The Fate of Singularities in Quantum Cosmology and the Application of Generalized Effective Equations to Constrained Quantum Systems

INAUGURAL-DISSERTATION

zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln

vorgelegt von

Barbara Sandhöfer
aus Saarbrücken

Köln 2009

Berichterstatter:	Prof. Dr. Claus Kiefer Prof. Dr. Frank Steiner
Tag der mündlichen Prüfung:	21. April 2009

Zusammenfassung

Meine Dissertation besteht aus zwei Teilen. Im ersten Teil befasse ich mich mit dem Schicksal klassischer, kosmologischer Singularitäten in der Quantenkosmologie. Der zweite Teil widmet sich der Extraktion von Vorhersagen aus der Quantenkosmologie.

Im ersten Teil habe ich zwei Klassen kosmologischer Modelle untersucht. Universen aus der ersten Klasse von Modellen beginnen oder enden ihre Entwicklung mit einer *big-rip*–Singularität. Hier divergieren Energiedichte, Druck und Skalenfaktor in endlicher Zeit. Diese Art von Singularität ist ein generischer Bestandteil von kosmologischen Modellen mit dunkler Energie, die durch ein Phantomfeld generiert wird. Für jedes dieser Modelle wurde auch das entsprechende Gegenstück mit normalem Skalarfeld betrachtet. Durch das gewöhnliche Skalarfeld wird eine *big-bang*–Singularität hervorgerufen. Die zweite Modellklasse, die ich untersucht habe, weist eine *big-brake*–Singularität auf. Am *big brake* wird die Ausdehnung des Universums durch eine unendlich große, negative Beschleunigung zum Stillstand gebracht.

Die Wahl all dieser Modelle ist motiviert duch das Auftreten einer Singularität bei großem Skalenfaktor. Die grundsätzliche Frage, die es zu beantworten galt, war, ob diese Singularitäten auf Quantenebene vermieden werden. Wenn eine solche Vermeidung tatsächlich stattfindet, so ist dies ein Beleg und Beispiel für das Auftreten von Quantengravitationseffekten im makroskopischen Universum.

Nach der Entwicklung von Modellen, die die gewünschten Singularitäten aufweisen, habe ich nämliche quantisiert. Quantisierung erfolgte im geometrodynamischen Zugang. Die zentrale Gleichung ist hier die Wheeler–DeWitt–Gleichung. Ich habe Lösungen zu dieser Gleichungen gefunden, in einem Fall konnte ich sogar eine exakte Lösung angeben. Aus diesen Lösungen habe ich Wellenpakete entlang Bahnen konstruiert, die auf klassischer Ebene in die Singularität führen würden. Die klassichen Bahnen konnten mit Hilfe des Prinzips der konstruktiven Interferenz aus den Paketen abgeleitet werden.

Als Kriterien für eine tatsächliche Vermeidung der Singularität auf dem Quantenlevel wurde das Verschwinden der Wellenfunktion am Ort der klassischen Singularität, sowie das Verschmieren von Wellenpaketen bei Annähe-

rung an die klassische Singularität benutzt. Dies entspricht einem Zusammenbruch der semi-klassischen Näherung und einer Auflösung der Raumzeit.

In allen Fällen konnte ich Singularitätsvermeidung nachweisen. Im Falle der *big-bang* und der *big-brake* Singularität, die beide bei endlichen Werten der Konfigurationsraumvariablen auftreten, verschwindet die Wellenfunktion. Für die *big-brake* Singularität zerfließt zusätzlich das Wellenpaket. Die ist nicht der Fall für die *big-bang* Singularität. Die *big-rip* Singularität liegt im Unendlichen. Das Wellenpaket zerläuft bei Annäherung an diese Singularität, verschwindet aber nicht.

Im zweiten Teil meiner Arbeit habe ich mich mit der Anwendung von generalisierten, effektiven Gleichungen auf Systeme mit Zwangsbedingungen beschäftigt. Generalisierte, effektive Gleichungen benutzen zur Beschreibung von Quantensystemen die Erwartungswerte fundamentaler Operatoren und die höheren Momente der Wellenfunktion anstelle der Wellenfunktion selbst. Dieser Formalismus ist damit wie geschaffen für die Extraktion von Vorhersagen aus der Quantenkosmologie, zum Beispiel in der Form von Korrekturen zu klassischen Bewegungsgleichungen. Der Formalismus ist auf einem, im allgemeinen Fall, unendlich-dimensionalen Quantenphasenraum aufgebaut.

Die erste Aufgabe bei der Anwendung auf Systeme mit Zwangsbedingungen war die Übertragung der Dirac'schen Quantisierungsregel für Zwangsbedingungen auf diesen Quantenphasenraum. Die resultierenden Zwangsbedingungen — es entstehen tatsächlich unendlich viele Zwangsbedingungen auf dem Quantenphasenraum — eliminieren die unphysikalischen Freiheitsgrade in der zu erwartenden Art und Weise. Dies wurde für den Fall einer einzelnen, linearen Zwangsbedingung gezeigt. Daraus folgt, daß auch für eine beliebige Zwangsbedingung die Freiheitsgrade, zumindest lokal, korrekt eliminiert werden.

In einem zweiten Schritt mußte die neugefundene Menge an Zwangsbedingungen konsistent genähert werden, so daß eine endliche Zahl an Bedingungen verbleibt. Nur dann kann man dem System überhaupt Informationen entziehen.

Ein solches Näherungsverfahren wurde für nicht-relativistische Systeme entwickelt und am Beispiel des parametrisierten, freien, nicht-relativistischen Teilchens demonstriert.

Abstract

This thesis consists of two parts. The first part is concerned with the fate of singularities in quantum cosmology. The second part addresses the derivation of predictions from quantum cosmology.

In the first part, I studied two classes of cosmological models. In the first class of models, the universe evolves to or emerges from a big-rip singularity. Here, energy density, pressure and scale factor diverge after a finite amount of time. This type of singularity arises rather generically in cosmological models with phantom dark energy. For each of these phantom-field models, the corresponding scenario with ordinary scalar field was studied. The scalar field induced a big-bang singularity. The second class of models studied was dominated by a big-brake singularity. At the big brake, the universe evolution comes to a halt due to an infinite deceleration.

The motivation behind this choice of models was the occurrence of a singularity at large scale factor. The major question pursued was whether these types of singularity were resolved on the quantum level. If such singularities were resolved in quantum cosmology, this would imply that quantum gravitational effects can occur in the macroscopic universe.

After devising classical models that contain the respective singularity, I subjected these models to quantization which was carried out in the geometrodynamical approach. The governing equation is then the Wheeler–DeWitt equation. I found solutions to the Wheeler–DeWitt equation, in one case even an exact solution. Wave packets were constructed around trajectories which, on the classical level, would lead into the singularity. I have then shown that the classical trajectory can indeed be recovered from these packets through the principle of constructive interference.

As criteria for singularity avoidance, the vanishing of the wave function at the location of the classical singularity, as well as the spreading of wave packets upon approach of this region, was used. Whereas the former ensures that the classical singularity does not contribute to the quantum theory, the latter signals a dissolution of the semi-classical approximation and thus of spacetime.

In all cases, I found singularity resolution. In the case of the big-bang and big-brake singularities, the wave function vanishes at the classical singularity. These two have in common that they occur at finite value of the

configuration space variables. A spreading of the wave packet is however only observed upon approach of the big-brake singularity. A strict vanishing cannot be found at the location of the classical big-rip singularity. This singularity is located at the infinite boundary of configuration space. The wave packet spreads upon approach of this singularity.

The second part of my thesis deals with the application of the generalized effective-equation scheme to constrained systems. Generalized effective equations describe a quantum system via expectation values of fundamental operators and higher moments of the wave function — instead of using the wave function itself. It is thus a very useful scheme for the derivation of predictions from quantum cosmology, e.g. in the form of corrections to classical equations of motion. The theory is formulated on a, generally, infinite-dimensional quantum phase space. The first task was to find a formulation of Dirac's constraint-quantization condition on this phase space. Such a formulation was found and proven to remove degrees of freedom appropriately in the case of a single linear constraint. This result ensures the correct removal of degrees of freedom for any singly constrained system at least locally. In a second step, the newly formulated constraints — there are actually infinitely many of them — had to be consistently approximated. Such an approximation is necessary to reduce the infinite number of constraints to a finite one. Only then can information be extracted from the system.

Such an approximation scheme for non-relativistic systems was developed. Its consistency was explicitly checked using the parametrised, free non-relativistic particle.

Contents

Notation and conventions

Starting point of all our derivations is a four-dimensional manifold $\mathcal{M}$ with metric g of signature $(-,+,+,+)$. We will consider spatial hypersurfaces Σ in $\mathcal{M}$ which are assumed to be compact throughout.

Greek indices range from $0,\ldots 3$ and latin indices from $1,\ldots 3$. Round brackets around indices denote symmetrisation, $x_{(ab)}$, and square brackets anitsymmetrisation, $x_{[ab]}$.

Poisson brackets will be denoted by $\{\cdot,\cdot\}$ and commutators by $[\cdot,\cdot]$.

We will use the abbreviation $\kappa^2 = \frac{8\pi G}{c^4}$ containing the gravitational constant G and the speed of light c. Also $C = \frac{3\mathcal{V}}{\kappa^2}$ will be used as abbreviation.

I decided to follow the conventions in the choice of variables for the different quantities. As a consequence, some letters occur with double meaning and in different styles, especially p and H. If the same letter is used for different quantities, then this happens in different contexts so that no confusion can arise.

Units are chosen such that $c = 1$. The constant c will only be retained in exceptional cases.

Geometrical quantities			
g	spacetime metric	h	metric on Σ
K	extrinsic curvature	$G_{\alpha\beta}$	Einstein tensor
x	points in Σ	X	points in $\mathcal{M}$

Derivatives			
$_{,\,\alpha}$	partial derivative	∇_α	covariant derivative of g
$_{\mid a}$	covariant derivative of h		

Canonical formulation			
h_{ab}	three-metric on Σ	p^{ab}	conjugate momentum of h_{ab}
N	lapse function	p_N	conjugate momentum of N
N^a	shift vector	p_{N^a}	conjugate momentum of N^a
ϕ	scalar field	p_ϕ	conjugate momentum of ϕ
$\mathcal{V}$	volume of Σ	$V(\phi)$	scalar-field potential
$\mathcal{H}_\perp$	Hamiltonian constraint	$\mathcal{H}_a$	diffeomorphism constraints
$\mathbf{\Gamma}$	phase space	Γ_c	constraint hypersurface
Q_x	space of three-metrics at x	$Q(\Sigma)$	configuration space of GR
$\mathcal{Q}$	superspace		

Cosmological quantities			
p	pressure	ρ	energy density
w	$= \frac{p}{\rho}$ equation of state		
Λ	cosmological constant	k	curvature index
λ	$= C\Lambda$	$\mathcal{K}$	$= Ck$

Effective equations			
q	canonical coordinate	p	conjugate momentum of q
$G^{a,b}_{c,d}$	quantum variable	C_Q	principal constraint

Chapter 1

Introduction

General relativity is the theory describing the gravitational interaction. Even though it is experimentally well verified, it has the disturbing feature that it predicts its own break-down in the form of spacetime singularities. Famous examples are the big-bang singularity arising in cosmological spacetimes and the black-hole singularity arising in spherically symmetric spacetimes.

As these are both highly symmetric solutions to Einstein's equations, one might hope that the singular nature is due to the high symmetry and not to general relativity itself. However, the singularity theorems of Hawking and Penrose squash this hope. Singularities are a generic feature of physically interesting spacetimes. So we really need a theory that goes beyond general relativity — this is at least the usual introduction found in many quantum-gravity reviews.

The general assumption is that 'the theory beyond' should be a quantum theory transcending general relativity. Depending on the own capacity of transcendation, one of the following two paths is pursued. Either, one can attempt to include general relativity in a quantum theory comprising all four known interactions, i.e. to devise a grand unified theory. Or, one can contend oneself with a mere quantization of general relativity.

We reserve the name quantum gravity for the latter approach.

Simply subjecting general relativity to quantization leads to quantum general relativity. Here, 'simply' refers to a straight forward application of canonical or path-integral quantization rules. In this thesis, we work in the functional Schrödinger picture of quantum general relativity, i.e. in the canonical formulation. This approach is known as quantum geometrodynamics and often treated as opposed to the corresponding path-integral formulation. But we see quantum geometrodynamics just as one side of the coin of quantum general relativity.

Quantum general relativity is the oldest candidate for a quantum theory of gravity. The problem with this theory is that it does not fit into the

mathematical corset of ordinary quantum theory. Therefore many other approaches have arisen over the years. On the canonical level, the most fervent opponent of quantum general relativity is loop quantum gravity. Being not only faced with problems but also with an alternative, it seems wise to review the path taken and examine it carefully. This will be the content of Chapters 2 – 4.

But let us come back to the singularity issue — which is actually intermingled with the list of problems of quantum general relativity.

As it was introduced above, the resolution of classical spacetime singularities forms a major motivation for the setup of quantum gravity. Let us take the time to check this motivation. Should we really expect that quantum general relativity resolves all classical singularities? Which arguments can be given in favour of this hope?

Another expectation that is generally raised in connection with the resolution of classical singularities is that quantum general relativity itself be a singularity-free theory. Note that this issue is a priori unrelated to the first one. Furthermore, it is our aim to elucidate whether the expectation of singularity-freedom is a valid and well-founded one. Chapter 5 is devoted to a critical review of these two expectations connected with singularities on the quantum gravitational level.

Apart from just checking whether these two expectations are justified, we want to study whether they are actually met. That means, we discuss whether classical singularities are resolved on the quantum level and whether quantum general relativity is itself a singularity-free theory. It is difficult to arrive at general statements on these issues. A tentative answer to the second question is, however, given in Chapter 5. Following the results of this chapter, criteria are developed which account for singularity resolution on the quantum level.

Armed with these criteria, we approach the first question, namely the question of singularity resolution on the quantum level. It is discussed for two different types of quantum-cosmological singularity in Chapters 6 and 7.

Quantum cosmology is the adaptation of quantum general relativity to cosmology. It has two main applications. Firstly, it is considered as the theory relevant for the study of the early Universe. Secondly, it is often used as a toy model for full quantum general relativity. This is the sense in which it is employed in Chapters 6 and 7.

Of course, the cosmological singularity of primary interest for *our* Universe is the big bang. But other, more exotic types of singularity also exist. The two singularities studied here are the big rip and the big brake. They both occur at large scale factor, i.e. in the macroscopic universe. We analyze what happens to these singularities upon quantization. Can we expect quantum gravitational effects in the macroscopic universe?

The last chapter is devoted to a rather new effective formulation developed

for quantum cosmology. This formulation is based on expectation values and higher moments of the wave function instead of the wave function itself. Therefore it is very well suited for the extraction of predictions from quantum cosmology. In Chapter 7 a formalism is developed that allows the application of this effective scheme to constrained, non-relativistic systems.

Chapter 2

Hamiltonian formulation of General Relativity

The Hamiltonian formulation is the classical starting point of canonical quantum general relativity. It is based on a $(3+1)$*-decomposition of spacetime. The corresponding action arising from the Einstein–Hilbert action is the so-called Arnowitt–Deser–Misner action. Most significantly, the Hamiltonian dynamics of general relativity turns out to be constrained, the Hamiltonian itself being one of the constraints.*

2.1 The Arnowitt–Deser–Misner action

Any canonical formulation rests on the isolation of velocities of configuration space variables. From these, the conjugate momenta are inferred which then span up phase space. Thus to obtain a canonical formulation one has to identify a time variable. For general relativity this implies that four-dimensional spacetime has to be decomposed into spatial hypersurfaces. On each hypersurface, canonical variables can then be defined which evolve along the foliation. The resulting formalism is implicitly generally covariant if it does not depend on a specific choice of time, i.e. foliation.

The decomposition of the Einstein–Hilbert action into time and space directions is called *Arnowitt–Deser–Misner action*, or ADM-action for short.

In the following, I want to present the $(3+1)$-decomposition and derive the ADM-action for pure gravity. Finally, some remarks are made on the inclusion of matter.

We start from a Lorentzian manifold $\mathcal{M}$ with spacetime metric g of signature $(-,+,+,+)$. This metric is a solution of Einstein's equations derived from the Einstein–Hilbert action

$$S_{\mathrm{EH}}[g] = \frac{c^4}{16\pi \mathrm{G}} \int\limits_{\mathcal{M}} \mathrm{d}^4x \, \sqrt{-\det g} \, (\mathrm{R} - 2\Lambda) \ .$$

The prefactor containing the gravitational constant G and the velocity of light c will be abbreviated through $\kappa^2 = \frac{8\pi \mathrm{G}}{c^4}$. The determinant of g will be written as $\det g$, R is the Ricci scalar of g. The cosmological constant is denoted by Λ.

2.1.1 The $(3+1)$-decomposition

Now we decompose this spacetime manifold $\mathcal{M}$ into three-dimensional, spacelike hypersurfaces. These hypersurfaces shall actually be Cauchy hypersurfaces: the specification of initial data on these hypersurfaces shall determine a (up to diffeomorphisms) unique solution to Einstein's equations. We know that such hypersurfaces do not exist for arbitrary spacetimes but only for globally hyperbolic ones.[1]

The foliation of spacetime

Globally hyperbolic spacetimes are diffeomorphic to a product manifold of a three-dimensional, spacelike manifold, Σ, with the real line. This diffeomorphism is just the foliation

$$\begin{array}{rccc} \mathcal{E}: & \Sigma \times R & \longrightarrow & \mathcal{M} \\ & (x,t) & \longmapsto & \mathcal{E}(x,t) \ . \end{array}$$

In the following, *we will assume Σ to be compact.* For each $t \in \mathbb{R}$, we get a special embedding of the hypersurface Σ into $\mathcal{M}$

$$\begin{array}{rccc} \mathcal{E}_t: & \Sigma & \longrightarrow & \mathcal{M} \\ & x & \longmapsto & \mathcal{E}_t(x) = \mathcal{E}(x,t) \ . \end{array}$$

In this way, a global time function can be defined. Namely let

$$\begin{array}{rccc} \mathcal{E}^{-1}: & \mathcal{M} & \longrightarrow & \Sigma \times \mathbb{R} \\ & X & \longmapsto & \mathcal{E}^{-1}(X) = (\sigma(X), \tau(X)) \ , \end{array}$$

where $\sigma : \mathcal{M} \longrightarrow \Sigma$ and $\tau : \mathcal{M} \longrightarrow \mathbb{R}$. Then $\tau(\mathcal{E}_t(x)) = t$. This map associates with each point in $\mathcal{M}$ a time coordinate in $\mathbb{R}$, the *foliation time.*

[1] For non-globally hyperbolic spacetimes, data on a three-hypersurface determine the metric only in a limited region, the domain of dependence.

For all $x \in \Sigma$, the map $\mathcal{E}_x : \mathbb{R} \longrightarrow \mathcal{M}$ defines a timelike curve in $\mathcal{M}$. Its tangent vector is $\dot{\mathcal{E}}_x$, where the dot denotes derivation with respect to t. As this holds at each $x \in \Sigma$, this actually defines a tangent vector field. It is called *deformation vector field*. At each point $X \in \mathcal{E}_t(\Sigma)$ it describes the change of the embedding as a function of t. As a vector field it thus specifies how one hypersurface $\mathcal{E}_t(\Sigma)$ is deformed into an infinitesimally neighbouring one $\mathcal{E}_{t+\delta t}(\Sigma)$, see Figure 2.1.

If we introduce coordinates X^α, $\alpha = 0, \ldots 3$, on $\mathcal{M}$ — which are not to be confused with the *point* $X \in \mathcal{M}$ — this field has components $\dot{\mathcal{E}}^\alpha_x$ where $\mathcal{E}^\alpha = X^\alpha(\mathcal{E}_t(x))$ at each point x in the hypersurface.

