter *et al.*, Measurements of Omega and Lambda from 42 high redshift supernovae, *Astrophys. J.* **517** 565-586 (1999).

[63] A.H. Guth, The Inflationary Universe: A Possible Solution to the Horizon and Flatness Problems, *Phys. Rev. D* **23** 347-356 (1981).

[64] K. Sato, First Order Phase Transition of a Vacuum and Expansion of the Universe, *Mon. Not. Roy. Astron. Soc.* **195** 467-479 (1981).

[65] K. Sato, Cosmological Baryon Number Domain Structure and the First Order Phase Transition of a Vacuum, *Phys. Lett. B* **99** 66-70 (1981).

[66] A.D. Linde, A New Inflationary Universe Scenario: A Possible Solution of the Horizon, Flatness, Homogeneity, Isotropy and Primordial Monopole Problems, *Phys. Lett. B* **108** 389-393 (1982).

[67] A. Albrecht and P.J. Steinhardt, Cosmology for Grand Unified Theories with Radiatively Induced Symmetry Breaking, *Phys. Rev. Lett.* **48** 1220-1223 (1982).

[68] A.D. Linde, Chaotic Inflation, *Phys. Lett. B* **129** 177-181 (1983).

[69] A.A. Starobinsky, Spectrum of relict gravitational radiation and the early state of the universe, *JETP Lett.* **30** 682-685 (1979).

[70] V.F. Mukhanov and G.V. Chibisov, Quantum Fluctuations and a Nonsingular Universe, *JETP Lett.* **33** 532-535 (1981).

[71] A.A. Starobinsky, Dynamics of Phase Transition in the New Inflationary Universe Scenario and Generation of Perturbations, *Phys. Lett. B* **117** 175-178 (1982).

[72] V.F. Mukhanov and G.V. Chibisov, The vacuum energy and large scale structure of the universe, *Sov. Phys. JETP* **56** 258-265 (1982).

[73] A.H. Guth and S.Y. Pi, Fluctuations in the New Inflationary Universe, *Phys. Rev. Lett.* **49** 1110-1113 (1982).

[74] S.W. Hawking, The Development of Irregularities in a Single Bubble Inflationary Universe, *Phys. Lett. B* **115** 295 (1982).

[75] J.M. Bardeen, P.J. Steinhardt, and M.S. Turner, Spontaneous Creation of Almost Scale - Free Density Perturbations in an Inflationary Universe, *Phys. Rev. D* **28** 679 (1983).

[76] J. Martin, C. Ringeval, and V. Vennin, Encyclopædia Inflationaris, *Phys. Dark Univ.* **5-6** 75-235 (2014).

[77] D.F. Torres, Slow roll inflation in nonminimally coupled theories: Hyperextended gravity approach, *Phys. Lett. A* **225** 13-17 (1997).

[78] A.R. Liddle, P. Parsons, and J.D. Barrow, Formalizing the slow roll approximation in inflation, *Phys. Rev. D* **50** 7222-7232 (1994).

[79] D. Burns, S. Karamitsos, and A. Pilaftsis, Frame-Covariant Formulation of Inflation in Scalar-Curvature Theories, *Nucl. Phys. B* **907** 785-819 (2016).

[80] K.A. Malik and D.R. Matravers, A Concise Introduction to Perturbation Theory in Cosmology, *Class. Quant. Grav.* **25** 193001 (2008).

[81] J. Weenink and T. Prokopec, Gauge invariant cosmological perturbations for the nonminimally coupled inflaton field, *Phys. Rev. D* **82** 123510 (2010).

[82] 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** 117-134 (1978).

[83] M. Abramowitz and I.A. Stegun, *Handbook of Mathematical Functions: With Formulas, Graphs, and Mathematical Tables* (U.S. Department of Commerce, National Bureau of Standards, 1972).