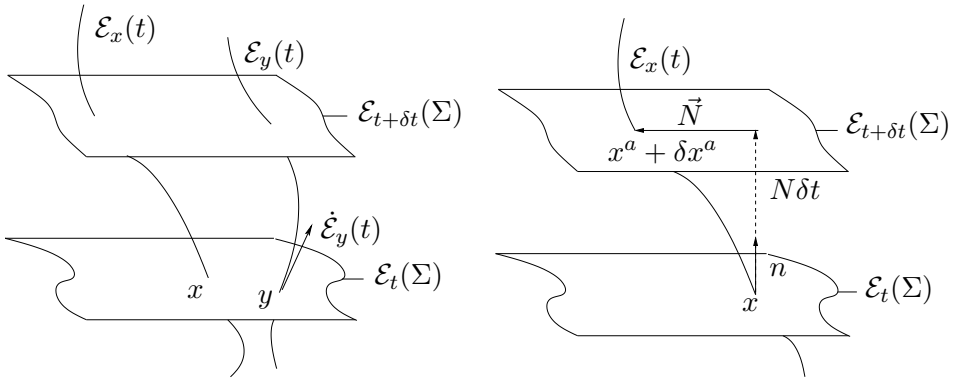

Figure 2.1: Two embeddings of the hypersurface Σ into $\mathcal{M}$ are shown. These embeddings are separated by a foliation-time interval δt. On the left-hand side, the timelike curves $\mathcal{E}_x(t)$, $\mathcal{E}_y(t)$ for two points x, $y \in \mathcal{E}_t(\Sigma)$ are shown. The deformation vector $\dot{\mathcal{E}}_y(t)$ is depicted at y. The figure on the right-hand side illustrates the geometrical interpretation of lapse and shift. Following the hypersurface normal n by an amount $N\delta t$ along a geodesic to the intersection point with $\mathcal{E}_{t+\delta t}(\Sigma)$ and then N^a by an amount $-\delta x^a$, we arrive at the intersection point of $\mathcal{E}_x(t)$ with $\mathcal{E}_{t+\delta t}(\Sigma)$.

Keeping in mind that we want a theory that is defined on spatial hypersurfaces, we have to strife for a way to separate quantities which lie in the hypersurface from those orthogonal to it. This can be done with the help of the hypersurface normal vector field n. For each embedding $\mathcal{E}_t$ it is defined via

$$n_\alpha(x)\mathcal{E}^\alpha_t{}_{,a}(x) = 0 \ ,$$

where $\mathcal{E}^\alpha_t{}_{,a}(x) = \frac{\mathrm{d}\mathcal{E}^\alpha(x,t)}{\mathrm{d}x^a}$ and $x \in \Sigma$. Furthermore, as we are interested in *spacelike* embeddings, we want this normal to satisfy

$$g^{\alpha\beta}(x)n_\alpha(x)n_\beta(x) = -1 \ .$$

Both equations imply that the normal vector field depends on the spacetime metric g as well as on the embedding $\mathcal{E}_t$. So, to be correct, we should write $n = n(x; \mathcal{E}_t, g]$. For the sake of simplicity, I will refrain from doing so.

With these quantities, $n_\alpha(x)$ and $\mathcal{E}_t^\alpha{}_{,a}(x)$, we can decompose each tensor into its hypersurface-orthogonal and -normal parts. For the deformation vector, we write

$$\dot{\mathcal{E}}_x^\alpha(t) = N(x,t) g^{\alpha\beta} n_\beta + N^a(x,t) \mathcal{E}_t^\alpha{}_{,a} \ .$$

We call the function $N(x,t)$ *lapse function* and the three-vector $\vec{N}(x,t)$ with components $N^a(x,t)$ *shift vector*. I use (x,t) as a short-hand to denote their dependence on the embedding and the location on the hypersurface Σ. The correct notation would be: $N = N(x; \mathcal{E}_t, g]$ and similarly for the shift vector.

As can be seen from their definition, they describe the change of the hypersurface with t in the orthogonal and tangential directions, respectively. More precisely, for two embeddings $\mathcal{E}_t(\Sigma)$ and $\mathcal{E}_{t+\delta t}(\Sigma)$, $N(x,t)\delta t$ gives the proper time separation between $\mathcal{E}_t(\Sigma)$ and $\mathcal{E}_{t+\delta t}(\Sigma)$ normal to $\mathcal{E}_t(\Sigma)$, $N(x,t)\delta t = \delta_\perp \tau(x)$. Similarly (but more complicated to phrase), $N^a(x,t)\delta t$ describes the displacement of $\mathcal{E}_{t+\delta t}(x)$ with respect to the intersection point of the geodesic normal to $\mathcal{E}_t(x)$ with $\mathcal{E}_{t+\delta t}(\Sigma)$. Let the intersection point have coordinates $x^a + \delta x^a$ on Σ, then we can write this as $N^a(x,t)\delta t = -\delta x^a$, see Figure 2.1 for clarification.

$(3+1)$-decomposition of the spacetime metric

We consider the pull-back $\mathcal{E}^*$ of the spacetime metric g under the foliation $\mathcal{E}$ to $\Sigma \times \mathbb{R}$. Coordinates on $\Sigma \times \mathbb{R}$ are adapted to the product structure of this manifold. Namely, the coordinates Y^α are chosen such that $Y^0 = t$ and $Y^a = x^a$, $a = 1, \ldots 3$, and x^a is the coordinate system on Σ. Then we get the following metric components under pull-back

$$\begin{aligned} (\mathcal{E}^* g)_{00}(x,t) &= h_{ab} N^a N^b - N^2 \\ (\mathcal{E}^* g)_{a0}(x,t) &= h_{ab} N^b \\ (\mathcal{E}^* g)_{ab}(x,t) &= h_{ab} \ . \end{aligned}$$

So lapse and shift are basically the time-space and time-time components of the space-time metric g. Often one finds the short-hand version of this decomposition, leaving out the details connected with the embedding. Then the spacetime metric is simply given in its ADM-form

$$g_{\alpha\beta} = \begin{pmatrix} -N^2 + h_{ab} N^a N^b & h_{ab} N^a \\ h_{ab} N^b & h_{ab} \end{pmatrix} \ , \tag{2.1}$$

also referred to as ADM-metric. As is obvious from the left-hand side of this equation, this formula depends on the choice of coordinate system on $\mathcal{M}$.

The extrinsic curvature

We want to make one last observation before turning to the decomposition of the action. This concerns the change of the metric on an initial surface $\mathcal{E}_t(\Sigma)$ under evolution to the neighbouring one, $\mathcal{E}_{t+\delta t}(\Sigma)$, projected along the hypersurface normal. With hindsight one might guess that this quantity will serve as metric-velocity. One finds

$$n^\alpha(x)\frac{\delta h_{ab}(y)}{\delta \mathcal{E}_t^\alpha(x)} = -2K_{ab}(x)\delta(x-y)\ . \tag{2.2}$$

K_{ab} is the extrinsic curvature of $\mathcal{E}_t(\Sigma)$ and describes the way the hypersurface is embedded into $\mathcal{M}$. As such it characterizes the *extrinsic geometry* of $\mathcal{E}_t(\Sigma)$. This can be seen from its definition

$$K_{ab} = -\mathcal{E}_t^\alpha{}_{,a}\,\mathcal{E}_t^\beta{}_{,b}\,\nabla_\alpha n_\beta$$

where ∇ is the covariant derivative of g. From this follows also that the extrinsic curvature has components only *in* the three-hypersurface. Using lapse and shift, we can write this as

$$K_{ab} = \frac{1}{2N}\left(-\dot{h}_{ab} + \mathcal{L}_{\vec{N}}h_{ab}\right)\ , \tag{2.3}$$

where $\mathcal{L}_{\vec{N}}h_{ab}$ is the Lie derivative of h_{ab} along the shift vector.

2.1.2 The Arnowitt–Deser–Misner action

The ADM-formulation of the Einstein–Hilbert action consists now in pulling back the Einstein–Hilbert Lagrangian with the help of $\mathcal{E}$ and expressing the result in terms of three-metric, extrinsic curvature, lapse and shift. This amounts to inserting (2.1) into the Einstein-Hilbert Lagrangian, making use of (2.3) and isolating as many total derivatives as possible.

One arrives at the following form of the action

$$S_{\rm ADM} = \frac{1}{2\kappa^2}\int \mathrm{d}t \int_\Sigma \mathrm{d}^3x N\sqrt{\det h}\left[K_{ab}K^{ab} - (K^a_a)^2 + \left({}^{(3)}R - 2\Lambda\right)\right] - \Delta S_{\rm ADM}\ , \tag{2.4}$$

where ${}^{(3)}R$ denotes the Ricci scalar of the three-metric h. The determinant of this metric is denoted by $\det h$ and

$$\begin{aligned}\Delta S_{\rm ADM} &= 2(\sqrt{\det h}K)_{,t} - (\sqrt{\det h}(KN^b - h^{ab}N_{,a}))_{,b} \\ &= \left(2\sqrt{-\det g}(n^\alpha \nabla_\beta n^\beta - n^\beta \nabla_\beta n^\alpha)\right)_{,\alpha}\ .\end{aligned}$$

This total-derivative term is usually dropped. Recall that we restricted the discussion to compact Σ.

Note that the ADM-action is first-order in time-derivatives of the metric. Moreover, all quantities occurring in this action — apart from lapse function and shift vector — lie in the three-hypersurface, i.e. they are three-covariant. We thus put the action into first-order form and reformulated it in terms of quantities defined entirely on the spatial hypersurface. These were the principles according to which Arnowitt, Deser and Misner derived the ADM-action originally.

2.2 The Hamiltonian of General Relativity

The canonical formulation is set up on the three-hypersurface Σ. As explained in the beginning, this hypersurface is a Cauchy surface of $(\mathcal{M}, g)$. The ADM-metric components and their foliation-time derivatives are thought of as being prescribed on this hypersurface. The components themselves serve as canonical coordinates, the time derivatives are thought of as velocities. The time dependence of both, coordinates as well as velocities, are not known but to be determined by the equations of motion.

Due to its first-order form and due to the fact that all quantities it depends on are defined on Σ, we can use the ADM-action as starting point of the canonical formulation. But we have to forget about the construction of S_{ADM} and just consider the hypersurface and the quantities defined on it as fundamental — and not as having arisen from a decomposition of a spacetime metric g.

We use (2.4) with $\Delta S_{\mathrm{ADM}} = 0$.

2.2.1 Phase space variables of General Relativity

We thus have as canonical coordinates lapse $N(x)$, shift $\vec{N}(x)$ and three-metric $h(x)$ at each $x \in \Sigma$. The momenta corresponding to lapse and shift, p_N, p_{N^a}, vanish as the action does not depend on their velocities (*if* the total derivative in (2.4) is removed[2]),

$$p_N \approx 0 \ , \qquad p_{N^a} \approx 0 \ . \tag{2.5}$$

These are primary constraints! The Hamiltonian dynamics of general relativity is thus constrained. This is a consequence of the fact that the ADM-Lagrangian is singular. The algorithm that tells us how to proceed with such a constrained system on the Hamiltonian level is the *Bergmann–Dirac algorithm*. It is explained in Appendix A. In the following, I will use the terminology of constrained systems as presented there. Equations (2.5) define the *primary constraint surface*.

[2]Otherwise the linear dependence of the action on the velocities of lapse and shift yields their canonical momenta as functions of the three-metric components, see Dirac's own derivation [2].

For the three-metric components, we find momenta

$$p^{ab} = \frac{\delta S_{\rm ADM}}{\delta \dot{h}_{ab}} = -\frac{\sqrt{\det h}}{\kappa^2}\left(K^{ab} - h^{ab}K^c_c\right) \ . \tag{2.6}$$

So whereas the three-metric describes the *intrinsic* geometry of Σ, its canonical momentum characterizes the *extrinsic* geometry.

The space $(N, N^a, h_{ab}; p_N, p_{N^a}, p_{ab})$ defines the phase space of general relativity. It can be equipped with a Poisson bracket in the conventional way. The only non-vanishing brackets are[3]

$$\begin{aligned} \{N_\alpha(x), p^{0\beta}(y)\} &= \delta^\beta_\alpha \delta(x-y) \ , \\ \{h_{ab}(x), p^{cd}(y)\} &= \delta^{(c}_{(a}\delta^{d)}_{b)}\delta(x-y) \ , \end{aligned}$$

where I introduced $N_\alpha = (N, N^a)$ and $p^{0\beta} = (p_N, p_{N^a})$ for the sake of shortness.

The geometrical structure of this phase space is, to my perception, best discussed in a review by Fischer and Marsden, [16]. There, the ADM-decomposition is described in differential geometric language. The main result is that the phase space of general relativity does form a cotangent bundle over configuration space. More precisely, let configuration space $Q(\Sigma)$ be the space of all C^∞, Riemannian metrics h on Σ. The tangent bundle $TQ(\Sigma)$ associates with each Riemannian metric a C^∞, 2-covariant symmetric tensor field on Σ. This can be thought of as the 'velocity' K of h. The cotangent bundle is defined as $T^*Q(\Sigma)$ and associates a C^∞ symmetric 2-contravariant tensor density of weight one with each metric. This can be thought of as canonically conjugate momentum of h. This cotangent bundle can then be equipped with the above Poisson bracket which turns it into a symplectic manifold.

2.2.2 The Hamiltonian of General Relativity

Equation (2.6) can be solved for the velocities,

$$K_{ab} = -\frac{\kappa^2}{\sqrt{\det h}}\left(p^{ab} - \frac{1}{2}h^{ab}p^c_c\right) \ ,$$

so that one arrives at the canonical Hamiltonian density through

$$\mathcal{H}(x; h, p, N, \vec{N}] = p_N\dot{N} + p_{N^a}\dot{N}^a + p^{ab}\dot{h}_{ab} - \mathcal{L}_{\rm ADM}[h, N, \vec{N}] \ ,$$

[3]The brackets on the right-hand side of the second Poisson bracket denote symmetrization on the product of Kronecker δ.

as

$$\begin{aligned}\mathcal{H}(x;h,p,N,\vec{N}] &= 2p^{ab}N_{a|b} \\ &\quad -\frac{N}{\sqrt{\det h}}\left(\frac{1}{2}(p^a_a)^2 - p^{ab}p_{ab} + \det h\left({}^{(3)}R - 2\Lambda\right)\right) ,\end{aligned}$$

where the vertical bar denotes the covariant derivative of h. After the neglection of a surface term,

$$\begin{aligned}\mathcal{H}(x;h,p,N,\vec{N}] &= N(x,t)\mathcal{H}_\perp(x;h,p] + N^a(x,t)\mathcal{H}_a(x;h,p] \\ &= N^\alpha\mathcal{H}_\alpha(x;h,p,N,\vec{N}] ,\end{aligned}$$

were we introduced $\mathcal{H}_\alpha = (\mathcal{H}_\perp, \mathcal{H}_a)$ for the sake of shortness. The hypersurface-orthogonal and -tangential parts of the Hamiltonian density are given by

$$\begin{aligned}\mathcal{H}_a(x;h,p] &= -2p^b{}_{a|b}(x) , \\ \mathcal{H}_\perp(x;h,p] &= \frac{\kappa^2}{2\sqrt{\det h}}\left[h_{ac}(x)h_{bd}(x) + h_{bc}(x)h_{ad}(x) - h_{ab}(x)h_{cd}(x)\right] \\ &\quad \times p^{ab}(x)p^{cd}(x) - \frac{\sqrt{\det h(x)}}{\kappa^2}\left({}^{(3)}R(x;h] - 2\Lambda\right) .\end{aligned}$$

The tensor

$$\mathcal{G}_{abcd}(x) := \frac{1}{2\sqrt{\det h(x)}}\left[h_{ac}(x)h_{bd}(x) + h_{bc}(x)h_{ad}(x) - h_{ab}(x)h_{cd}(x)\right]$$

is called *DeWitt metric*. We will come to its meaning in Chapter 5. The Hamiltonian itself will be denoted by

$$\mathrm{H}^C = \int \mathrm{d}^3x\, N^\alpha\mathcal{H}_\alpha = \int \mathrm{d}^3x\, \mathcal{H}^C .$$

2.3 Constraints of General Relativity

The occurence of primary constraints entails more constraints (otherwise one would hardly call them primary). These arise from the requirement that the primary constraint surface be conserved under evolution through the primary Hamiltonian. From this, an iteration procedure results which shrinks the primary constraint surface to the constraint surface and terminates when all constraints are conserved under evolution.

Furthermore, constraints generate transformations on phase space.

In this section, the complete set of constraints for general relativity will be determined and the transformations they generate analyzed. Finally, we will introduce the concept of the partially reduced phase space which is the starting point for quantization.

2.3.1 The set of constraints

The primary Hamiltonian is obtained from the canonical one through the addition of the primary constraints,

$$\mathcal{H}^P(x; h, p, N, \vec{N}] = \mathcal{H}(x; h, p, N, \vec{N}] + v_\alpha p^{0\alpha} \ , \tag{2.7}$$

with arbitrary functions v_α. In the following, I will drop the dependence of the Hamiltonian on $(x; h, p, N, \vec{N}]$.

We require the primary constraints (2.5) to be conserved in time. This yields the consistency conditions

$$\dot{p}^{0\alpha} = \{p^{0\alpha}, \mathcal{H}^P\} = \mathcal{H}^\alpha \approx 0 \ .$$

Thus, conservation of the primary constraints yields the Hamiltonian components as secondary constraints,

$$\mathcal{H}_\perp(x; h, p] \approx 0 \ , \qquad \mathcal{H}_a(x; h, p] \approx 0 \ .$$

The total Hamiltonian is therefore constrained to vanish.[4] The constraints $p_N \approx 0$, $p_{N^a} \approx 0$, $\mathcal{H}_\perp \approx 0$ and $\mathcal{H}_a \approx 0$ define the *secondary* constraint surface. Requiring conservation of constraints on it, yields no new constraints. The set $p_N \approx 0$, $p_{N^a} \approx 0$, $\mathcal{H}_\perp \approx 0$ and $\mathcal{H}_a \approx 0$ is the complete set of constraints for general relativity, the secondary constraint surface is *the* constraint hypersurface.

We will denote this hypersurface in phase space by Γ_c.

2.3.2 Transformations generated by the constraints

Evaluating Poisson brackets between all constraints, they are found to vanish on Γ_c. This follows trivially for Poisson brackets involving the primary constraints. Poisson brackets between secondary constraints are more complicated to evaluate. They are given by

[4] Alternatively, one can arrive at this result from the Hamiltonian form of the action

$$S_{\mathrm{ADM}} = \int \mathrm{d}t \int\limits_\Sigma \mathrm{d}^3x \left(p^{ab}\dot{h}_{ab} - N\mathcal{H}_\perp - N^a\mathcal{H}_a \right)$$

through variation with respect to lapse and shift.

$$\{\mathcal{H}_\perp(x), \mathcal{H}_\perp(y)\} = 2\mathcal{H}^b(x)\delta,_b(x-y) + \mathcal{H}^b,_b(y)\delta(x-y) \ , \quad (2.8)$$

$$\{\mathcal{H}_\perp(x), \mathcal{H}_a(y)\} = \mathcal{H}_\perp(y)\delta,_a(x-y) \ , \quad (2.9)$$

$$\{\mathcal{H}_a(x), \mathcal{H}_b(y)\} = \mathcal{H}_b(x)\delta,_a(x-y) + \mathcal{H}_a(y)\delta,_b(x-y) \ . \quad (2.10)$$

This algebra is called *Dirac algebra*. It will be discussed in Section 2.5.

Thus, Poisson brackets between secondary constraints vanish also on Γ_c. We conclude that the set of constraints is first class. First-class constraints generate gauge transformations. Which transformations are generated by the constraints of general relativity?

The primary ones, $p_N \approx 0$, $p_{N^a} \approx 0$, correspond to arbitrary lapse and shift. But these just specify the coordinate system *off* the hypersurface Σ. Thus, these primary first-class constraints signal that the canonical formulation does not depend on the coordinate system *surrounding* Σ. Gauge transformations generated by these constraints are just coordinate transformations that leave the coordinates on Σ untouched.

The transformations generated by secondary constraints, $\mathcal{H}_\perp \approx 0$, $\mathcal{H}_a \approx 0$, are somewhat more subtle to derive. We can approach the problem from the total Hamiltonian $\mathrm{H} = \int \mathrm{d}^3x\mathcal{H}$. This clearly defines transformations on phase space mapping $p^{ab} \to p^{ab\prime}$ and $h_{ab} \to h'_{ab}$ such that both, (h_{ab}, p^{ab}) and $(h'_{ab}, p^{ab\prime})$ describe solutions to Einstein's equation. That means they both correspond to the same spacetime metric g, but arise from a different slicing of the latter. The points on the orbit of (h_{ab}, p^{ab}) generated by H^C correspond to one and the same solution of Einstein's equation, but expressed in different coordinate systems. If one defines a physical state in general relativity as a spacetime, then $\mathrm{H}_\perp$ and H_a are rightly called the generators of gauge transformations in the sense defined in Appendix A.[5]

Splitting the generator of four-diffeomorphisms into its tangential and orthogonal parts, we conclude that $\mathcal{H}_\perp$ relates spatial hypersurfaces that arise from different leaves of the foliation. The tangential components, on the other hand, generate diffeomorphisms in the hypersurface. For a fixed foliation, the orthogonal part clearly determines the dynamics whereas the tangential parts just change the coordinate system in the hypersurface.