[84] Y. Akrami *et al.*, Planck 2018 results. X. Constraints on inflation, (2018). (arXiv: 1807.06211 [astro-ph.CO])

[85] C. Møller, The energy-momentum complex in general relativity and related problems, *Colloq. Int. CNRS* **91** 15-29 (1962).

[86] K. Eppley and E. Hannah, The necessity of quantizing the gravitational field, *Foundations of Physics* **7**(1), 51–68 (1977).

[87] J. Mattingly, Why Eppley and Hannah's thought experiment fails, *Phys. Rev.* **D73** 064025 (2006).

[88] M. Albers, C. Kiefer, and M. Reginatto, Measurement Analysis and Quantum Gravity, *Phys. Rev.* **D78** 064051 (2008).

[89] A. Kent, Simple refutation of the Eppley–Hannah argument, *Class. Quant. Grav.* **35**(24), 245008 (2018).

[90] S.W. Hawking and R. Penrose, The Singularities of gravitational collapse and cosmology, *Proc. Roy. Soc. Lond.* **A314** 529-548 (1970).

[91] S.W. Hawking, Particle Creation by Black Holes, *Commun. Math. Phys.* **43** 199-220 (1975). ([,167(1975)])

[92] C. Kiefer, *Quantum Gravity* (Oxford University Press, New York, 2007).

[93] T. Banks, TCP, Quantum Gravity, the Cosmological Constant and All That..., *Nucl. Phys. B* **249** 332-360 (1985).

[94] B.S. DeWitt, Quantum Theory of Gravity. 1. The Canonical Theory, *Phys. Rev.* **160** 1113-1148 (1967).

[95] C. Kiefer and T.P. Singh, Quantum gravitational corrections to the functional Schrödinger equation, *Phys. Rev. D* **44** 1067-1076 (1991).

[96] P.A.M. Dirac, *Lectures on Quantum Mechanics* (Dover Publications, 1964).

[97] A.O. Barvinsky, Unitarity approach to quantum cosmology, *Phys. Rept.* **230** 237-367 (1993).

[98] J.J. Halliwell, Introductory lectures on quantum cosmology, In 7th Jerusalem Winter School for Theoretical Physics: Quantum Cosmology and Baby Universes Jerusalem, Israel, 27 December 1989 - 4 January 1990 (1990).

[99] K. Kuchař, Canonical Quantization of Gravity, In Relativity, Astrophysics and Cosmology (Springer Netherlands, Dordrecht, 1973).

[100] J.A. Wheeler, On the Nature of quantum geometrodynamics, *Annals Phys.* **2** 604-614 (1957).

[101] C.M. DeWitt and J.A. Wheeler, *Battelle rencontres - 1967 lectures in mathematics and physics (seattle)* (W. A. Benjamin, 1969).

[102] T. Thiemann, *Modern Canonical Quantum General Relativity* (Cambridge University Press, 2007).

[103] P.W. Higgs, Integration of Secondary Constraints in Quantized General Relativity, *Phys. Rev. Lett.* **1** 373-374 (1958). (Erratum: Phys. Rev. Lett.3,66(1959))

[104] J.B. Hartle and S.W. Hawking, Wave Function of the Universe, *Phys. Rev. D* **28** 2960-2975 (1983).

[105] A. Vilenkin, Quantum Creation of Universes, *Phys. Rev. D* **30** 509-511 (1984).

[106] A.D. Linde, Quantum Creation of the Inflationary Universe, *Lett. Nuovo Cim.* **39** 401-405 (1984).

[107] V. Moncrief and C. Teitelboim, Momentum constraints as integrability conditions for the hamiltonian constraint in general relativity, *Phys. Rev. D* **6** 966-969 (1972).

[108] C. Kiefer, Continuous Measurement of Minisuperspace Variables by Higher Multipoles, *Class. Quant. Grav.* **4** 1369 (1987).