Therefore, the tangential parts are usually referred to as *diffeomorphism constraints*, whereas the orthogonal part is somewhat ambiguously called *Hamiltonian constraint*.

[5]The term 'gauge' is used in different meanings and contexts in physics. Here, we use the terms *local* and *rigid gauge transformations* for those transformations to which Noether's first and second theorem applies. The term *gauge transformation*, on the other hand, is reserved for the transformations generated by first-class constraints that leave the physical state invariant.

2.3.3 The partially reduced phase space

Usually, one fixes the coordinate system off Σ through a specific choice of N, N^a and assumes the first-class primary constraints to be satisfied. This choice of coordinate system in the canonical formulation corresponds to a choice of gauge. It just fixes the foliation. Covariance is preserved as any choice is possible.

Then $\mathcal{H}_\perp$, $\mathcal{H}_a$ describe how the hypersurface Σ 'moves' through this 'background'. We will assume that such a gauge fixing has been carried out *before quantization*. This does not restrict our formalism or makes it less covariant as lapse and shift will not enter on the quantum level, as will be seen later.

2.4 Equations of motion

2.4.1 Hamiltonian equations of motion

The Hamiltonian equations of motion follow from a variation of the Hamiltonian. For lapse and shift, we just get the trivial statement

$$\dot{N}^\alpha = \{N^\alpha, \mathrm{H}^P\} = v^\alpha \ .$$

This confirms the previous statement that they are superfluous degrees of freedom that can be freely specified.

Variation with respect to three-metric and momentum yields

$$\dot{p}^{ab} = \{p^{ab}, \mathrm{H}^P\} = -\frac{\delta \mathrm{H}^P}{\delta h_{ab}} \ , \tag{2.11}$$

$$\dot{h}_{ab} = \{h_{ab}, \mathrm{H}^P\} = \frac{\delta \mathrm{H}^P}{\delta p^{ab}} \ . \tag{2.12}$$

N and N^a thus enter as arbitrary functions. Here it becomes obvious that they have to be prescribed in order to give meaning to the equations of motion. We also see that N, N^a specify the time parameter t, i.e. the foliation. The freedom to choose N, N^a arbitrarily insures that the theory does not depend on a specific foliation — it holds for any foliation.

The equations of motion (2.12) for h_{ab} can be solved for $p^{ab} = p^{ab}(\dot{h}_{cd})$. Using the defining equation for p^{ab} in terms of extrinsic curvature, (2.6), one can express the extrinsic curvature in terms of the time derivative of the metric h — which just yields (2.3).

The equations (2.11) for $\dot{p}^{ab}$ are dynamical equations and complete the set of Hamiltonian equations of motion for general relativity.

2.4.2 Relation to Einstein's equations

We expect these equations in some sense to be equivalent to Einstein's equations as they arise from the ADM-action which in turn arises from a rewriting

of the Einstein–Hilbert action. Due to the constraints, some of the Hamiltonian equations should be redundant. Both expectations are met. First of all, the following equivalence can be shown

> **From Einstein to ADM:**
> Let a spacetime with metric, $(\mathcal{M}, g)$, be given, where g is a solution to Einstein's equations. Now foliate this spacetime with a one-parameter family of spacelike embeddings $\mathcal{E}_t : \Sigma \longrightarrow \mathcal{M}$ of a three-hypersurface Σ into $\mathcal{M}$. These embeddings define lapse and shift. Furthermore, we get a family of induced three-metrics $t \longmapsto h(t)$, and similarly for the momenta $t \longmapsto p(t)$. Here, p is defined in terms of the extrinsic curvature.
>
> It then follows that the pair $h(t)$ and $p(t)$ satisfies Hamilton's equations of motion, (2.11) and (2.12) and the constraints. It is thus shown that we obtain solutions to Hamilton's equations of motion and the constraints through the foliation of a spacetime obeying Einstein's equations, [8].
>
> **From ADM to Einstein:**
> On the other hand, let a spacetime with metric $(\mathcal{M}, g)$ be given — but let g be an arbitrary metric that is not a priori required to satisfy Einstein's equations. Carry out a foliation as described above. But now assume that $h(t)$ and $p(t)$ satisfy the Hamiltonian equations of motion and the constraints.
>
> It then follows that the spacetime metric g satisfies Einstein's equations. So, a solution to Hamilton's equations and the constraints is associated with a solution to Einstein's equations through whose foliation it arises, [8].

Actually, (2.11) and (2.12) are the projections of the vacuum Einstein equations into the surface: $G_{ab} = 0 \Leftrightarrow$ (2.11), (2.12), where $G_{\alpha\beta}$ is the Einstein tensor. And the constraints are its projections perpendicular to the hypersurface, namely

$$\mathcal{H}_\perp \approx 0 \quad \Leftrightarrow \quad n^\alpha n^\beta G_{\alpha\beta} = 0 \ , \qquad \mathcal{H}_a \approx 0 \quad \Leftrightarrow \quad n^\alpha G_{\alpha a} = 0 \ .$$

But we also suspected that some of the equations of motion are redundant due to the existence of constraints. This is the subject of the *interconnection theorems*, for a full list see [8]. I only quote the two which are important for the justification of the quantum scheme we will use.

The first statement was discovered by Dirac and can be found in his Lectures, [3]. It guarantees that the constraints are conserved by the equations of motion. More precisely, if the constraints and the equations of motion

for p^{ab} and h_{ab} are satisfied by a pair (p^{ab}, h_{ab}) for all points on $\mathcal{E}_t(\Sigma)$, then the constraints will be satisfied for all $\mathcal{E}_T(\Sigma)$ with $T > t$. This 'conservation law' is a consequence of the Bianchi identities.

Reversely, if the constraints are satisfied by a pair (p^{ab}, h_{ab}) on $\mathcal{E}_t(\Sigma)$ for all parameters $t \in \mathbb{R}$, i.e. on all hypersurfaces of a foliation, then the equations (2.11), (2.12) for p^{ab} and h_{ab} are automatically satisfied. This statement is due to Kuchar, [7].

From these two statements we learn that *the entire dynamics of the Hamiltonian theory is contained in the Hamiltonian constraints*. This is the reason why it is believed to be sufficient to quantize these constraints alone.

2.5 Dirac algebra

I want to come back to the algebra generated by the secondary constraints, the Dirac algebra. It characterizes the transformations generated by the secondary constraints.

Before discussing several interpretations of this algebra, note that the Dirac algebra is not a Lie algebra[6] and consequently, the generators cannot be exponentiated to form a group. However, the $\mathcal{H}_a(x)$ alone do form a Lie algebra, namely the algebra of three-diffeomorphisms in the hypersurface.

Last but not least it should be noted that the signature of the embedding spacetime metric can be read off from this algebra. If the right-hand side of (2.8) is positive, the spacetime metric has signature $(-,+,+,+)$, whereas a negative sign on the right-hand side comes from a Euclidean four-metric.

2.5.1 Reparametrization-invariant theories

It was already observed by Dirac that any diffeomorphism-invariant theory necessarily has weakly vanishing Hamiltonian.

Theories with weakly vanishing Hamiltonian are called *reparametrization invariant*.[7] Any theory can be brought into such a form. To this end, one has to introduce an unphysical 'time' parameter that elevates physical time to a dynamical variable. The resulting theory is then obviously invariant under reparametrizations of the unphysical 'time' parameter. This procedure can of course be carried out with any of the coordinates. The resulting formulation is then invariant under reparametrizations of the artificially introduced parameters. But general relativity *is* already invariant under 'reparametrizations' of coordinates. This is just the essence of general covariance, or diffeomorphism invariance. Thus, general relativity is

[6] Note that in the first bracket, (2.8), the indices are raised using the three-metric on the right-hand side.

[7] In the literature one also finds the name *parametrization invariant*. But to my understanding, the theories are parametrization *independent* and therefore reparametrization *invariant*.

naturally in reparametrization invariant form. Other such theories are the relativistic free particle and the free bosonic string.

Irrespective of whether or not the theory has been artificially turned into a reparametrization-invariant one or is reparametrization invariant from the very beginning, the canonical Hamiltonian will always be of the form

$$\mathrm{H} = v_\mu \mathrm{H}_\mu \ , \qquad \mathrm{H}_\mu \approx 0 \ .$$

Decomposing H_μ into hypersurface-orthogonal and -tangential parts, one always arrives at the Dirac algebra, (2.8) — (2.10), see [3].

2.5.2 The hypersurface-deformation algebra

It has been shown by Kuchar, Hojman and Teitelboim that — for artificially as well as naturally reparametrization-invariant theories — the Dirac algebra can be interpreted as the algebra of hypersurface deformations. Here, the H_a generate the deformations in the hypersurface which are just equivalent to a change of coordinate system in the hypersurface. And $\mathrm{H}_\perp$ generates the deformations perpendicular to the hypersurface and thus the dynamics.[8]

In this way, the Dirac algebra can be derived from first principles, [12]. In spirit, this is however very close to the way Dirac himself interpreted the algebra in his Lectures, [3].

2.5.3 Relation to diffeomorphisms

The question is now how the transformations generated by the Hamiltonian constraints relate to invariance transformations on the Lagrangian level.[5] This issue has been studied by Bergmann and Komar, [14]. First of all, note that a hypersurface deformation is equivalent to the transformation of one coordinate system specifying a spacelike hypersurface into another one describing the same hypersurface. Generators of such a transformation would just be the generators of four-diffeomorphisms. But due to the canonical form and the projection tangential and orthogonal to the hypersurface, we are here not dealing with the generators of four-diffeomorphisms. This can be immediately argued from the fact that such a projection is not unique. Therefore the set of hypersurface deformations is larger than the set of four-diffeomorphisms. This argument is made precise by Bergmann and Komar, [14]. They showed that general relativity (i.e. the Einstein–Hilbert action) is not only invariant under four-diffeomorphisms

$$x^\mu \longrightarrow \quad x^{\mu\prime} = f^\mu(x) \ , \tag{2.13}$$

[8] It is also a result of Kuchar's work that the Hamiltonian constraints of artificially reparametrization-invariant theories depend linearly on the canonical momenta whereas those of the truly reparametrization-invariant theories have quadratic dependence on the momenta, [9].

but under the larger set of transformations

$$x^\mu \longrightarrow \quad x^{\mu\prime} = f^\mu[x, g_{\rho\sigma}(x)] \ . \tag{2.14}$$

How do these two groups relate to the transformations generated by the Hamiltonian constraints? First of all, note that (2.13) is a subgroup of (2.14). The Dirac algebra is a proper subalgebra of the Lie algebra of (2.14) only if N^α depend on the canonical variables h and p. The group generated by (2.13), on the other hand, is a subgroup of the group generated by $\mathcal{H}_a$ — but only for spatial diffeomorphisms.

2.6 Gravity coupled to matter

If we go beyond vacuum gravity, the matter action has to be added to the Einstein–Hilbert action,

$$S = S_{\rm EH} + \int\limits_{\mathcal{M}} \mathrm{d}^4x \, \sqrt{-\det g} \, \mathcal{L}_{\rm matter} \ .$$

The decomposition of the Einstein–Hilbert part was carried out in the previous sections. Now we also have to decompose the matter Lagrangian. Essentially, we have to express the metric determinant as $\det g = -N^2 \det h$ and split derivatives into timelike and spatial ones. Note that because the matter Hamiltonian describes the matter density, what we have to do, is to decompose the energy-momentum tensor into its timelike and spatial parts and express the resulting quantities in terms of the field and its momenta.

The total Hamiltonian is likewise constrained to vanish. Tangential and orthogonal components include now

$$\mathcal{H}_\perp^{\rm matter} = N\sqrt{\det h}\,\rho \ , \qquad \mathcal{H}_a^{\rm matter} = N\sqrt{\det h}\, j_a \ ,$$

where ρ and j_a are energy density and energy-flow components, respectively.

References: I heavily relied on Isham's presentation of the (3 + 1)-decomposition given in 'Canonical Quantum Gravity and the Problem of Time', [11]. The relation between the canonical equations of motion and Einstein's equations is explained in Kuchar's papers on canonical quantum gravity, [8] as well as in a review by Isham, [10]. I also used Kuchar's lecture notes from 1976, [7], as well as the review article of Bergmann and Komar in the first volume of 'General Relativity and Gravitation', [13].

Chapter 3

Dirac Constraint Quantization

The road to quantum general relativity is a rather bumpy one — and the final destination is still not in sight. Different authors blame different problems for the deprivations of the enduring voyage. Moreover, with the advent of loop quantum gravity, a shortcut with speedy vehicle is claimed to exist, putting someone chugging in a rusty old car along a run-down road into a position of defence. I therefore want to get out of the car and walk the road to quantum general relativity step by step, trying to separate the problems of the path (Dirac constraint quantization) from those of the vehicle (general relativity), trying to identify those potholes which are natural and those which one should be able to avoid. The present chapter is entirely dedicated to the path, i.e. Dirac's constraint quantization scheme.

3.1 Canonical Quantization

3.1.1 Canonical Quantization of Classical Mechanics

Canonical Quantization is a scheme originally devised for the quantization of classical, mechanical systems with simple phase-space structure $\mathbb{R}^{2n}$. It was mainly developed by von Neumann, Weyl and Dirac. The idea is to give a clear prescription of how to pass from a given classical system to its quantum counterpart. In such a way one tries to axiomatize the heuristic steps that led to the discovery of the Schrödinger equation. Such an effort results in five axioms, given in Appendix B. Their essence is the following instruction: promote phase-space functions to operators on a Hilbert space of square integrable functions over configuration space such that all Poisson

bracket relations are turned into commutators. Schematically, this can be written as

$$f \longmapsto \hat{f} \qquad \text{such that} \qquad \{\cdot,\cdot\} \longmapsto -\frac{1}{\mathrm{i}\hbar}\,[\cdot,\cdot] \ .$$

The problem with this system of axioms is that it is inconsistent, see Appendix B. Moreover, the Groenewold–van Hove theorem shows that not all phase-space functions can be elevated to operators such that the Poisson algebra is preserved in the form of a commutator algebra. Another problem is that not all phase spaces have the simple structure $\mathbb{R}^{2n}$. It would be desirable to generalize the quantization procedure to systems whose phase space is a general symplectic manifold $\mathbf{\Gamma}$. Both of these problems have been tackled in various ways.

I just want to emphasize here that *the canonical quantization scheme does not provide a consistent, unique, axiomatic prescription.* One of the major draw-backs is that it does not tell us which variables have to be implemented under preservation of their Poisson algebra. This is a direct consequence of Groenewold–van Hove's work and will be of some importance in the quantization of gravity.

Dynamics in canonical quantization in the Schrödinger picture is given by the Schrödinger equation. According to it, any state vector, i.e. element of the Hilbert space, evolves in external time t through the Hamiltonian operator,

$$\mathrm{i}\hbar\frac{\partial}{\partial t}|\Psi(t)\rangle = \hat{H}|\Psi(t)\rangle,$$

where $\hat{H}$ is the operator corresponding to the system's Hamiltonian.

3.1.2 Canonical Quantization of Field Theories

Generalizing the canonical scheme to infinitely many degrees of freedom yields the so-called *functional Schrödinger picture of quantum field theory.* Essentially, the rules of the theory with finite degrees of freedom are taken over to field theory. This is done, of course heuristically, through a discretization of fields — thus truncating the field theory at finite orders. After employment of the quantization rules, the continuum limit to infinitely many degrees of freedom is taken. For details, see Appendix B.

The functional Schrödinger picture is a generalization of the usual Fock-space representation. The latter is tied to the particle interpretation, whereas the former also applies on curved hypersurfaces — which are just the type of hypersurfaces we are dealing with in canonical general relativity.[1]

[1]For a free field theory in Minkowski space, one usually uses the Fock space representation. It is invariant under Poincaré transformations, that means 1) the vacuum is invariant under action of the Poincaré group, 2) creation and annihilation operators

Despite its intuitive nature, there are some subtleties involved in the setup of the functional Schrödinger picture. For example, it is not obvious to which function space the canonical field should belong, i.e. it is not clear what the correct choice of configuration space $\mathcal{Q}$ shall be. Related to this is the fact that an infinite analogue of Lebesgue measure does not exist. In fact, for a non-interacting scalar field, one has to choose a Gaussian measure μ. It turns out that as configuration space not only functions but also distributions have to be included, [10].

A more serious problem are the divergences arising from the fact that one is dealing with infinitely many degrees of freedom. The products of non-commuting operators taken at the same spacetime point are generally ill-defined as they produce $\delta(0)$-divergences. In a flat background, one uses operator ordering to get rid of these infinities.

The divergences stemming from the products of non-commuting operators arise on the same level as the UV-divergences in the path-integral. Renormalization of these is generally done through the introduction of a regulator. In the end, the regulator is removed after the potentially divergent terms have been rendered finite. If the divergences cannot be dealt with, the theory is meaningless — at least it is not valid as a *fundamental* theory.[2]

So note that *infinities associated with operator products at the same space point arise in any quantum field theory in the functional Schrödinger picture*. A particular problem of quantum *gravity* is the perturbative non-renormalizability of general relativity and the lack of background structure.

Analogous to the finite-dimensional case, one arrives at the representation

$$\hat{\phi}(x,t)\Psi\left[\phi\right] = \phi(x,t)\cdot\Psi\left[\phi\right] \ , \qquad \hat{\pi}(x,t)\Psi\left[\phi\right] = -\mathrm{i}\hbar\frac{\delta\Psi\left[\phi\right]}{\delta\phi(x,t)}$$

for fundamental variables ϕ, π, where $\Psi\left[\phi\right] \in L^2(\mathcal{Q},\mu)$. They satisfy the equal-time commutation relations

transform covariantly amongst themselves and 3) an n-particle state is mapped onto an n-particle state under the action of the Poincaré group. This can be summarized by the statement that Fock space carries a unitary representation of the Poincaré group. This does no longer hold if the theory acquires an interaction term. Then Poincaré invariance is generally broken which means that a particle interpretation does no longer exist. The same effect occurs if one considers field theory on curved instead of flat surfaces. From this we see that the Schrödinger representation is a more fundamental concept than the Fock space representation. The latter is tied to the particle interpretation whereas the former incorporates the field aspects.

[2]But as general relativity does not couple to fermions, one might expect that its quantization cannot yield a fundamental theory anyway.

$$\begin{aligned}\left[\hat{\phi}(x,t),\hat{\phi}(y,t)\right] &= 0\ ,\\ [\hat{\pi}(x,t),\hat{\pi}(y,t)] &= 0\ ,\\ \left[\hat{\phi}(x,t),\hat{\pi}(y,t)\right] &= \mathrm{i}\hbar\,\delta^3(x-y)\end{aligned}$$

Dynamics of the system is given here as well through the (functional) Schrödinger equation,

$$\mathrm{i}\hbar\frac{\partial}{\partial t}|\Psi[\phi]\rangle = \hat{H}|\Psi[\phi]\rangle\ ,$$

where $\hat{H}$ now contains functional derivatives, e.g. in the position representation with respect to the field $\phi(x,t)$ for fixed t.

3.2 Constraint Quantization

Any theory with local gauge invariance is constrained on the Hamiltonian level. Even though we did not obtain the ADM-action through a gauging of a rigid symmetry, i.e. even though we are not dealing with a gauge theory of general relativity, we still encounter constraints. These are due to the diffeomorphism invariance of general relativity. More generally, we saw that any reparametrization-invariant theory obeys a constrained Hamiltonian dynamics.

So in order to quantize reparametrization-invariant theories — as well as gauge theories — canonically, we need a prescription for the implementation of constraints in the functional Schrödinger formulation of quantum field theory.[3]

3.2.1 Dirac Constraint Quantization

There are several ways in which one could try to realize such a formulation. For a given system with a set of first-class constraints, the following approaches towards quantization can be found in the literature:

1. One can solve the constraints on the classical level and quantize only the true degrees of freedom. This quantization method is also referred to as *true canonical quantization*, [10]. But often the true degrees of freedom are only known implicitly through the constraint equations. In general relativity, for a general solution to Einstein's equations, it is not known *which* the true degrees of freedom are. Several attempts have been made but all in one or the other way failed due to the complexity of the constraints.

[3] Diffeomorphism-invariance is just a special case of reparametrization-invariance as they are called in Chapter 2 and Appendix A.

2. One can gauge-fix the freedom generated by the constraints. This is done by adding gauge-fixing conditions until no gauge freedom is left. This turns the first-class into a second-class system. These second-class constraints can then be implemented as operator identities, provided that Dirac brackets instead of Poisson brackets are used to derive commutators.

   When applied to general relativity, this procedure faces the same problems as do Yang–Mills theories. No global gauge choices are possible and one has to be content with local coordinate conditions.

3. A third possibility is to interpret the constraint operators as restrictions on the quantum state space. This is done via the requirement that they annihilate *admissible states*. I use the term *admissible* instead of *physical* as additional conditions may arise that prevent some admissible states from being physical.

Another condition that points nonetheless in the same direction as (3.) is the requirement that the matrix elements of the constraint operators with respect to admissible states do vanish. This criterion presupposes that the spectrum of the constraints is known before admissible states can be identified. The impracticability of this procedure is surely the reason why I found this criterion solely in the book by Sundermeyer, [23].

In any case, (2.) or (3.), the quantum state space should not be more restricted than the classical state space. This poses restrictions on the commutators of constraints.[4] So in case (3.), the constraint algebra has to annihilate all admissible states. Case (3.) is referred to as *Dirac constraint quantization*.

Note that so far, these quantization procedures are, apart from the first one, only suggestions. They do not follow from some principle. It remains to be proven that they yield the correct results. Correct results here mean that the canonical constraint quantization has to yield results which are in accord with experiment for those theories where experiments exist or simply coincide with the established path-integral quantization for those theories where path-integral results are trusted.

In general relativity, a mixture of the above recipes is used. The primary first-class constraints $p^{\mu 0} \approx 0$ are solved on the classical level. Lapse and shift are specified before quantization. The remaining, secondary constraints which encode the dynamics, are subject to Dirac constraint quantization. The two axioms of this quantization scheme (3.) are

(q6) The classical first-class constraints are promoted to operators in such a way that the classical Poisson algebra of the constraints is preserved

[4] Actually, it is this restriction that requires second-class constraints to be implemented as operator identities.

as commutator algebra of the constraint *operators* with the constraint operators ordered to the *right* of the operators of structure functions.

(q7) The thus constructed operators select all physically admissible states through

$$\hat{K}_\alpha \Psi = 0 \ ,$$

for all classical first-class constraints K_α and $\Psi \in \mathcal{F}$, where $\mathcal{F}$ is the representation space of the operators.

I will use simply (q6) and (q7) whenever I refer to the respective requirement. Condition (q7) will also be referred to as *Dirac's constraint quantization condition*.[5]

3.2.2 Dirac's constraint quantization condition: an example

Once we accept (q7), (q6) is a logical consequence. But the constraint quantization condition (q7) itself is harder to justify. It is not obvious that it should yield the same result as (1.).

That is why I want to present an example that shows how and that the above prescription (q7) works. Despite this example, there are other cases in which Dirac constraint quantization does *not* yield the same result as the path-integral method.

A strong support in favour of Dirac constraint quantization in the special case of general relativity, however, is provided by its equivalence to the path-integral, see Chapter 4.

The free, parametrized non-relativistic particle

This example is due to Komar, [25]. I quote it here, because it exhibits the general structures and problems arising in the quantization of a reparametrization-invariant system. As general coordinate invariance is a form of reparametrization invariance, this is of particular interest for the quantization of any general-relativistic theory.

Classical theory On the classical level, we are dealing with a particle of mass m, described by canonical coordinates q^μ and momenta p_ν. These satisfy Poisson-bracket relations $\{q^\mu, p_\nu\} = \delta^\mu_\nu$.

The system is subject to a constraint $K = p_0 + \frac{p^a p_a}{2m}$ which arises because the time-coordinate has been turned into a dynamical variable through the introduction of an unphysical time-parameter, $q^0 = q^0(\tau)$. The system is invariant under reparametrization of τ. The constraint restricts dynamics to a hypersurface of phase space. It also generates transformations of phase-space variables in the constraint surface

[5] The axioms (q1) — (q5) can be found in Appendix B.

$$q^\mu \quad \to \quad q^\mu + \epsilon\delta q^\mu = 0 \ , \qquad p_\mu \quad \to \quad p_\mu + \epsilon\delta p_\mu = 0 \ ,$$
$$\text{where} \qquad \delta q^\mu = (\delta^\mu_0 + \delta^\mu_i \frac{p^i}{m}) \ , \qquad \delta p_\mu = 0 \ .$$

One recognizes that these are just the classical trajectories.

Observables are those quantities which commute (weakly) with the constraint $\{A, K\} \approx 0$, see Chapter 4, Section 4.3.2 and Appendix A. An over-complete set of observables is e.g. given by the components of linear momentum p_ν, orbital angular momentum $q_i p_j - q_j p_i$ and $q_i - \frac{p_i}{m} q_0$. These observables can be used to label the trajectories. As the trajectories are one-dimensional on the seven-dimensional constraint hypersurface, a minimum of six observables is needed to characterize a trajectory uniquely. Under the action of observables, trajectories are mapped upon trajectories. Note that the action of the constraint (namely along the trajectory) is perpendicular to the action of all observables (namely from trajectory to trajectory). Thus the induced symplectic structure on the constraint surface is singular.[6]

Quantum theory On the quantum level, the theory is described by operators $\hat{q}^\mu$, $\hat{p}_\nu$ with only non-trivial commutator $[\hat{q}^\mu, \hat{p}_\nu] = \mathrm{i}\hbar\,\delta^\mu_\nu$. These are realized on a Hilbert space of states $|\Psi\rangle$. Out of this space, the constraint projects a hypersurface of admissible states via $\hat{K}|\Psi\rangle = 0$. In the Schrödinger representation, this equation turns out to be just the time-dependent Schrödinger equation.

The first point of irritation is that *the constraint cannot be realized as a self-adjoint operator*. If it could, one would arrive at the following paradox conclusion

$$\begin{aligned} \delta\langle\Psi|\hat{q}^i|\Psi\rangle &= \langle\Psi|\left[\hat{q}^i, \hat{K}\right]|\Psi\rangle \\ &= \frac{\mathrm{i}\hbar}{m}\langle\Psi|\hat{p}^i|\Psi\rangle = 0 \ , \end{aligned}$$

as the constraint annihilates the bra as well as the ket, if we take the expectation value in an admissible state. As the constraint thus cannot be represented by a *self-adjoint* operator, consider the action of a state under the adjoint $\hat{K}^\dagger$,

$$\delta|\Psi\rangle = -\frac{\mathrm{i}}{2}\epsilon\hat{K}^\dagger|\Psi\rangle \ .$$

[6] This can be rigorously proven using the Lagrange bracket. One can show that the Lagrange bracket vanishes and thus a symplectic structure on the constraint surface is singular.

If $\hat{K}$ is a normal operator, i.e. if the inner product is chosen such that $\left[\hat{K}, \hat{K}^\dagger\right] = 0$, *the adjoint can be interpreted as generating the gauge transformations in the quantum theory.* This can be seen from the fact that the transformation $|\Psi\rangle \longmapsto |\tilde{\Psi}\rangle = |\Psi\rangle + \delta|\Psi\rangle$ for $\hat{K}|\Psi\rangle = 0$

- preserves the inner product of admissible states: $\langle\tilde{\Psi}|\tilde{\Psi}\rangle = \langle\Psi|\Psi\rangle$,
- preserves the expectation values of observables if these are defined via $\left[\hat{A}, \hat{K}\right]|\Psi\rangle = 0$ for each admissible state: $\langle\tilde{\Psi}|\hat{A}|\tilde{\Psi}\rangle = \langle\Psi|\hat{A}|\Psi\rangle$,
- preserves admissible states, i.e. leaves the admissible state space (the 'quantum constraint hypersurface') invariant: $\hat{K}|\tilde{\Psi}\rangle = 0$.

This is a first step towards a classical-to-quantum correspondence.

The second point that deserves special mention is that the inner product on the quantum constraint hypersurface is singular. As a consequence, *the inner product of states is at best indefinite and can become negative.* Singularity on the quantum constraint hypersurface follows from the fact that $\hat{K}^\dagger|\Psi\rangle$ is orthogonal to every admissible state. This is the quantum analogue of the fact that the symplectic structure on the classical constraint surface is singular.

Komar further discusses the quantum-to-classical correspondence. He takes offence in the fact that for any non-linear constraint operator, as e.g. $\hat{K}$, expectation values do not satisfy the classical constraint equations. Therefore, he argues, we do not recover Ehrenfest's correspondence principle for the constraints. Driven by this critique, he introduces the concept of *quasi-observables.* These are operators that commute with $\hat{K} - \hat{K}^\dagger$. This is his attempt at an interpretation of the constrained quantum theory.

The parametrized, free, non-relativistic particle exhibits already all the problems related to Dirac's constraint quantization condition.

First of all, the measure includes integration over all q^μ, also the unphysical q^0. This is a restriction on the Hilbert space on which the theory is set up. Furthermore, the theory contains unphysical states and therefore has no positive definite inner product. Moreover, the Hamiltonian should not be represented as a self-adjoint operator on these states. Lastly, not all operators $\hat{q}^\mu$, $\hat{p}_\mu$ should correspond to self-adjoint operators. Komar distinguishes the relevant operators through the concept of a quasi-observable.

The parametrized, free particle is a simple system not only in the mathematical sense. But also because we *know* what is actually going on. We know that q^0 is no dynamical variable and can therefore interpret the theory accordingly. General relativity, on the other hand, is reparametrization invariant by nature and thus we *do not know* which are the dynamical variables.

3.3 Summary

From all this, one can see that the canonical quantization scheme does not provide a unique, axiomatic prescription that turns a classical into a quantum theory. We will list the encountered ambiguities and problems hierarchically, starting from the general quantization problems, via the specific problems encountered in the quantization of field theory, to the difficulties arising for constrained field theories, ending with the specific problems of reparametrization-invariant theories.

We will link each problem to its counterpart in the covariant approach.

Most generally, as a consequence of the Gronewold–van Hove theorem, we do not know which Poisson bracket relations to enforce as commutator relations. Thus we are deprived of a means that tells us how to quantize more complicated phase-space functions. This problem is known as *factor-ordering problem*. On the covariant level, it is reflected in the ambiguity in the definition of the path-integral measure. Namely, one has to choose a specific slicing of time in the construction of the path-integral measure. Different choices correspond to different factor orderings of the theory's Hamiltonian.

In going from ordinary quantum mechanics to field theory, one automatically encounters divergences related to operator products at the same space point — in the canonical formulation *on curved surfaces*. Because in this case, we can no longer take refuge to the Fock-space representation but have to use the functional Schrödinger picture. The divergences here arise on the same level as UV-divergences in the path-integral approach and should be removed through one of the familiar renormalization schemes.

But we are not just trying to quantize an arbitrary field theory, but we are trying to quantize a *constrained system*. This entails a whole series of alterations, if not complications, compared to the quantization of an unconstrained system. The difference to the unconstrained case becomes particularly severe as, in general relativity, we cannot easily and unambiguously isolate the true degrees of freedom. Therefore we are led to Dirac's constraint quantization condition. The application of this condition entails that not only the physical, but *all* degrees of freedom are promoted to operators. As a consequence, a positive inner product cannot be defined. From the point of view of the path integral, the arisal of negative norm states is necessary to cancel the contribution of unphysical states to the overall transition amplitude. The corresponding procedure on the canonical level would be to add gauge-fixing conditions on the classical level and quantize the resulting second-class system.

We saw in the example of the parametrized, free particle that the resulting theory is very difficult to interpret.

But general relativity is not simply governed by constraints, here the Hamiltonian itself is constrained to vanish. This is the case for any reparame-

trization-invariant theory. For a theory of this type, we expect that the Hamiltonian cannot be defined as a self-adjoint operator.

I emphasize again that, apart from the last problem, all other problems occur in the canonical quantization of *any* constrained field theory. Moreover, these problems are well-known problems on the covariant level where *in principle* methods exist that allow to cope with them — albeit on a very formal level.

References:

As sources for the general scheme of canonical quantization served the article by Giulini [18] and the review on quantization methods written by Ali and Englis, [17]. The methods of constraint quantization are described in Sundermeyer's book, [23].

Chapter 4

Quantum General Relativity

The application of the canonical quantization scheme is referred to as quantum general relativity. This application is not without flaws. Therefore, I will outline quantum general relativity as a programme. This programme still has mathematical as well as interpretational problems. Despite these, the theory has some successes to its name. These will be described in the last section of this chapter.

4.1 The Programme

Now we will apply the just presented 'quantization scheme' to general relativity. We put 'scheme' in quotation marks to remind the reader that we are not using an axiomatic system that is *guaranteed* to yield the correct quantum theory.

4.1.1 The quantization map Q

As fundamental variables serve the three-metric h and its canonically conjugate momentum p. These are represented by operators

$$h_{ab}(x) \longrightarrow \hat{h}_{ab}(x) \ , \quad p^{ab}(x) \longrightarrow \hat{p}^{ab}(x)$$

such that

$$\begin{aligned} \left[\hat{h}_{ab}(x), \hat{p}^{cd}(y)\right] &= \mathrm{i}\hbar\, \delta^{(c}_{(a}\delta^{d)}_{b)}\delta(x-y) \ , \\ \left[\hat{h}_{ab}(x), \hat{h}_{cd}(y)\right] &= 0 \ , \\ \left[\hat{p}^{ab}(x), \hat{p}^{cd}(y)\right] &= 0 \ . \end{aligned}$$

We use the Schrödinger representation as is required by the axioms of the canonical quantization scheme, see (q4) in Appendix B,

$$\hat{h}_{ab}(x)\Psi[h] = h_{ab}(x)\cdot\Psi[h]\ , \quad \hat{p}^{ab}(x)\Psi[h] = -\mathrm{i}\hbar\frac{\delta\Psi[h]}{\delta h_{ab}(x)}\ . \tag{4.1}$$

Here, $\Psi[h]\in\mathcal{F}$ is the state vector. It can be any functional of the metric sufficiently well-behaved to allow differentiation but not restricted otherwise. Especially, it does not need to be square-integrable with respect to h. So the *representation space* $\mathcal{F}$ is the vector space of all functionals over configuration space $Q(\Sigma)$.

4.1.2 The Dirac Algebra

A minimalistic approach requires the realization of the classical Hamiltonian constraints as operators on $\mathcal{F}$. This realization is restricted by the condition that commutators of constraints shall not generate new constraints. This limits the choice of factor ordering of the constraints. Namely, it has to be chosen such that metric operators occur to the *left* of constraints operators in the operator algebra corresponding to (2.8) — (2.10).

The search for such an ordering is beset with problems and an open issue. The general ordering of the constraint operators, containing parameters α, β and γ labelling the ambiguities in the ordering of the operators, is given by

$$\begin{aligned}
\hat{\mathcal{H}}_a(x) &= -2\left[\hat{h}_{ab}(x)\hat{p}^{bc}{}_{,c}\,(x) + \hat{\Gamma}_{abc}(x)\hat{p}^{bc}(x)\right] + \alpha\delta_{,a}\,(0)\ ,\\
\hat{\mathcal{H}}_\perp(x) &= -2\kappa^2\frac{1}{\sqrt{\widehat{\det h}}}\left[\frac{1}{2}\hat{h}_{ab}(x)\hat{h}_{cd}(x) - \hat{h}_{ac}(x)\hat{h}_{bd}(x)\right]\hat{p}^{ab}(x)\hat{p}^{cd}(x)\\
&+ \frac{\sqrt{\widehat{\det h}}}{2\kappa^2}\left({}^{(3)}\hat{R} - 2\Lambda\right)\\
&+ \beta\,\delta^3(0)\frac{1}{\sqrt{\widehat{\det h}}}\hat{h}_{ab}(x)\hat{p}^{ab}(x) + \gamma\left[\delta^3(0)\right]^2\frac{1}{\sqrt{\widehat{\det h}}}\ .
\end{aligned}$$

where $\delta^3(0)$ denotes the usual, three-dimensional Dirac delta-distribution evaluated on zero, [32], and Γ^a_{bc} is the Christoffel connection of h.

The problems related to the consistent representation of the constraints as operators in Section 4.2.3.

4.1.3 The solution space $\mathcal{F}_0$

Assume that we were in some way able to fix the ambiguities in the previous form of the constraint operators. Any admissible state then has to satisfy

$$\hat{H}_\perp\Psi[h] = 0\ , \quad \hat{H}_a\Psi[h] = 0\ . \tag{4.2}$$

We will denote the space of solutions to these equations *solution space* $\mathcal{F}_0$. The quantum analogue of the Hamiltonian constraint is called Wheeler–DeWitt equation. The quantum diffeomorphism constraint has no special name. Whether a non-trivial solution exists at all as well as the interpretation of the constraint, depends on the choice of ordering, again see Section 4.2.3.

4.2 Mathematical Problems

4.2.1 Positivity requirement for the three-metric h

The first step, (4.1), looks rather innocent but already at this level, difficulties arise. Namely, the metric h is not simply a symmetric tensor but a *metric*. It has to yield a positive norm for any vector field on Σ. If this geometric property of the metric is implemented on the quantum level, the operator $\hat{p}^{cd}$ is no longer (formally) self-adjoint (with respect to the naive Schrödinger measure, see Section 4.3.3), see [11]. A proposed solution to this problem is to generalize the canonical commutation relations to *affine* commutation relations, [24].

The general question behind this is which geometrical properties of operators should be implemented on the quantum level.

4.2.2 Factor ordering

Classically equivalent functions of phase-space variables are not equivalent on the quantum level when phase-space variables have been turned into non-commuting operators. This problem, known as *factor-ordering problem*, always arises if one quantizes a classical system. One has to find other conditions that fix the thus arising ambiguities. One such condition is that all Poisson-bracket relations be realized as commutators. But it is just the content of the Groenewold–van Hove theorem that this is not possible: either one has to limit the space of quantizable observables or one has to alter the commutator relations.

Kuchar emphasizes that the factor-ordering problem cannot be underestimated. In fact, the choice of factor ordering determines the quantum theory, [19]. Changes in the factor ordering produce order $\hbar$-corrections. Of course, the factor ordering is most severe in quantum gravity and there has the biggest impact on the factor ordering of constraints.

4.2.3 Consistent constraint operators

The constraint operators decide which state is to count as an admissible state in quantum gravity. Therefore, their representation decides about the physical content of the theory. More importantly, it depends on their

realization whether or not a (sensible) canonical theory of quantum gravity does exist at all.

As discussed above, the constraints, being non-linear in h and p, are subject to operator-ordering ambiguities. One can try to fix them through the following requirements:

1. **Consistency**: The constraints have to satisfy the Dirac algebra with the structure-function operators to the left. If such an ordering cannot be found, the theory does not incorporate general coordinate invariance and has to be dismissed.

2. **Regularity**: The constraints contain products of non-commuting operators at the same space point and thus have to be regularized. Such a regularization has to be in accord with the constraint algebra. Note that the corrections obtained from different regularization schemes are on equal footing to those produced by different factor orderings. (They are simply divergences.)

3. **Transformation properties**: The diffeomorphism constraints should be ordered such that they describe the transformation of the wave function under diffeomorphisms on Σ. Such a rule does not apply to the Hamiltonian constraint as it is not generator of a Lie algebra.

4. **Formal self-adjointness**: As was pointed out by Komar, [25], the Hamiltonian operator cannot be realized as a symmetric operator. The criterion of formal self-adjointness which was pursued in the early days of quantum gravity thus does *not* apply for this constraint — but it does, of course, for the diffeomorphism constraints.

If one searches the literature on this topic, one quickly comes to realize that the number of articles is inversely proportional to the importance of the problem. A first factor ordering was given by Anderson — but it did not satisfy (1.), [27]. To the contrary, he found that no symmetric factor ordering of the Hamiltonian constraint can exist such that (1.) holds, [30]. Irritatingly, Schwinger claimed that Anderson's orderings were actually consistent with (1.), [31]. This discrepancy was cleared up independently by Tsamis and Woodard, [32], and Friedman and Jack, [33], more than twenty years later. I will follow the argument of Tsamis and Woodard. Namely they show that if the constraints are taken at face value and not regularized, one can achieve any result from formal manipulations of the operators in the Dirac algebra. This is due to the fact that in shifting the constraint operators to the left after the commutators were calculated, one has to use distributional identities which are not well-defined when acting on operator

products.[1]

It thus follows that the problem of consistency of the constraints is not even well-posed before the constraints are not regularized.