[109] D. Brizuela, C. Kiefer, and M. Krämer, Quantum-gravitational effects on gauge-invariant scalar and tensor perturbations during inflation: The de Sitter case, *Phys. Rev.* **D93**(10), 104035 (2016).

[110] D. Brizuela, C. Kiefer, and M. Krämer, Quantum-gravitational effects on gauge-invariant scalar and tensor perturbations during inflation: The slow-roll approximation, *Phys. Rev. D* **94**(12), 123527 (2016).

[111] A. Vilenkin, The Interpretation of the Wave Function of the Universe, *Phys. Rev.* **D39** 1116 (1989).

[112] B.S. DeWitt, Point transformations in quantum mechanics, *Phys. Rev.* **85** 653-661 (1952).

[113] A.O. Barvinsky and A. Yu. Kamenshchik, Selection rules for the Wheeler-DeWitt equation in quantum cosmology, *Phys. Rev. D* **89**(4), 043526 (2014).

[114] S.W. Hawking, The Quantum State of the Universe, *Nucl. Phys. B* **239** 257 (1984).

[115] V.A. Rubakov, Particle creation in a tunneling universe, *JETP Lett.* **39** 107-110 (1984).

[116] Ya. B. Zeldovich and A.A. Starobinsky, Quantum creation of a universe in a nontrivial topology, *Sov. Astron. Lett.* **10** 135 (1984).

[117] A. Vilenkin, Quantum Cosmology and the Initial State of the Universe, *Phys. Rev. D* **37** 888 (1988).

[118] A.O. Barvinsky and A. Yu. Kamenshchik, Quantum scale of inflation and particle physics of the early universe, *Phys. Lett. B* **332** 270-276 (1994).

[119] A. Vilenkin, The Quantum cosmology debate, *AIP Conf. Proc.* **478**(1), 23-29 (1999).

[120] A.O. Barvinsky, A. Yu. Kamenshchik, C. Kiefer, and C.F. Steinwachs, Tunneling cosmological state revisited: Origin of inflation with a nonminimally coupled Standard Model Higgs inflaton, *Phys. Rev. D* **81** 043530 (2010).

[121] G. Calcagni, C. Kiefer, and C.F. Steinwachs, Quantum cosmological consistency condition for inflation, *JCAP* **1410**(10), 026 (2014).

[122] U.H. Gerlach, Derivation of the ten Einstein field equations from the semiclassical approximation to quantum geometrodynamics, *Phys. Rev.* **177** 1929-1941 (1969).

[123] V.G. Lapchinsky and V.A. Rubakov, Canonical quantization of gravity and quantum field theory in curved space-time, *Acta Phys. Polon. B* **10** 1041-1048 (1979).

[124] R. Brout, On the Concept of Time and the Origin of the Cosmological Temperature, *Found. Phys.* **17** 603 (1987).

[125] R. Brout and G. Venturi, Time in Semiclassical Gravity, *Phys. Rev. D* **39** 2436 (1989).

[126] C. Bertoni, F. Finelli, and G. Venturi, The Born-Oppenheimer approach to the matter - gravity system and unitarity, *Class. Quant. Grav.* **13** 2375-2384 (1996).

[127] A. Yu. Kamenshchik, A. Tronconi, and G. Venturi, The Born-Oppenheimer Method, Quantum Gravity and Matter, *Class. Quant. Grav.* **35**(1), 015012 (2018).

[128] M. Born and J.R. Oppenheimer, Zur Quantentheorie der Molekeln, *Annalen der Physik.* **389** **(20)** 457–484 (1927).

[129] A. Peres, On Cauchy's problem in general relativity - II, *Nuovo Cimento* **XXVI** 53-62 (1962).

[130] C. Kiefer and D. Wichmann, Semiclassical approximation of the Wheeler-DeWitt equation - arbitrary orders and the question of unitarity, *Gen. Rel. Grav.* **50**(6), 66 (2018).