This problem is aggravated by the fact that general relativity is a perturbatively non-renormalizable theory. The regulator thus *cannot* be removed. Obviously, we cannot demand that the regulated theory has the same symmetry as the original one. Thus one can require closure of the constraint algebra only in a limiting sense for the regulated theory.

A rather simple regularization was proposed by DeWitt in 1967, [4]. He just advocated that operators at the same space point should commute. Then of course, consistency of the constraints holds if it holds on the classical level (which it does).

A mostly used *pragmatic* attitude is to use the Laplace–Beltrami factor ordering for the kinetic part of the Hamiltonian constraint. This ensures that this part is invariant under three-coordinate transformations.

Choosing the factor ordering such that all momenta stand to the right of the metric operators, the Wheeler–DeWitt equation and the diffeomorphism constraints become

$$\begin{aligned}\left[-2\kappa\hbar^2 G_{abcd}\frac{\delta^2}{\delta h_{ab}\delta h_{cd}} - \frac{\sqrt{\widehat{\det h}}}{2\kappa^2}\left({}^{(3)}R - 2\Lambda\right)\right]\Psi[h] &= 0\ ,\\ -2\frac{\hbar}{\mathrm{i}}\frac{\delta\Psi}{\delta h^{ab}{}_{|b}} &= 0\ .\end{aligned}$$

I will refer to this ordering as *naive* ordering. It is just the ordering that results from DeWitt's regularization condition.

4.3 Interpretational Problems

Even *if* we assume that we were able to find consistent, well-defined constraint operators and moreover solve the ensuing constraint equations — we would not be much better off. Because there is no agreement how to

[1] A very simple example of such manipulations is actually already given by Dirac, [34]. Differentiate the identity

$$x\delta(x) = 0 \tag{4.3}$$

to obtain

$$x\delta'(x) = -\delta(x)\ . \tag{4.4}$$

Multiplying (4.3) by $\delta'(x)$, one finds $x\delta(x)\delta'(x) = 0$. On the other hand, multiplying (4.4) by $\delta(x)$, we get $x\delta(x)\delta'(x) = -\left[\delta(x)\right]^2$, obviously a contradiction. Dirac ended his explications on the quantization of the gravitational field with this line and *never came back to this topic again.*

proceed: Neither do we know how to identify the physically relevant states in the solution space, nor do we know how to extract any predictions from the physical state space. One can distinguish the following aspects of the problem.

4.3.1 Identification of physical state space

The identification of the space of physical states is a difficult problem, quantum gravity shares with any constrained, canonical theory. This problem is due to the loss of Hilbert-space structure on $\mathcal{F}$ coming with the constraints.[2] But it is the existence of an inner product that establishes the equivalence between momentum and position space representation. This is thus lost and the theory depends on our choice of representation. Therefore, the space of solutions to the constraints, $\mathcal{F}_0$, depends on the representation.

There are additional conditions that prevent the identification of the solution space $\mathcal{F}_0$ with the physical state space. First of all, no boundary conditions have been imposed so far. In fact, there is no agreement on the conditions that should be imposed and so $\mathcal{F}_0$ is surely larger than the physical state space.

Several proposals for boundary conditions exist. In general, they are connected with the issue of singularity avoidance in quantum gravity. In principle, they try to remove singularities from the quantum framework. This is most obvious in DeWitt's boundary proposal, [4]. He requires $\Psi\left[{}^{(3)}\mathcal{G}\right] = 0$ for all three-geometries ${}^{(3)}\mathcal{G}$ that constitute a barrier to quantum evolution. The relation to singularity avoidance is equally obvious in the 'no-boundary' proposal of Hartle and Hawking, [39], which was raised in the cosmological context as a condition on the path-integral.

Apart from boundary conditions, further conditions might arise. For example, in the Klein–Gordon theory, only half of the solution space consists of physical states. A similar splitting has been tried in quantum general relativity, but cannot be carried out because no timelike Killing vector field exists in superspace. Also, there might be other restrictions necessary which we do not know of.

Thus, even if we *were* able to set up the theory as required, we would not able to extract the set of physical states — at least not with our today standard of knowledge.

[2] The inner product on $\mathcal{F}$ is used to define operators. As it contains unphysical degrees of freedom, it is not connected to measurement results. This Hilbert space on $\mathcal{F}$ is therefore called kinematical or auxiliary Hilbert space. This is opposed to the physical Hilbert space on $\mathcal{F}_{\text{phys}}$.

4.3.2 Observables

A huge problem in quantum gravity, or at least a point of diverging opinion, is the question of how to extract information from a (yet to establish) quantum theory of gravity.

So far, the only variables we have quantized, are the three-metric and its momentum, the Hamiltonian and diffeomorphism constraints. But if we want to measure other quantities, as e.g. volume, we have to introduce the corresponding operators. An ongoing debate is caused by the question as to which quantities might be observable. This problem has nothing to do with the quantization of gravity but arises already on the classical level.

The community is divided into two factions: one faction which claims that *observables* are those quantities which have vanishing Poisson brackets with all the constraints, whereas the opposite camp reserves the name *observables* for those quantities which commute with the diffeomorphism but not with the Hamiltonian constraint. I will in the following adopt the terminology of Kuchar and refer to quantities that commute with *all* constraints as *perennials* and those which only commute with the diffeomorphism constraints as *observables*.

The differing points of view arise from a different characterization and understanding of a physical state. Whereas the notion of *perennial* characterizes a state as an entire trajectory in phase space (labelled by its constants of motion), in the *observable* picture, a physical state is characterized by a point in phase space which corresponds to a fixed instant of time. So I do not see any contradiction between these definitions — they just refer to different concepts.

A problem for those sticking with perennials is that, so far, no perennials for general relativity are known, neither is it clear whether such quantities exist at all. To the contrary, it has been shown that *globally*, i.e. on the entire manifold $\mathcal{M}$, such perennials cannot exist, [29].

Whatever opinion one has on this point concerning the classical theory, the realization on the quantum level is given by

$$\begin{aligned} \left[\hat{A}, \hat{H}_\mu\right] \Psi &= 0 \ , \quad \text{for perennials and} \\ \left[\hat{A}, \hat{H}_i\right] \Psi &= 0 \ , \quad \text{for observables} \ , \end{aligned}$$

where Ψ is an element of the *physical state space*, $\Psi \in \mathcal{F}_{\text{phys}}$.

4.3.3 The physical state space $\mathcal{F}_{\text{phys}}$

A point of similar dissent is the Hilbert-space structure. In ordinary quantum field theory, a Hilbert space is used to define observables, namely as self-adjoint operators on that space, and their expectation values with respect to physical states. These are then interpreted as possible outcomes of

measurements. Thus the Hilbert space is the essential ingredient connecting quantum theory with experiment. It may seem therefore weird to call this structure into question. As excuse let it be said that this structure was not called into question voluntarily but challenged itself.

The Wheeler–DeWitt equation as it stands is a timeless equation. Our understanding of quantum theory, however, is tightly knit to the existence of a time variable. Hilbert spaces define probabilities that are conserved *in time* and measurement outcomes *at fixed instants of time*. The difficulty to interpret such a timeless theory is called the *problem of time*. Attempts have been made to identify one of the configuration space variables as 'internal time'. But these did not yield consistent results so far. Note also that this concept clearly transcends our understanding of quantum theory where measurements always take place in a background of space and time (which here is lost, i.e. quantized).

Moreover, there are indications that we might not need such a Hilbert space structure. These are discussed in the following Section 4.4.1.

4.4 Successes of quantum General Relativity

4.4.1 The semi-classical limit

In quantum theory, different approximation schemes are used to determine what is then called the *semi-classical limit*. Here, I want to present one such scheme.

It is an expansion in terms of the Planck mass $m_{\rm Pl}$ in the limit $m_{\rm Pl} \to \infty$. Through this expansion, *quantum field theory on curved spacetimes can be derived from quantum general relativity*. This is an important and non-trivial result. It gives confidence in the correctness of the general scheme. But *it also bears upon the interpretation of the overall quantum gravity scheme*, as will be explained below.

A second approximation that will be used later on is the Born–Oppenheimer one, see Chapter 8. It is presented in Appendix E.

Derivation of quantum field theory in curved spacetimes

Quantum field theory in curved spacetimes can be recovered from the four quantum constraint equations, $\hat{\mathcal{H}}_\mu \Psi = 0$, through an expansion in the Planck mass in the limit $m_{\rm Pl} \to \infty$. The underlying assumption is that all other masses in the scheme are masses of fundamental particles and thus much smaller than the Planck mass.

Furthermore, one has to assume that the wave functional is complex and consists just of one branch given by

$$\Psi[h, \phi] = e^{\frac{\rm i}{\hbar} S[h,\phi]} \ , \tag{4.5}$$

where ϕ is a minimally coupled scalar field representing a matter degree of freedom. The form (4.5) is a suitable approximation if the two branches $\Psi[h,\phi] = e^{\frac{\mathrm{i}}{\hbar}S[h,\phi]}$ and $\Psi[h,\phi] = e^{-\frac{\mathrm{i}}{\hbar}S[h,\phi]}$ decohere from each other. That this is indeed the case, at least for the first order $S = S_0 \in \mathbb{R}$, has been shown only for a minisuperspace model, [38].

Moreover, the matter content just consists of a single field. Fermionic matter, however, cannot be taken into account as it cannot consistently be coupled to general relativity already on the classical level.

If these conditions are satisfied, the expansion scheme is completely analogous to the expansion of the free Klein–Gordon equation in the velocity of light c in the limit $c \to \infty$. From the thus obtained (special) relativistic correction terms to the one-particle Schrödinger equation, relativistic corrections to the energy (as expectation value of the corrected Hamiltonian) can be calculated. This can be done despite the fact that the Klein–Gordon equation itself has no well-defined interpretation as a one-particle equation. But see [37] for more details on the non-relativistic limit of the Klein–Gordon equation.

Similarly, predictions in terms of corrections to the energy of the matter content can be derived from the Wheeler–DeWitt equation in the $m_{\mathrm{Pl}} \to \infty$ limit, *despite* the fact that the full equation cannot be given a sensible interpretation due to a lacking Hilbert space.

Note that the following calculations are formal in the sense that no regularization scheme has been employed to render the operator products finite. Furthermore, the factor-ordering problem has not been addressed. Rather, a specific ordering is chosen. In the end, the dependence of the results on the factor ordering should be discussed. The factor ordering in the Wheeler–DeWitt equation is the following

$$\left[-2\hbar^2\kappa^2 G_{abcd}\frac{\delta^2}{\delta h_{ab}\delta h_{cd}} - \frac{1}{\kappa^2}\sqrt{\widehat{\det h}}\left({}^{(3)}R - 2\Lambda\right) + \hat{\mathcal{H}}_m\right]\Psi[h,\phi] = 0 \ , \tag{4.6}$$

where $\mathcal{H}_m$ is the scalar-field Hamiltonian. So basically, we work with the DeWitt-regularized Wheeler–DeWitt equation. Introducing the parameter

$$M = \frac{1}{\kappa^2} = \frac{1}{8\pi\hbar}m_{\mathrm{Pl}}^2 \tag{4.7}$$

as expansion parameter, one can rewrite the Wheeler–DeWitt equation as

$$\left[-\frac{\hbar^2}{2M}G_{abcd}\frac{\delta^2}{\delta h_{ab}\delta h_{cd}} + MV[h] + \hat{\mathcal{H}}_m\right]\Psi[h,\phi] = 0 \ ,$$

where $V[h] = -\frac{1}{\kappa^2}\sqrt{\det h}\left({}^{(3)}R - 2\Lambda\right)$ was introduced as a short-hand for the potential. As listed above, a first assumption is that the solution is of the form (4.5) where

$$S[h] = MS_0 + S_1 + M^{-1}S_2 + \dots \ .$$

At the highest order, $\mathcal{O}(M^2)$, one thus derives

$$\left(\frac{\delta S_0}{\delta \phi}\right)^2 = 0 \ .$$

From this follows that S_0 depends only on the gravitational degrees of freedom. This is not the case if more than one matter field is present and one of these can have negative kinetic energy. Similarly, the above conclusion cannot be drawn if S_0 is not purely imaginary or real.

The interesting order is the next one. Here, at $\mathcal{O}(M)$, one arrives at the Hamilton–Jacobi equation of general relativity,

$$\frac{1}{2} G_{abcd} \frac{\delta S_0}{\delta h_{ab}} \frac{\delta S_0}{\delta h_{cd}} + V[h] = 0 \ . \tag{4.8}$$

This equation provides a classical spacetime satisfying Einstein's equations. Assume that a solution S_0 to the Hamilton–Jacobi equation (4.8) was found. The momentum can be read off from the Hamilton–Jacobi equation to be

$$p^{ab} = M \frac{\delta S_0}{\delta h_{ab}} \ . \tag{4.9}$$

Now insert on the left-hand side the defining equations for the gravitational momentum, (2.6) and (2.3). These make the connection to general relativity. On the right-hand side insert the solution to (4.8). Solve the resulting equation for $\dot{h}_{ab}$ which enters (4.9) through the definition of p^{ab},

$$\dot{h}_{ab} = 2N G_{abcd} \frac{\delta S_0}{\delta h_{cd}} + N_{(a|b)} \ . \tag{4.10}$$

To solve this equation, one has to specify lapse and shift. Having done so, an initial three-geometry $h_{ab}(t_0)$ will be evolved through (4.10) yielding a spacetime solving Einstein's equations.[3]

There are two observations to make at this point: First of all, as input one does not only need the initial three-geometry but also lapse and shift, i.e. one has to specify a foliation. As a result, and that is the second point worth mentioning, *a solution to Einstein's equation comes with a specific foliation in the semi-classical expansion scheme.* Lastly, as can be retraced using DeWitt's paper, [4], p. 1127, it is necessary that the diffeomorphism constraint is also satisfied at this order of the expansion scheme. Otherwise, one cannot prove that the resulting spacetime indeed satisfies Einstein's equations.

[3] That the spacetime obtained in this way is indeed a solution to Einstein's equations can be deduced from the calculations of DeWitt's paper [4], p. 1127.

Having thus recovered classical general relativity from the quantum framework, we can proceed to the next order, $\mathcal{O}(M^0)$. The equation at this order contains S_0 as well as S_1. If one rewrites

$$\chi[h,\phi] = D[h]e^{\frac{\mathrm{i}}{\hbar}S_1[h,\phi]} \ ,$$

where the prefactor D shall satisfy the 'conservation law'

$$G_{abcd}\frac{\delta}{\delta h_{ab}}\left(\frac{1}{D^2}\frac{\delta S_0}{\delta h_{cd}}\right) = 0 \ ,$$

one arrives at the following equation

$$\mathrm{i}\hbar\, G_{abcd}\frac{\delta S_0}{\delta h_{ab}}\frac{\delta}{\delta h_{cd}}\chi[h,\phi] = \hat{\mathcal{H}}_m\chi[h,\phi] \ .$$

This takes the form of a Schrödinger equation for the matter part (at each point in the three-hypersurface), if one introduces the time parameter $\tau(x;h]$ in the following way

$$\frac{\delta}{\delta\tau} = G_{abcd}\frac{\delta S_0}{\delta h_{ab}}\frac{\delta}{\delta h_{cd}} \ ,$$

i.e.

$$G_{abcd}\frac{\delta S_0}{\delta h_{ab}(y)}\frac{\delta\tau(x;h]}{\delta h_{cd}(y)} = \delta(x-y) \ .$$

Because we thus have a time-parameter at each point in the hypersurface, this parameter τ is also called many-fingered time, or, due to the way it is derived, WKB-time. The interpretation of this parameter becomes obvious if one contracts the equation of motion for h, (4.10), with $\frac{\delta}{\delta h_{ab}}$, yielding

$$\begin{aligned}\dot{h}_{ab}\frac{\delta}{\delta h_{ab}} &= 2NG_{abcd}\frac{\delta S_0}{\delta h_{cd}}\frac{\delta}{\delta h_{ab}} + N_{(a|b)}\frac{\delta}{\delta h_{ab}} \\ &= 2N\frac{\delta}{\delta\tau} + N_{(a|b)}\frac{\delta}{\delta h_{ab}} \ .\end{aligned}$$

Thus we obtain, using the definition of the extrinsic curvature, (2.2),

$$\frac{\delta}{\delta\tau} = K_{ab}\frac{\delta}{\delta h_{ab}} = n^\alpha\frac{\delta}{\delta\mathcal{E}^t_\alpha(x)} \ .$$

Thus the parameter τ has a clear interpretation: it labels the foliation of the spacetime. Note that $\tau(x;h]$ is determined by the three-geometry!

At higher orders, $\mathcal{O}(M^{-1})$, correction terms to the functional Schrödinger equation occur. Real correction terms imply an energy shift for the matter part. But imaginary corrections also occur, i.e. unitarity in τ is violated for the matter wave functional. From this perspective, unitarity violation

seems to be a rather natural consequence of the fact that the (approximate) Schrödinger equation takes account only of one part of the full wave functional, namely the matter part.

The semi-classical approximation as presented here thus underpins the Dirac quantization scheme as it is applied to general relativity: at least, classical general relativity can be recovered from this scheme. Moreover, it is possible to extract predictions through this approximation which — in principle — could be measured. In order to extract numerical values, of course, it is necessary to regularize the relevant quantities. Moreover, factor-ordering issues may change the results, even though the claim is that these are not altered by a reordering of the kinetic term of the Wheeler–DeWitt equation, [37].

Lastly, note that the DeWitt metric diverges for $\det h = 0$. At these points the Wheeler–DeWitt equation breaks down and no semi-classical time τ can be derived at this point anyway. But see the Chapter 5 for more details on this.

Interpretational consequences

We saw that the semi-classical approximation connects quantum general relativity with quantum field theory on curved spacetimes.

This suggests the following attitude towards the problem of time. We accept that $\mathcal{F}_{\text{phys}}$ has *no* inner product. We accept further that quantum general relativity is a timeless theory — i.e. it does not have a time parameter. In the semi-classical limit, however, we recover our ordinary notion of time in the form of WKB-time.

Therefore we take the most conservative point of view that the usual interpretation of quantum theory applies only as long as a (classical or semi-classical) background of space and time exists. As soon as the semi-classical approximation breaks down, spacetime and especially time dissolves. This is not only the end of the world as we know it, but also the boundary of the region in which the ordinary interpretation of quantum theory does apply.

In the full quantum regime, no time and therefore no Hilbert-space structure exists.

4.4.2 Relation to path integral

It was shown that the path-integral of the Einstein–Hilbert action, when constructed properly, solves the Wheeler–DeWitt equation. This holds formally and was checked explicitly up to the one-loop approximation, [36].[4]

Therefore, canonical quantum general relativity has an *equivalent* covariant formulation. Or, to formulate it the other way round, covariantly and canonically quantizing general relativity yields two equivalent formulations

[4]Of course, the factor ordering has to be chosen appropriately.

of the same theory. So I could (and actually should) extend the name *quantum general relativity* to include the covariant formulation. Indeed, this is how the name is often used in the literature. I will, however, refrain from doing so, because this thesis is only concerned with the canonical scheme.

The equivalence to the covariant formulation is a second assurance that the approach we pursue is not completely wrong. Moreover, anyone who is put off by the breaking of (explicit) covariance that is necessary for any canonical formulation, should now rest assured.

Canonical and covariant quantization are in general relativity — as in any other known quantum field theory — just two sides of the same coin.

References:

A concise list of goals and failures of quantum general relativity is given by Kuchar, [19]. A review of the semi-classical approximation in quantum general relativity is given by Kiefer in [37] which I used as source for the first section. All information on the relation between the canonical approach and the path-integral was drawn from the rather extensive review by Barvinsky, [36].

Chapter 5

Singularities in quantum General Relativity

One of the major motivations that stand behind the attempt to quantize general relativity is the fact that the classical theory itself predicts its own break-down in the form of spacetime singularities. The hope is that, on the quantum level, these singularities are, in some still to be specified way, overcome. But more than that, the resulting quantum-gravity theory is supposed to be singularity free itself, i.e. it should not break down at some point. There are two major questions pursued in this chapter. First of all, are the expectations of singularity resolution and singularity freedom well justified? Secondly, irrespective of our expectations, what happens to classical singularities in quantum general relativity? Is quantum general relativity singularity free? Following these questions, we will derive criteria which will be used to account for singularity resolution of the singularities of classical general relativity in the following chapters.

5.1 Singularities in classical General Relativity

It is a well-known fact that solutions to Einstein's equations rather generically form singularities. Best known examples of these are the big-bang singularity arising in the past of cosmological Friedmann–Robertson–Walker models and the black-hole singularity arising in spherically symmetric spacetimes. These singularities are both associated with gravitational collapse of either the entire spacetime or only a portion of it.

The big-bang singularity is characterized by infinite scalar curvature, energy density and pressure. The volume of spacetime shrinks to zero. The metric becomes degenerate. 'At' the black-hole singularity scalar curvature

likewise diverges, as does gravitational redshift. The metric becomes degenerate here as well.

Even though these are the most popular representatives of spacetime singularities, other types also exist — and not all of them are captured by the criterion of diverging curvature scalars. For example, plane gravitational waves as vacuum solutions of Einstein's equations have vanishing curvature scalars but diverging curvature tensor components. Another example are conical singularities.

An accepted criterion which encompasses most but not all types of singularity is that of *geodesic incompleteness*: a spacetime contains a singular point if at least one geodesic cannot be extended to its infinite parameter range. This is the definition used in the singularity theorems of Hawking and Penrose. These theorems pose certain conditions on spacetime and matter. I will quote here the most general of the theorems. Matter has to satisfy the strong energy condition. Spacetime has to be free of closed timelike curves and the timelike and null generic conditions have to be satisfied.[1] Then the singularity theorem states that *if* spacetime contains a trapped surface *or* a point from which on the expansion of a null geodesic congruence becomes negative along each geodesic *or* the spacetime is closed, *then* spacetime contains at least one incomplete causal geodesic.

One can conclude that rather generically spacetimes arise in which an observer (or light) can reach the end of spacetime in a finite time. As the singularity theorems make use of Einstein's equations only to employ the strong energy condition, not much can be learned from them as to *what happens* upon approach to the classical singularity.

5.2 What do we expect from quantum General Relativity?

The question is then what happens to these singularities if we go from classical to quantum general relativity. There are two questions that intermingle here. The first question is whether classical general relativity singularities persist in quantum general relativity. The second, a priori unrelated, question is whether quantum general relativity is itself a singularity-free theory or not.

5.2.1 Resolution of classical singularities?

The general expectation is that the singularities of classical general relativity are — in a yet to be specified way — resolved through quantum general

[1] A spacetime satisfies the timelike (null) generic condition if each timelike (null) geodesic contains at least one point at which $R_{\alpha\beta\gamma\delta}v^\alpha v^\delta \neq 0$. Here v is the tangent of the geodesic, $R_{\alpha\beta\gamma\delta}$ is the Riemann curvature tensor.

relativity.

But let us be very careful and ask *why* we actually should expect this. There are scarcely any arguments given in the literature. Singularity resolution through the quantization of general relativity seems to be more of a silent consensus than a well-argued expectation. But looking more closely at the problem, it seems very hard to find any *argument* for it.

Singularity resolution in quantum general relativity takes its cue from the quantization of the Coulomb problem. At least, this problem is regularly quoted when singularity resolution is discussed. Therefore, I reviewed the motion of an electron in the potential of a nucleus in the non-relativistic as well as special-relativistic setting on the classical and quantum level in detail in Appendix C. It turns out that this problem is actually a bad example for singularity resolution. Because in the non-relativistic case, no singularity exists on the classical level and on the special-relativistic level, the classical singularity *persists* in the quantum theory. So, in fact, there is no singularity resolution in the Coulomb potential case — either there is no singularity or there is no resolution.

Moreover, classical general relativity singularities are characterized by geodesic incompleteness. These are trajectories of point particles. But before classical general relativity breaks down, surely the point-particle approximation breaks down. Thus before looking at quantum general relativity effects, one should exclude that the singularity problem of general relativity could not be solved by the quantum aspects of matter propagating in spacetime.

5.2.2 Singularity-free quantum General Relativity?

Turning now to the second question, namely whether or not quantum general relativity itself is free of singularities or not, we first have to specify what we mean by this statement. Generalizing the term *singularity* here as referring to any point of configuration space of a theory at which the laws of physics do not yield unique predictions, singularities occur in physics at all ends. Usually they signal the breakdown of the framework one is using. This may be an effective description, as is the case e.g. in molecular physics and physical chemistry where singular potentials are used to mimick complicated molecular interactions. Or this may be a physical theory which is encompassed by a more fundamental one.

Mathematically, singularities are points at which the equations of the system break down.

Physically, a theory with singularities does not yield unique predictions throughout configuration space. In quantum theory, this can be discussed on two levels. Either in the language of self-adjoint operators. Then a unique solution exists if the Hamiltonian operator is essentially self-adjoint. Or in the language of partial differential equations. Then a unique solution

exists if the initial- or boundary-value problem (depending on the form of the quantum equation) is well-defined.

In quantum general relativity, the Hamiltonian is no self-adjoint operator — and should not be, so we cannot resort to operator language. Rather, we have to check whether the theory produces (functional) partial differential equations which have a well-defined initial- or boundary-value problem.

But we saw above that the Coulomb potential is singular in special relativity and remains so on the quantum level. Obviously, this is no reason to completely dismiss either the Coulomb potential or the Klein–Gordon equation. Rather, this singularity signals that the problem at hand — the motion of an electron in the potential of a nucleus — can, in the special-relativistic domain and on the quantum level, no longer be described by an instantaneous interaction as is suggested by the Coulomb potential. So the singularity of the Klein–Gordon equation with Coulomb potential stirred no crisis in physics because, anyway, the Coulomb potential is not believed to be a *fundamental* description of the problem at hand.

This bridges the gap to another idea that may underlie the assumption that quantum general relativity should be free of singularities. Namely, the idea that *quantum general relativity is a fundamental theory*. But this can be ruled out straight away. How can a perturbatively non-renormalizable theory like quantum general relativity be fundamental? Even more so, if the theory cannot be consistently coupled to fermionic matter?[2]

So actually, quantum general relativity is *not* a fundamental theory — it cannot be, and therefore singularities occurring on the quantum-gravitational level do not necessarily signal that the theory itself is meaningless. They merely stress the fact that quantum general relativity is not the final and fundamental theory.

5.2.3 Expectation

So we should not expect that quantum general relativity is free of singularities. Nor do we know what happens to classical singularities on the quantum level. It might be that some classical singularities persist in quantum general relativity. This is at least what we can learn from the Coulomb example. But then, what happens to classical general relativity singularities in quantum general relativity? And is quantum general relativity free of singularities in the sense that it predicts unique results or not?

[2]However, in Blagojevic's textbook the Hamiltonian formulation of Einstein–Cartan and general Poincare gauge theories are given, [40].

5.3 Singularities in quantum cosmology

Before turning to the full theory, I want to discuss the cosmological Friedmann–Robertson–Walker model, which will be introduced thoroughly in Chapter 6. The quantum-cosmological model arises through the quantization of the symmetry-reduced Friedmann–Robertson–Walker metric. Such a model is called *minisuperspace model* for reasons that will become clear in Section 5.4.1. The reduction makes this model much easier to deal with mathematically than the full theory. Nonetheless, relevant features survive in this model. Moreover, the cosmological case is central to this thesis and therefore *has* to be discussed.

Let us now pursue the question whether the quantum-cosmological model is singular or not, i.e. whether unique predictions can be retrieved from this model or not.

5.3.1 The Wheeler–DeWitt equation with cosmological constant

Due to the symmetry, no diffeomorphism constraint arises. We just have to deal with the Wheeler–DeWitt equation. To include all possible factor orderings, we write the kinetic term in the general form

$$\left[a^{r+m-1}\frac{\partial}{\partial a}a^{-r}\frac{\partial}{\partial a}a^{-m}\right] ,$$

where r and m are any real numbers and a is the scale factor, but see Chapter 6. The naive factor ordering where momenta are ordered to the right is then recovered for $r = 0$, $m = 0$. The 'Laplace–Beltrami' factor ordering is given by $m = 0$, $r = \frac{1}{2}$.[3]

The Wheeler–DeWitt equation in the vacuum case is given by

$$\frac{b}{a}\psi'' - \frac{b}{a^2}(r+2m)\psi' + \left[\frac{bm}{a^2}\left(r + \frac{m-1}{a}\right) + \lambda a^3 - \mathcal{K}a\right]\psi = 0 , \qquad (5.1)$$

where the following short-hands were used

$$b = \frac{\hbar^2\kappa^2}{12} , \qquad \lambda = \frac{\Lambda}{\kappa^2} , \qquad \mathcal{K} = \frac{k}{\kappa^2} ,$$

and Λ denotes the cosmological constant, $k = 0, \pm 1$ the spatial curvature and $\psi = \psi(a)$ is the wave function. The derivatives with respect to a are written as primes.

This is a well-defined equation on the interval $a \in I =]0; \infty[$. The requirements on ψ shall be minimal: it is supposed to be twice continuously

[3] The Laplace–Beltrami ordering is defined for two-dimensional surfaces. Carrying over this definition to the one-dimensional case, we obtain the given ordering.

differentiable in I, $\psi \in C^2(I)$. On I, the Wheeler–DeWitt equation can be brought into the form

$$\psi'' + \frac{P(a)}{a}\psi' + \frac{Q(a)}{a^2}\psi = 0 \ , \tag{5.2}$$

with

$$P(a) = -(r+2m) \ , \qquad Q(a) = m(ra+m-1) + \frac{\lambda}{b}a^6 - \frac{\mathcal{K}a^4}{b} \ . \tag{5.3}$$

From this, one can immediately read off that the left boundary of the interval, $a = 0$, is a *weakly singular* point of the equation.[4] This means that $p(a) = \frac{P(a)}{a}$ has a pole of at most first, and $q(a) = \frac{Q(a)}{a^2}$ a pole of at most second order at $a = 0$.

On the other hand, the right boundary of I, $a = \infty$, is in general *no* weakly singular point. To determine the behaviour there, make a transformation onto $\zeta = \frac{1}{a}$ and consider $\zeta = 0$. Using $\Phi(\zeta) := \psi(\frac{1}{\zeta})$ and dots to denote derivatives with respect to ζ, (5.2) becomes

$$\ddot{\Phi} + \frac{\tilde{P}(\zeta)}{\zeta}\dot{\Phi} + \frac{\tilde{Q}(\zeta)}{\zeta^2}\Phi = 0 \ ,$$

where

$$\tilde{P}(\zeta) = 2 + 2m + r \ , \qquad \tilde{Q}(\zeta) = m(m-1) + \frac{1}{\zeta}mr + \frac{\lambda}{b}\frac{1}{\zeta^6} - \frac{\mathcal{K}}{b}\frac{1}{\zeta^4} \ .$$

Obviously, $\tilde{q}(\zeta) = \frac{\tilde{Q}(\zeta)}{\zeta^2}$ has poles of at most second order only for vanishing curvature, vanishing cosmological constant and $r = 0$. The boundary $a = \infty$ is thus a *strongly singular* point of the Wheeler–DeWitt equation (5.2).

A basis that spans the two-dimensional solution space of the Wheeler–DeWitt equation (5.2) is obtained through Frobenius' method, [47]. It allows to determine a solution *around* a weakly singular boundary point. This solution may then be extended *to* the weakly singular boundary. The quantum cosmological theory yields unique predictons if a *unique* extension to the weakly singular point exists.

I will do this in some detail for the left boundary $a = 0$. The same methods and calculations can be applied to $a = \infty$ in the case that this point is weakly singular as well.

[4] Often, one also finds the expression 'regular singular point'.

The weakly singular point $a = 0$

Before Frobenius' method can be applied to (5.2), the equation has to be *normalized*. That means it has to be brought to the form

$$a^2\psi'' + aP(a)\psi' + Q(a)\psi = 0 \ , \tag{5.4}$$

where P and Q are analytical functions around the origin $a = 0$. Generally, they are therefore given in terms of power series', $P(a) = \sum_{j=0}^{\infty} p_j a^j$, $Q(a) = \sum_{j=0}^{\infty} q_j a^j$. Here, of course, the sum is finite and the coefficients can be directly read off from (5.3). Frobenius' method then simply consists of the ansatz

$$\psi(a) = a^\mu \sum_{n=0}^{\infty} c_n a^n \ , \qquad c_0 \neq 0 \ .$$

This solution can be complex valued, $\mu \in \mathbb{C}$, and $a^\mu = e^{\mu \ln a}$ for $a > 0$. The power series $\sum_{n=0}^{\infty} c_n a^n$ converges on $0 < a < \infty$. Inserting this ansatz into (5.4), one obtains, after dividing by a^μ,

$$\sum_{n=0}^{\infty} \left[(n+\mu)(n+\mu-1)c_n a^n + \sum_{j+k=n} c_k \left(p_j(k+\mu) + q_j \right) \right] = 0 \ ,$$

where Cauchy's product rule has been applied. Comparing coefficients, one arrives at

$$[(n+\mu)(n+\mu-1) + p_0(n+\mu) + q_0] + \sum_{k=0}^{n-1} c_k \left(p_{n-k}(k+\mu) + q_{n-k} \right) = 0 \ .$$

For $n = 0$, one arrives at the *indicial equation* determining the characteristic exponents μ

$$\mu(\mu-1) + p_0\mu + q_0 = 0 \ .$$

In our case, the characteristic exponents are given by

$$\mu_{1,2} = \frac{r+2m+1}{2} \pm \sqrt{\frac{(r+2m+1)^2}{4} - m(m-1)} \ .$$

For clarity, I will here assume first that $\mu_{1,2} \in \mathbb{R}$ and the labelling is done such that $\mu_1 \geq \mu_2$. I will come to the case of complex μ_1, μ_2 subsequently. Note that whether or not $\mu_{1,2}$ are real, depends on the factor ordering. Setting $c_0 = 1$ (this is done conventionally), the series is determined through the recurrence relation

$$c_n = -\frac{1}{D(n+\mu)} \sum_{k=0}^{n-1} c_k \left(p_{n-k}(k+\mu) + q_{n-k} \right) \ , \tag{5.5}$$

for each characteristic exponent. Here, $D(\nu) = \nu(\nu-1) + p_0\nu + q_0$. This works, of course, only if $D(n+\mu) \neq 0$. For μ_1, $D(n+\mu_1) \neq 0$ always — it is the larger of the two roots. For μ_2, $D(n+\mu_2) = 0$ for $n = 0$ and for $n = \mu_1 - \mu_2$. So for the characteristic exponent μ_1, Frobenius' method *always* provides a generalized power series solution. For the second characteristic exponent μ_2, this is also the case *as long as* $\mu_1 - \mu_2 \notin \mathbb{N}$. But *if* $\mu_1 - \mu_2 \in \mathbb{N}$, the second, linearly independent solution has to be found from the series expansion with slight modifications. In the respective cases, a basis of the solution space is given by

Case A: $\mu_1 - \mu_2 \notin \mathbb{N}$

In this case, the recurrence relation (5.5) can be used to arrive at a solution for both characteristic exponents. The two linearly independent solutions are given by

$$\psi_1(a) = a^{\mu_1} \sum_{n=0}^{\infty} c_n a^n \ , \qquad \psi_2(a) = a^{\mu_2} \sum_{n=0}^{\infty} d_n a^n \ .$$

Here, $c_0 = 1$, $d_0 = 1$ and all higher coefficients follow from (5.5) with $\mu = \mu_1$ for c_n and $\mu = \mu_2$ for d_n, respectively.

Case B: $\mu_1 - \mu_2 \in \mathbb{N}$

In this case, the recurrence relation (5.5) can be used to arrive at a solution only for the characteristic exponent μ_1, $\psi_1(a) = a^{\mu_1} \sum_{n=0}^{\infty} c_n a^n$. The second solution is of the form

$$\psi_2(a) = a^{\mu_2} \sum_{n=0}^{\infty} d_n a^n + \gamma \, \psi_1 \ln a \ .$$

The values of d_n and γ depend on whether $\mu_1 - \mu_2 = 0$ or not.

Case B. I: $\mu_1 - \mu_2 = 0$

Here, $d_0 = 0$ and higher coefficients follow from

$$d_n = -\frac{1}{D(n+\mu_1)} \sum_{k=0}^{n-1} d_k \left(p_{n-k}(k+\mu_1) + q_{n-k} \right) + \frac{\lambda_n}{D(n+\mu_1)} \ ,$$

where λ_n follow from a series expansion of

$$a^{\mu_1} \sum_{n=0}^{\infty} \lambda_n a^n = (1 - P(a))\psi_1 - 2a\psi_1' \ .$$

Moreover, $\gamma = 1$. Thus, in this case, *a logarithmic contribution and thus a divergence at $a = 0$ is unavoidable.*

Case B. II: $\mu_1 - \mu_2 \neq 0$

Set $\mu_1 - \mu_2 = M$. Then $d_0 = 1$ and $d_M = 0$. Higher coefficients follow from

$$d_n = -\frac{1}{D(n+\mu_2)} \sum_{k=0}^{n-1} d_k \left(p_{n-k}(k+\mu_2) + q_{n-k}\right) \ ,$$

for $n < M$. The equation for $n = M$ is used to determine γ,

$$-M\gamma = \sum_{k=0}^{M-1} d_k \left(p_{M-k}(k+\mu_2) + q_{M-k}\right) \ ,$$

whereas all coefficients for $n > M$ follow from

$$d_n = -\frac{1}{D(n+\mu_2)} \sum_{k=0}^{n-1} d_k \left(p_{n-k}(k+\mu_2) + q_{n-k}\right) + \gamma \frac{\lambda_{n-M}}{D(n+\mu_2)} \ .$$

And λ_n is given as in Case B.I. Here, we cannot determine straight away whether or not a logarithmic divergence occurs. It depends on the specific form of the equation and has to be inferred from the recurrence relation.

Depending on the values of r and m, several cases arise. These are depicted in Table 5.1.

Let the difference $\mu_1 - \mu_2 = r + 2m + 1$ be an arbitrary number, $r + 2m + 1 \notin \mathbb{N}$. Then a *unique*, well-defined solution exists for $r + 2m + 1 > 0$ and $m \in]0; 1[$, as well as for $r + 2m + 1 \leq 0$.

If, on the other hand, $r + 2m + 1 > 0$ and $m = 0$ or $m = 1$, then the *additional condition* $\psi \to 0$ as $a \to 0$ is needed to pick a unique solution.

However, if $r + 2m + 1 > 0$ and $m(m-1) > 0$, *no solution is picked by this condition as both vanish at the origin.*[5]

If $r + 2m + 1 \in \mathbb{N}$, ψ_2 may acquire a logarithmic divergence.

Therefore, we conclude that for factor orderings $m \in]0; 1[$, a *unique* solution exists which can be continuously extended to the origin and vanishes there. For $m \in [0; 1]$ and $r + 2m + 1 \leq 0$, ψ_1 is the only contiuously extendable solution but does not necessarily vanish.

The form of the two linearly independent solutions depends thus very decisively on the factor ordering.

[5] We assume that m, r are chosen such that μ_i are real.

	$A > 0$	$A = 0$	$A < 0$
$B > 0$	$\mu_1 > 0$ $\mu_2 > 0$	no real solution	$\mu_1 < 0$ $\mu_2 < 0$
$B = 0$	$\mu_1 > 0$ $\mu_2 = 0$	$\mu_1 = 0$ $\mu_2 = 0$	$\mu_1 = 0$ $\mu_2 < 0$
$B < 0$	$\mu_1 > 0$ $\mu_2 < 0$	$\mu_1 > 0$ $\mu_2 < 0$	$\mu_1 > 0$ $\mu_2 < 0$

Table 5.1: This table shows which factor-orderings allow which kind of solutions. The factor-orderings are parametrized by r and m. The relations $A = r+2m+1$ and $B = m(m-1)$ determine the nature of the solution. In the table, white cells denote factor-orderings for which a single, regular solution exists. These are the cases we are looking for: the Wheeler–DeWitt equation can be uniquely extended to the origin. There are two cases in which *two* regular solutions might exist. The corresponding cells are coloured light-grey. For these factor-orderings, no *unique* solution to the Wheeler–DeWitt equation can be found. For $A < 0$ and $B > 0$, no regular solution exists. The Wheeler–DeWitt equation is singular.