[131] C. Kiefer, Nonminimally Coupled Scalar Fields and the Initial Value Problem in Quantum Gravity, *Phys. Lett. B* **225** 227-232 (1989).

[132] A.O. Barvinsky, A. Yu. Kamenshchik, and I.P. Karmazin, The Renormalization group for nonrenormalizable theories: Einstein gravity with a scalar field, *Phys. Rev. D* **48** 3677-3694 (1993).

[133] I.L. Shapiro and H. Takata, One loop renormalization of the four-dimensional theory for quantum dilaton gravity, *Phys. Rev. D* **52** 2162-2175 (1995).

[134] A. Yu. Kamenshchik and C.F. Steinwachs, Question of quantum equivalence between Jordan frame and Einstein frame, *Phys. Rev. D* **91**(8), 084033 (2015).

[135] M.S. Ruf and C.F. Steinwachs, One-loop divergences for $f(R)$ gravity, *Phys. Rev. D* **97**(4), 044049 (2018).

[136] N. Ohta, Quantum Equivalence of $f(R)$ Gravity and Scalar-tensor Theories in the Jordan and Einstein Frames, *PTEP* **2018**(3), 033B02 (2018).

[137] C. Kiefer and M. Kraemer, Quantum Gravitational Contributions to the CMB Anisotropy Spectrum, *Phys. Rev. Lett.* **108** 021301 (2012).

[138] D. Bini, G. Esposito, C. Kiefer, M. Kraemer, and F. Pessina, On the modification of the cosmic microwave background anisotropy spectrum from canonical quantum gravity, *Phys. Rev. D* **87**(10), 104008 (2013).

[139] A. Yu. Kamenshchik, A. Tronconi, and G. Venturi, Inflation and Quantum Gravity in a Born-Oppenheimer Context, *Phys. Lett. B* **726** 518-522 (2013).

[140] A. Yu. Kamenshchik, A. Tronconi, and G. Venturi, Signatures of quantum gravity in a Born–Oppenheimer context, *Phys. Lett. B* **734** 72-78 (2014).

[141] C. Kiefer, The semiclassical approximation to quantum gravity, In Canonical Gravity: From Classical to Quantum (Springer Berlin Heidelberg, Berlin, Heidelberg, 1994).

[142] D. Brizuela, C. Kiefer, M. Kraemer, and S. Robles-Pérez, Quantum-gravity effects for excited states of inflationary perturbations, *Phys. Rev. D* **99**(10), 104007 (2019).

[143] J.C. Hwang, Quantum fluctuations of cosmological perturbations in generalized gravity, *Class. Quant. Grav.* **14** 3327-3336 (1997).

[144] V.F. Mukhanov, H.A. Feldman, and R.H. Brandenberger, Theory of cosmological perturbations. Part 1. Classical perturbations. Part 2. Quantum theory of perturbations. Part 3. Extensions, *Phys. Rept.* **215** 203-333 (1992).

[145] J. Martin, V. Vennin, and P. Peter, Cosmological Inflation and the Quantum Measurement Problem, *Phys. Rev. D* **86** 103524 (2012).

[146] L. Chataignier, Gauge Fixing and the Semiclassical Interpretation of Quantum Cosmology, (2019). (arXiv:1901.07104 [gr-qc])

[147] H. Noh and J.C. Hwang, Inflationary spectra in generalized gravity: Unified forms, *Phys. Lett. B* **515** 231-237 (2001).

[148] P.A.R. Ade *et al.*, Planck 2015 results. XVII. Constraints on primordial non-Gaussianity, *Astron. Astrophys.* **594** A17 (2016).

[149] S. Renaux-Petel, Primordial non-Gaussianities after Planck 2015: an introductory review, *Comptes Rendus Physique* **16** 969-985 (2015).

[150] A.Y. Kamenshchik, A. Tronconi, and G. Venturi, Quantum Cosmology and the Evolution of Inflationary Spectra, *Phys. Rev.* **D94**(12), 123524 (2016).