As examples, I want to discuss the naive ordering and the 'Laplace–Beltrami' one. For the naive factor ordering, $\mu_1 = 1$, $\mu_2 = 0$. Thus, we have a Case–B.II scenario: there *may* be a logarithmic contribution. Direct calculation yields that $\gamma = -q_1 = 0$, and so *no logarithmic dependence occurs*. The two solutions are continuous as $a \to 0$. But whereas $\psi_1 \to 0$ in this limit, $\psi_2 \to const.$ there. The boundary condition $\psi(0) = 0$ would then pick a unique solution.

For the 'Laplace–Beltrami' ordering, one obtains $\mu_1 = \frac{3}{2}$, $\mu_2 = 0$ and therefore a Case–A scenario. Again, only ψ_1 vanishes at the origin, but both solutions remain bounded.

So for *both* factor orderings, a *unique* solution is chosen by the requrirement that the wave function vanish at the origin $a = 0$. In both cases, *both* solutions can be continuously extended to the weakly singular boundary point $a = 0$.

Note that because we are *not* working with self-adjoint operators here, the requirement $\psi(0) = 0$ does *not* arise in a natural way.[6]

The procedure remains essentially unaltered if complex characteristic exponents are taken into account. Instead of comparing μ_1 and μ_2, one now has to compare $e^{2\pi i\mu_1}$ and $e^{2\pi i\mu_2}$.

[6]Recall that e.g. for the Schrödinger equation with Coulomb potential the condition $r\psi(r) \to 0$ as $r \to 0$ arises as a consequence of the self-adjointness of the Hamiltonian operator. The radial coordinate is denoted by r here.

The Case–A scenario is then recovered if $e^{2\pi i\mu_1} \neq e^{2\pi i\mu_2}$. In this case, the recurrence relation can be used to obtain two linearly independent solutions

$$\psi_1(a) = a^{\mu_1}\,(1 + h_1(a)) \ , \qquad \psi_2(a) = a^{\mu_2}\,(1 + h_2(a)) \ , \tag{5.6}$$

where h_i are analytical functions vanishing for $a = 0$. For $e^{2\pi i\mu_1} = e^{2\pi i\mu_2}$, the general solution can contain a logarithmic dependence. This is just the Case–B scenario. Solutions are again of the form

$$\psi_1(a) = a^{\mu_1}\,(1 + h_1(a)) \ , \qquad \psi_2(a) = a^{\mu_2}\,(1 + h_2(a)) + \gamma\,\psi_1(a)\ln a \ . \tag{5.7}$$

These solutions are unique on the punched disk $U(a)$ around the origin. Out of the system of solutions, those can be picked for which the limit $\lim_{a\to 0}\psi$ exists and is continuously differentiable. Let this solution space be called $\{f_0\psi_0(a)|f_0 \in \mathbb{R}\}$.

The strongly singular point at $a = \infty$

Infinity is a *strongly* singular point of the equation — if $\Lambda \neq 0$, $k \neq 0$, $r \neq 0$. Only if these terms vanish, infinity is a *weakly* singular point. Thus, counter-intuitively, the problematic point of the cosmological Wheeler–DeWitt equation is not the origin but infinity. If infinity is a weakly singular point, as above, the solution can be determined through a power-series expansion. So solutions are of the form (5.6) or (5.7), depending on the relation between the two characteristic exponents. Similarly, conditions have to be posed on the wave function at infinity in order to pick a unique solution. Let this solution space be called $\{f_\infty\psi_\infty(a)|f_\infty \in \mathbb{R}\}$.

The overall solution is uniquely defined only when the two solutions $\psi_0(a)$ and $\psi_\infty(a)$ differ at most by a constant.

More specifically, from the indicial equation follow the characteristic exponents

$$\nu_{1,2} = -\frac{1}{2} - m \pm \sqrt{\frac{1}{4} + 2m} \ .$$

Thus $\nu_1 - \nu_2 = \sqrt{1 + 8m}$. For the naive ordering, we have $m = 0$ and thus $\nu_1 - \nu_2 = 1$. We therefore have a Case–B.II scenario. Because $\gamma = 0$ and $c_n = 0$, $d_n = 0$ for all $n > 0$, we get the two solutions

$$\Phi_1(\zeta) = 1 \ , \qquad \Phi_2(\zeta) = \zeta^{-1} \ .$$

On the other hand, the same choice of constants $\Lambda = 0$, $k = 0$, $r = 0$ produces the following solutions around $a = 0$

$$\psi_1(a) = a \ , \qquad \psi_2(a) = 1 \ .$$

If we require that $\psi \to 0$ as $a \to 0$ and $\Phi \to 0$ as $\zeta \to 0$, no overall solution is found. If we require $\psi \to const.$ as $a \to 0$ and $\Phi \to const.$ as $\zeta \to 0$, the only solution is the trivial one, the solution is simply constant. So it seems that no overall solution exists! This is a general result and does not depend on the simplifying assumption $m = 0$ made here.

What I want to emphasize at this point, however, is that here the regularity condition *at the origin alone* suffices to fix the solution. If we simply require that $\psi \to 0$ as $a \to 0$ and choose the solution at infinity so that it matches with the one obtained at the origin, we get the unique solution $\psi(a) = a$.[7] This can be justified by the observation that infinity is not a singular point like any other. It is much more natural to accept a solution which diverges as the argument goes out to infinity than to accept a divergence at a finite value like the origin.

For example, the Legendre differential equation has two weakly singular points at ± 1. The solution is found as outlined above and the extensions to both singular points are made in a continuous way. The Laguerre differential equation, on the other hand, is weakly singular at the origin and strongly singular at infinity — just like the vacuum Wheeler–DeWitt equation encountered here. Whereas one requires continuity of the solution at the origin, no such condition is imposed at infinity. Instead, one subjects solutions to the requirement that they be square-integrable with respect to a certain measure. Again, this condition results from the inner product which we have not at hand here. So, infinity is subjected to much weaker conditions than a finite singular point.

This brings us back to the conditions on the cosmological wave function at infinity. I would suggest to 'ignore' the fact that infinity is a strongly singular point of the equation and simply use the solution as determined at $a = 0$. Recall that, in any case, $a = \infty$ will not be a weakly singular point in general. This solution will then diverge at infinity and, of course, will not be square-integrable with respect to the measures conventionally used in quantum theory.

But, anyway, we are not interested in square-integrable wave functions with respect to a. Moreover, a wave function that vanishes for large radii of the universe would suggest an untimely end for the corresponding classical world. It would imply that no universe could expand forever — which really would be too bad. Lastly, note that $\psi \to 0$ as $a \to 0$ *and* as $a \to \infty$ would imply a fixing of the wave function at both ends and thus quantization

[7] This is *no* general result but depends on the choice of factor ordering. Note that the existence of a unique, regular solution and the factor-ordering question are tightly interwoven. The fact that a certain factor ordering provides a unique solution *could* be used as an argument in favour of this specific ordering.

conditions.

So, to conclude, for the class of factor orderings $m \in]0;1[$, a unique solution exists which can be continuously extended to the origin and vanishes there. Moreover, in all cases except $r+2m+1 > 0$ and $r+2m+1 \notin \mathbb{N}$, the condition of mere *continuous extendability* to the origin picks $\psi_1(a)$ as unique solution.

5.3.2 The Wheeler–DeWitt equation containing matter

Even though the cosmological-constant case is the easiest to deal with, it is surely not the most relevant scenario. Actually, little effort has to be made to arrive at some conclusions including matter. But to this end, some assumptions have to be made. First of all, note that minimally coupled fields have a kinetic term $\mathcal{L}_{\rm kin} \sim N\sqrt{\det h}\frac{(\dot{\phi})^2}{2}$, where ϕ represents *any* bosonic field component. Therefore, the inverse of the determinant $\det h$ enters in the kinematical part of the Hamiltonian $\mathcal{H}_{\rm kin} \sim \frac{p_\phi^2}{2\sqrt{\det h}}$ with momentum p_ϕ of the field. Defining $\mathcal{H}'_{\rm matter}$ through $\mathcal{H}_{\rm matter} = \frac{1}{\sqrt{\det h}}\mathcal{H}'_{\rm matter}$ and using that $\sqrt{\det h} = a^3$, the Wheeler–DeWitt equation becomes

$$\left[\hat{\mathcal{H}}_{\rm grav} + \frac{1}{a^3}\hat{\mathcal{H}}'_{\rm matter}\right]\psi = 0 \ .$$

This is only helpful if we decompose the wave functional into $\psi(a,\phi) = C_n(a)\varphi_n(a,\phi)$ where

$$\hat{\mathcal{H}}'_{\rm matter}\varphi_n(a,\phi) = E_n(a)\varphi_n(a,\phi) \ .$$

Then the Wheeler–DeWitt equation reads

$$\left[\hat{\mathcal{H}}_{\rm grav} + \frac{1}{a^3}E_n(a)\right]\psi(a,\phi) = 0 \ ,$$

or,

$$\psi'' + \frac{P(a)}{a}\psi' + \frac{Q(a)}{a^2}\psi = 0 \ , \tag{5.8}$$

with

$$P(a) = -(r+2m) \ , \qquad Q(a) = m(ra+m-1) + \frac{\lambda}{b}a^6 - \frac{\mathcal{K}a^4}{b} + \frac{E_n(a)}{b} \ . \tag{5.9}$$

To apply the previously used Frobenius' method, it is necessary to evaluate the derivatives, $\psi' = \varphi'_n C_n + \varphi_n C'_n$ and similarly the higher ones. The resulting equation for C_n has to be brought into the form (5.8), with C_n replacing ψ. From the newly determined $P(a)$ and $Q(a)$ one can then infer whether or not the boundaries $a = 0$ and $a = \infty$ are weakly singular. If they

are, Frobenius' method can be employed and solutions can be determined. As the result depends very crucially on the matter wave function, no general conclusion can be drawn on this level.

The Wheeler–DeWitt equation coupled to matter will be an arbitrarily complicated partial differential equation. In a specific case, it will therefore be much more useful to take refuge to the known existence and uniqueness proofs for partial differential equations.

Wheeler–DeWitt equation with free scalar field

A simple example, however, is the cosmological model with free scalar field. With Laplace–Beltrami factor ordering, the Wheeler–DeWitt equation reads

$$\frac{\hbar^2\kappa^2}{12}\left(\frac{1}{a^2}\frac{\partial\psi}{\partial a}+\frac{1}{a}\frac{\partial^2\psi}{\partial a^2}\right)-\frac{\hbar^2}{2}\frac{1}{a^3}\frac{\partial^2\psi}{\partial\phi^2}+\left(-\mathcal{K}a+\frac{\lambda}{3}a^3\right)\psi=0\ .$$

Making the ansatz $\psi(a,\phi)=C_n(a)\varphi_n(a,\phi)$ with

$$-\frac{\hbar^2}{2}\frac{\partial^2\varphi_n}{\partial\phi^2}=n^2\,\varphi_n\ ,$$

as suggested above, yields

$$C''+\frac{P(a)}{a}C'+\frac{Q(a)}{a^2}C=0\ ,$$

where here $P(a)=1$ and $Q(a)=\frac{n^2}{b}-\frac{\mathcal{K}}{b}a^4+\frac{\lambda}{3b}a^6$ and primes still denote differentiation with respect to a. At $a=0$, we find the characteristic exponents

$$\mu_{1,2}=\pm\frac{n}{\sqrt{b}}\ .$$

Therefore $\mu_1-\mu_2=\frac{2|n|}{\sqrt{b}}$. Thus, it depends on the values of n which scenario we encounter. But note that any solution which is free of a logarithm vanishes at the origin. So *we always have one solution which vanishes at the origin*, namely $\psi_1(a)=a^{\mu_1}\sum_m c_m a^m$ with $\mu_1=\frac{|n|}{\sqrt{b}}$.

Note that $\mu_1-\mu_2=0$ only for the trivial $\varphi_n=const.$ solution obtained for $n=0$.

For arbitrary values of n, $\mu_1-\mu_2\notin\mathbb{N}$. Then ψ_2 diverges as the origin is approached because $\mu_2=-\frac{|n|}{\sqrt{b}}$. If $|n|=M\frac{\sqrt{b}}{2}$, $M\in\mathbb{N}$, then $\mu_1-\mu_2\in\mathbb{N}$. In this case ψ_2 diverges at least as $\psi_2\sim a^{-M}$. A logarithmic divergence may ensue, depending also on the precise value of M.

Therefore, for $n\neq 0$ a unique solution exists which can be continuously extended to the origin and vanishes there. For $n=0$, again ψ_1 is the only solution that can be extended to the origin as for $\mu_2=0$, $\gamma=1$ and ψ_2

contains a logarithmic divergence. But in this case, the only continuously extendable solution is *constant* at the origin.

5.3.3 Conclusions

The Wheeler–DeWitt equation with cosmological constant *is singular*. It has two singular points, $a = 0$ and $a = \infty$. The singularity at $a = 0$ is, however, a *weak* singularity. That means solutions can be extended to this point. We showed that for the class of factor orderings $m \in]0; 1[$, a *unique* solution exists which can be contiuously extended to the origin. The same holds for $r + 2m + 1 \leq 0$. Whereas in the first case the only continuously extendable solution vanishes at the origin, it may remain constant in the second scenarios.

The singularity at infinity is a strong one. Generally, this implies that solutions diverge as $a \to \infty$. I suppose that there is no physical argument against such behaviour.

The weak singularity $a = 0$ of the quantum equation has, of course, a counterpart on the classical level: $a = 0$ is just the 'location' of the big-bang singularity. So this singularity which is associated with the collapse of the entire three-volume to one point pertains to the quantum level in the form of a weakly singular point of the Wheeler–DeWitt equation with cosmological constant.

It is hard to draw any conclusions for the Wheeler–DeWitt equation with arbitrary matter content. The reason for this is that additional degrees of freedom turn the Wheeler–DeWitt equation into a partial differential equation whose precise form depends on the matter Hamiltonian. One can, however, expect that the singular point $a = 0$ remains problematic as the second derivative with respect to a *always* occurs with a prefactor a^{-1}. Moreover, singularities in the matter potential are candidates for problematic points of the Wheeler–DeWitt equation.

In the simple case of a free scalar field, the singular nature of $a = 0$ does not hinder the existence of a unique solution which is constant at $a = 0$.

We will see that this method generalizes also to scalar fields with potential term — after a suitable approximation has been carried out, see Chapter 7 and Chapter 8.

5.4 Singularities in full quantum General Relativity

We will now try to find out how these features of the quantum-cosmological Wheeler–DeWitt equation look like in *full* quantum general relativity. The

Wheeler–DeWitt equation will contain only a cosmological constant. Again, not much can be said about other matter content.

In full quantum general relativity, the governing equations are the Wheeler–DeWitt equation *and* the quantum diffeomorphism constraints. The Wheeler–DeWitt equation in the naive factor ordering is given by (4.5). In the following, I will disregard the infinities associated with the functional character of the equations. Assume that these are regularized in some way.

It is immediately obvious from (4.5) that the Wheeler–DeWitt equation as it stands is no well-defined equation. Due to the divergence of the DeWitt metric for $\det h = 0$, the kinetic term diverges — irrespective of the choice of factor ordering — at all points where $\det h = 0$. We thus have no existence and uniqueness proof for the solution at these points. So *the quantum theory contains a singularity which is a priori not connected with the singularities of classical general relativity.*

This is just the same phenomenon as in the quantum-cosmological case where $\det h = a^6 = 0$. There, we could use Frobenius' method to ensure that at least the cosmological constant and free scalar-field solutions could be continuously extended to $\det h = a^6 = 0$. If we deal with the full Wheeler–DeWitt equation, our mathematical tools are much more limited.

To discuss this singularity properly, it is useful to take a closer look at the domain manifold of the wave functional.

5.4.1 Superspace

At each point of Σ, the space of all three-metrics is given. The DeWitt-metric defines a metric on this space, turning it into a metric manifold — again at each $x \in \Sigma$. Denote this manifold by Q_x. The configuration space, $Q(\Sigma)$, is then a product manifold

$$Q(\Sigma) = \prod_{x \in \Sigma} Q_x \ .$$

The quantum diffeomorphism constraints insure that the admissible wave functional depends only on the three-geometry, not the three-metric. Assuming that the diffeomorphism constraints have been solved, the Wheeler–DeWitt equation can be considered as an equation defined on $Q(\Sigma)$ *modulo the orbits of three-diffeomorphisms.* The resulting orbifold is the space of all three-geometries, ${}^{(3)}\mathcal{G}$, denoted by $\mathcal{Q}$. This space is called *superspace* and often also referred to as the configuration space of general relativity.

DeWitt discovered that the manifold Q_x can be mapped onto a six-dimensional manifold with Lorentzian signature. The timelike variable is given by $\xi = \sqrt{\frac{32}{3}}(\det h)^{\frac{1}{4}}$. This manifold is geodesically incomplete with a curvature singularity at ${}^{(6)}R \sim -\xi^{-2}$, where ${}^{(6)}R$ denotes the six-dimensional Ricci scalar.

In the new coordinates, the Wheeler–DeWitt equation becomes

$$\left[\frac{\hbar^2}{2M}\left(-\frac{\delta^2}{\delta\xi^2}+\frac{32}{3\xi^2}G^{AB}\frac{\delta^2}{\delta\xi^A\delta\xi^B}\right)+\frac{3M}{32}\xi^2\left({}^{(3)}R-2\Lambda\right)\right]\Psi[{}^{(3)}\mathcal{G}]=0\ . \tag{5.10}$$

The metric G^{AB} is positive definite and a function of the three-metric and its inverse; ξ^A, $A=1,\dots 5$ are the spatial coordinates with respect to the DeWitt metric. It is obvious that on the spatial hypersurface (in Q_x) $\xi=0$ the Wheeler–DeWitt equation is *not* well-defined. This means that unique solutions for $\xi<0$ and $\xi>0$ can be joined to each other in a unique way *only if some additional condition is provided*. The singular hypersurface *in* $\mathcal{Q}$ defined through $\det h=0$ (or $\xi=0$) is called *quantum barrier* and denoted by $\mathcal{B}_{\rm quant}$ following the notation of DeWitt. Note that $\det h=0$ for the cosmological big-bang singularity as well as for the black-hole singularity. But this does not mean that $\det h=0$ always corresponds to a geometrical singularity. Σ and the foliation may just have been chosen poorly.

So, as in the quantum-cosmological case, we encounter a singularity of the quantum equation also for the full theory. This singular region is connected with the vanishing of the determinant of the three-metric. That means, we encounter a singularity of the Wheeler–DeWitt equation whenever the three-metric becomes degenerate.

The structure of this singular region is much more complicated than in the cosmological case.[8] Here, the singular region arises from the product manifold

$$F(\Sigma)=\prod_{x\in\Sigma}F_x\ ,$$

where F_x contains all elements from Q_x which have vanishing determinant. But then we still have to divide out the diffeomorphisms. In this way we arrive at $\mathcal{B}_{\rm quant}$.

I want to make two remarks at this point. First of all, note that in quantum cosmology $\det h=0$ was equivalent to $a=0$ and thus the singularity lie at the *boundary* of configuration space (minisuperspace). Here, the region $\det h=0$ does not lay at the boundary because we have not taken care of the positivity condition of the three-metric, recall Chapter 4, Section 4.2.1. If we would in some way take care of this property of the three-metric, we would exclude $\det h<0$, i.e. negatively oriented volume, and move the singular region to the 'boundary' of configuration space.

Secondly, in the quantum-cosmological case a unique solution exists that vanishes at the origin. Here, we encounter the problem that we do not know

[8] It is hard to beat a point in simplicity.

how to join together the wave functions in the regions $\det h < 0$ and $\det h > 0$ across $F(\Sigma)$. But we could alleviate the problem by simply *imposing* some condition there. This condition should, of course, be chosen such that it simultaneously makes the Wheeler–DeWitt equation on $\det h = 0$ finite.

5.4.2 DeWitt's boundary proposal

Such a condition was proposed by DeWitt. He supposes that a well-defined problem arises for the Wheeler–DeWitt equation when one subjects the wave functional to the additional condition

$$\Psi[{}^{(3)}\mathcal{G}] = 0 \ , \qquad \forall \quad {}^{(3)}\mathcal{G} \in \mathcal{B}_{\mathrm{quant}} \ . \tag{5.11}$$

On configuration space $Q(\Sigma)$ instead of superspace, this condition reads

$$\Psi[h] = 0 \ , \qquad \forall \quad h \in F(\Sigma) \ .$$

This condition is known as DeWitt's boundary proposal. There are two facts which support that this proposal might actually work. First of all, note that the term $\frac{1}{\sqrt{\det h}}\Psi[h]$ that causes the divergence in the Wheeler–DeWitt equation vanishes if $\Psi[h] = 0$ for $\det h = 0$. Secondly, consider the Klein–Gordon form, (5.10), of the Wheeler–DeWitt equation. From the *form* of this equation, one would suppose that a unique solution needs specification of Ψ and $\frac{\delta\Psi}{\delta\xi}$ on a $\xi = const.$ hypersurface. The core idea of DeWitt's proposal is then to fix Ψ on two hypersurfaces of constant ξ — *instead of fixing the first derivative*. One of these hypersurfaces corresponds just to some arbitrary $\xi = const.$ but the second one should be $\xi = 0$ and the condition on it $\Psi[\xi = 0] = 0$.

Thus fixing the wave function at each x for two spatial hypersurfaces F_x and $\mathcal{S}_x$ in Q_x fixes the wave function on surfaces $F(\Sigma)$ and $\mathcal{S}(\Sigma)$ in $Q(\Sigma)$. Note that $\dim F(\Sigma) = \dim \mathcal{S}(\Sigma) = 5 \times \infty^3$. We can thus expect that this suffices to fix the wave function completely for all three-geometries.

DeWitt carries this construction further to superspace. Let $\mathcal{S}(\Sigma)$ modulo three-diffeomorphisms be denoted by $\mathcal{S}$. Then we fix $\Psi[{}^{(3)}\mathcal{G}]$ on $\mathcal{B}_{\mathrm{quant}}$ and $\mathcal{S}$.

5.4.3 Regularizing the Wheeler–DeWitt equation

A way of regularizing the Wheeler–DeWitt equation is the transformation on a new variable

$$X = \ln\left(\frac{\det h}{\det h_0}\right) \ ,$$

where $\det h_0$ is some constant with unit length to the sixth power, used to keep the dimensions right. The $\det h = 0$ singularity is moved from the finite

value $\det h = 0$ to the infinite $X = -\infty$. As $\det h \sim e^X \neq 0$, the Wheeler–DeWitt equation can, after this transformation of variables, be multiplied by e^X to yield a finite, singularity-free (functional) partial differential equation.

More than just regularizing the Wheeler–DeWitt equation, the new variable implements the *positivity condition* on the metric: $\det h > 0$. This was the original context in which this proposal arouse. We will use this method to obtain a well-defined partial differential equation in the cosmological models discussed in the following chapters.

5.4.4 Coupling to matter

Note that so far, we only considered pure gravity. As discussed in Section 5.3, coupling to matter introduces another singularity stemming from the fact that minimally coupled fields have a kinetic term $\mathcal{L}_{\rm kin} \sim N\sqrt{\det h}\frac{(\dot{\phi})^2}{2}$. Writing $\mathcal{H}_{\rm matter} = \frac{1}{\sqrt{\det h}}\mathcal{H}'_{\rm matter}$, the Wheeler–DeWitt equation reads

$$\left[\hat{\mathcal{H}}_{\rm grav} + \frac{1}{\sqrt{\det h}}\hat{\mathcal{H}}'_{\rm matter}\right]\Psi = 0 \ .$$

One can then use the same tricks as in the cosmological case, simply replacing a by h_{ab}. However, now we have no mathematical tool to deal with the resulting equation,

$$\left[\hat{\mathcal{H}}_{\rm grav} + \frac{1}{\sqrt{\det h}}E_k(h)\right]\Psi = 0 \ .$$

But we see that the coupling to matter does not alleviate, but to the contrary rather aggravates the problem associated with the $\det h = 0$ singularity.

5.5 Criteria for singularity resolution

We thus encountered two sets of singularities. First, there are the classical singularities of general relativity. Secondly, the Wheeler–DeWitt equation is singular on the set $\mathcal{B}_{\rm quant}$ of *quantum* general relativity singularities.

Because classical singularities are defined via geodesic motion in spacetimes, it is not easy to track down the classical singularities in the quantum framework. Considering Einstein's equations as evolution equations, the idea is that a *regular* three-geometry evolves into a *singular* three-geometry. The set of singular three-geometries is a subset of superspace. This set is denoted by $\mathcal{B}_{\rm class}$. It seems that the set of classical singularities comprises that of quantum singularities, $\mathcal{B}_{\rm quant} \subset \mathcal{B}_{\rm class}$.[9] Thus out of the total set

[9] Arguments can be found in [4]. In my opinion, the relation between $\mathcal{B}_{\rm class}$ and $\mathcal{B}_{\rm quant}$ is not very well understood. A connection between both sets can be made through geodesic motion *in* $\mathcal{Q}$. Any geodesic hits the quantum barrier $\mathcal{B}_{\rm quant}$ in finite proper time. On the other hand, one can recover Einstein's equations from a WKB-ansatz to the Wheeler–

of classical singularities, we have those which do *not* occur on the quantum level and those, as the big bang, for example, which *do* occur through $\det h = 0$.

Singularities of the first class are thus simply resolved through the fact that classical spacetime is dissolved. The corresponding criterion is the break-down of the semi-classical approximation. Those of the second class, that is the hope, are resolved if the wave functional vanishes on the respective region $\mathcal{B}_{\text{quant}}$ of superspace. The question remains whether this has to be put in by hand, as in DeWitt's boundary proposal, or is insured by some other mechanism.

I will use the two criteria described above as indicators for singularity resolution. In the following, I will briefly describe the picture standing behind these criteria and also comment shortly on other criteria that can be found in the quantum-cosmology literature.

5.5.1 Break-down of semi-classical approximation

Through the quantization of three-metric and extrinsic curvature, in the form of conjugate momentum, we loose classical spacetime. Loosely speaking, we either have three-geometries but no prescription of how to stack these together to a spacetime or we have a prescription of how to stack three-geometries together but do not have any three-geometry. This follows in a hand-wavy way from the uncertainty relations of three-metric and momentum.

Of course, the choice of criteria depend on the attitude one takes towards the interpretation of the Wheeler–DeWitt equation and its solutions. Recall that we adopt the semi-classical interpretation here.

Only in the semi-classical realm can we recover spacetime as a history of three-geometries ('geometrodynamics'), recall Chapter 4. Whenever the semi-classical approximation breaks down, no spacetime exists which can be probed for geodesic incompleteness. No spacetime, no singularity.

A breakdown of the semi-classical approximation is signalled by the spreading of initially tightly peaked wave packets.

5.5.2 Vanishing of the wave function

As discussed already in the previous section, the singularities associated with $\mathcal{B}_{\text{quant}}$ require a vanishing wave functional. We accept the vanishing of the wave functional at the classically singular point in configuration space as a valid criterion for singularity avoidance. If the wave functional vanishes

DeWitt equation. It turns out that Einstein's equations are just geodesic equations in superspace — plus an extra force term. The question is then whether the force term is strong enough to prevent the three-geometry from hitting the $\det h = 0$ frontier. An answer to this question is not known. Neither is it clear under which conditions the singular three-metric $\det h = 0$ corresponds to a true geometrical singularity in spacetime.

at some region in configuration space, this region cannot contribute to the quantum theory. However, we do not want to put this vanishing in by hand, as is done in DeWitt's boundary condition. Rather, we want to see whether other mechanisms already enforce this vanishing.

5.5.3 Other criteria

As another citerion for singularity avoidance the finiteness of expectation values of classically diverging quantities has been used. As we have no well-defined inner product to define such expectation values and as expectation values defined outside the realm of the semi-classical approximation surely transcend our understanding of ordinary quantum theory, we do not apply such a criterion here. But see also Chapter 9, Section 9.1.1

Another criterion is that of a *quantum bounce* in which the wave packet remains tightly peaked but deviates from the classical trajectory. Here the classical trajectory is altered through quantum effects. That means a semi-classical approximation holds throughout but yields a different evolution than the purely classical one. Usually, the wave packet falls off exponentially in the region of the classical singularity. Therefore this scenario is comprised in Section 5.5.2.

References: I used the textbook of Robert Wald for the brief discussion of the classical singularities in general relativity, [41]. For the singularities of quantum general relativity, I used DeWitt's paper, [4]. A very readable introduction to the theory of ordinary differential equations is given in Jänich's textbook, [43]. Further reading is provided in the textbook of Fischer and Kaul, [47].

Chapter 6

Quantum Cosmology

In this chapter the canonical quantization scheme will be applied to the Friedmann–Robertson–Walker metric. Quantization on such a highly-symmetric superspace is called minisuperspace quantization. This name alludes to the fact that the symmetry actually reduces the infinite number of degrees of freedom of the full theory to a finite one. Quantum cosmology thus provides a simple testing ground for quantum general relativity concepts. Moreover, it can be used to explore qualitative features.

6.1 Hamiltonian formulation of classical cosmology

6.1.1 Motivation

The assumption of spatial homogeneity and isotropy yields the well-known Friedmann–Robertson–Walker line element. This is given by

$$\mathrm{d}s^2 = -\,\mathrm{d}t^2 + a^2(t)\,\mathrm{d}\Omega_k^2 \ ,$$

where $\mathrm{d}\Omega_k^2$ is the line element of a three-space with constant curvature $k = 0, \pm 1$,

$$\mathrm{d}\Omega_k^2 = \begin{cases} \mathrm{d}\psi^2 + \sin^2\psi\left(\mathrm{d}\theta^2 + sin^2\theta\,\mathrm{d}\phi^2\right) & k = 1 \\ \mathrm{d}x^2 + \mathrm{d}y^2 + \mathrm{d}z^2 & k = 0 \\ \mathrm{d}\psi^2 + \sinh^2\psi\left(\mathrm{d}\theta^2 + sin^2\theta\,\mathrm{d}\phi^2\right) & k = -1 \end{cases} .$$

There are several problems with the assumption of homogeneity and isotropy. First of all, this approximation applies only to large scales. It is very questionable why such a model — in which points correspond to galaxies! — could be used as starting point for a quantum framework. Secondly, already

on the classical level, it is not clear whether such an approximation can hold throughout the universe's evolution. It is surely not applicable in the vicinity of the big-bang singularity. Besides these theoretical concerns, there is the plain fact that the Friedmann–Robertson–Walker model is not able to explain the current, observed accelerated expansion of the Universe. To reproduce such an acceleration, one has to introduce exotic forms of matter — or a mysterious cosmological constant.

Nonetheless, I want to use this model in the following — not as a realistic device to extract predictions but as a playground to study the quality of quantum gravitational effects. Quantum general relativity models with a finite number of degrees of freedom, as the one described here, are called *minisuperspace models*.

6.1.2 Gravitational Hamiltonian

As discussed above, imposing spatial homogeneity and isotropy, the metric reduces to

$$g_{\alpha\beta} = \begin{pmatrix} -1 & 0 \\ 0 & a^2(t)\tilde{h}_{ab} \end{pmatrix} , \tag{6.1}$$

where $\tilde{h}_{ab}$ is the metric of a space of constant curvature. Comparing this to the ADM-metric (2.1), we find the ADM-form of the Friedmann–Robertson–Walker metric (6.1), the only non-vanishing components of which are

$$g_{00} = -N(t) , \qquad g_{ab} = a^2(t)\tilde{h}_{ab} = h_{ab} .$$

This can now be used to calculate the ADM-action (2.4). Using

$$K_{ab} = -\frac{\dot{a}}{aN} h_{ab} , \qquad {}^{(3)}R = \frac{6k}{a^2} , \qquad \sqrt{\det h} = a^3 ,$$

where a dot denotes derivatives with respect to t, one obtains

$$S^{grav}_{\mathrm{ADM}} = \frac{3}{\kappa^2} \int \mathrm{d}t \int\limits_{\Sigma} \mathrm{d}^3x \, N \left(-\frac{1}{N^2}\dot{a}^2 a + ka - \frac{\Lambda}{3}a^3 \right) - \Delta S_{\mathrm{ADM}} ,$$

where $\kappa^2 = \frac{8\pi G}{c^4} = 8\pi G$ as before. In the following, we will assume that the boundary term is made to vanish. For a closed FRW model, this follows trivially. For the flat case, we want to assume that space has been compactified to a torus. Carrying out the integration over a volume $\mathcal{V}$ of three-space, we arrive at the Lagrangian density

$$\mathcal{L}^{grav}_{\mathrm{ADM}} = CN \left[-\frac{1}{N^2}\dot{a}^2 a + ka - \frac{\Lambda}{3}a^3 \right] ,$$

where $C = \frac{3\mathcal{V}}{\kappa^2}$. The gravitational momenta are thus

$$p_a = \frac{\partial \mathcal{L}}{\partial \dot{a}} = -\frac{2C}{N}\dot{a}a \ , \quad p_N = \frac{\partial \mathcal{L}}{\partial \dot{N}} \approx 0 \ ,$$

yielding the Hamiltonian

$$\mathcal{H}^{grav} = N\mathcal{H}_\perp^{grav} = NC\left[-\frac{p_a^2}{4aC^2} - ka + \frac{\Lambda}{3}a^3\right] \ .$$

Note that the index on p_a here is no space index but just stands for the momentum of the scale factor a.

Conservation of the primary constraint $\dot{p}_N \approx 0$ yields the Hamiltonian constraint $\mathcal{H}_\perp^{grav} \approx 0$.

6.1.3 Coupling to scalar-field matter

We use a scalar field as toy model for matter. It thus replaces the perfect fluid usually used in cosmology. Such a replacement is necessary as a perfect fluid is just an effective description of matter. But on the quantum level, we need some fundamental field. Of course, a scalar field is not the best experimentally verified sort, but it is easiest to deal with. Moreover, we are only interested in the qualitative behaviour of the model, anyway, and in this respect a scalar field is an appropriate substitute for bosonic field components. The action for such a scalar field ϕ with arbitrary potential $V(\phi)$ is given by

$$S_{\mathrm{ADM}}^{mat} = \frac{1}{2}\int \mathrm{d}t \int\limits_\Sigma \mathrm{d}^3x\, N\sqrt{\det h}\left[-\ell g^{\alpha\beta}\frac{\partial\phi}{\partial x^\alpha}\frac{\partial\phi}{\partial x^\beta} - 2V(\phi)\right] \ ,$$

where the constant $\ell = \pm 1$ is introduced for later convenience. For an ordinary scalar field, $\ell = 1$. If $\ell = -1$, the field is a so-called phantom field, see Chapter 7. Using the ADM-form of the Friedmann–Robertson–Walker metric, this can be written as

$$S_{\mathrm{ADM}}^{mat} = \frac{1}{2}\int \mathrm{d}t \int\limits_\Sigma \mathrm{d}^3x\, Na^3\left[\ell\frac{\dot{\phi}^2}{N^2} + \ell h^{ab}\frac{\partial\phi}{\partial x^a}\frac{\partial\phi}{\partial x^b} - 2V(\phi)\right] \ .$$

After integration over the volume $\mathcal{V}$ of Σ, the scalar field momentum is derived

$$p_\phi = \ell\mathcal{V}a^3\frac{\dot{\phi}}{N} \ . \tag{6.2}$$

The matter part of the Hamiltonian is thus

$$\mathcal{H}^{mat} = N\mathcal{H}_{\perp}^{mat} = \frac{N\mathcal{V}a^3}{2}\left[\ell\frac{p_\phi^2}{\mathcal{V}^2a^6} + h^{ab}\frac{\partial\phi}{\partial x^a}\frac{\partial\phi}{\partial x^b} + 2V(\phi)\right] .$$

The energy-momentum tensor for the scalar field is defined in the usual way as

$$T_{\alpha\beta} = -\frac{2}{\sqrt{-\det g}}\frac{\delta\mathcal{L}^{mat}}{\delta g^{\alpha\beta}} .$$

In the *homogeneous* case, it has only four non-vanishing components

$$T_{00} = \ell\frac{\dot\phi^2}{2} + NV(\phi) , \qquad T_{bb} = a^2\left(\ell\frac{\dot\phi^2}{2N^2} - V(\phi)\right) .$$

Comparing this with the perfect fluid energy-momentum tensor for a co-moving observer[1]

$$T_{\alpha\beta} = \rho n_\alpha n_\beta + p\left(g_{\alpha\beta} + n_\alpha n_\beta\right) ,$$

we can assign an energy density ρ and a pressure p to the scalar field

$$\rho = \ell\frac{\dot\phi^2}{2} + V(\phi) , \qquad p = \ell\frac{\dot\phi^2}{2} - V(\phi) . \tag{6.3}$$

The equation of state is then defined as the ratio of pressure to energy density $w = \frac{p}{\rho}$. Do not confuse the pressure p with the momenta p_a and p_ϕ. Assuming a constant equation of state parameter w for the scalar field, we can use the equations for ρ and p to find the relation between the scalar field and its potential,

$$V(\phi(t)) = \frac{\ell}{2}\frac{1-w}{1+w}\dot\phi^2(t) , \tag{6.4}$$

where $w \neq -1$. This is analogous to the virial theorem in which the kinetic energy is proportional to the potential energy of the field.

6.1.4 Total Hamiltonian for Friedmann–Robertson–Walker cosmology coupled to a homogeneous scalar field

To fit the symmetry conditions of the Friedmann–Robertson–Walker model, the scalar field should not depend on the spatial coordinates, $\phi = \phi(t)$. The total Hamiltonian is then given by

[1] Recall that $n = (-1, 0, 0, 0)$ is the hypersurface normal and N has to be fixed to $N = 1$, see Section 6.1.5 below.

$$\mathcal{H} = N\mathcal{H}_\perp = N\left[-\frac{p_a^2}{4aC} + \ell\frac{p_\phi^2}{2\mathcal{V}a^3} - \mathcal{K}a + \frac{\lambda}{3}a^3 + a^3\mathcal{V}V(\phi)\right] , \qquad (6.5)$$

where $\mathcal{K} = Ck$ and $\lambda = C\Lambda$. The total Hamiltonian is constrained to vanish as a consequence of the conservation of the primary constraint. From the kinetic term, we can read off the inverse of the two-dimensional analogue of the DeWitt metric

$$G^{AB} = \begin{pmatrix} -\frac{1}{4aC} & 0 \\ 0 & \frac{\ell}{2\mathcal{V}a^3} \end{pmatrix} .$$

Note that this inverse exists everywhere except at $a = 0$.

6.1.5 Hamiltonian equations of motion

To make the connection to Einstein's equations, the equations of motion in the Hamiltonian framework are briefly discussed. In order that the equations of motion can be solved — or acquire any meaning in the first place, it is necessary to fix the foliation. This is done by fixing the lapse function. For $N = 1$ we obtain Friedmann time t, whereas the choice $N = a$ yields conformal time τ. We will use $N = 1$.

Now note that the Hamiltonian itself, when expressed in terms of velocities, is just the Friedmann equation

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{\kappa^2}{3}\rho - \frac{k}{a^2} + \frac{\Lambda}{3} . \qquad (6.6)$$

Note that this constraint is of the form $E_{\rm kin} + E_{\rm pot} = 0$ with

$$E_{\rm kin} = -\left(\frac{\dot{a}}{a}\right)^2 + \ell\frac{\kappa^2}{6}\frac{\dot{\phi}^2}{2} , \qquad E_{\rm pot} = \frac{\kappa^2}{6}V(\phi) - \frac{k}{a^2} + \frac{\Lambda}{3} . \qquad (6.7)$$

For $\ell = -1$, the 'kinetic energy' is thus negative. Consequently, the 'potential energy' therefore has to be positive (or zero), $E_{\rm pot} \geq 0$. The region $E_{\rm pot} < 0$ is classically forbidden for $\ell = -1$.

The equations of motion determining the evolution of the scale factor are

$$\begin{aligned} \dot{p}_a &= \{p_a, \mathcal{H}_\perp\} = -\frac{p_a^2}{4a^2C} + \mathcal{K}a - \lambda a^2 + 3\ell\frac{p_\phi^2}{2\mathcal{V}a^4} - 3\mathcal{V}a^2V(\phi) , \\ \dot{a} &= \{a, \mathcal{H}_\perp\} = -\frac{\dot{p}_a}{2aC} - \frac{p_a^2}{4a^3C^2} . \end{aligned}$$

Differentiating the second equation with respect to t and using the first yields the well-known evolution equation

$$\frac{\ddot{a}}{a} = \frac{\Lambda}{3} - \frac{\kappa^2}{6}(\rho + 3p) \ . \tag{6.8}$$

For the scalar field, one obtains in the same way

$$\dot{p}_\phi = -a^3 \mathcal{V} V'(\phi) \ , \qquad \dot{\phi} = \ell \frac{p_\phi}{\mathcal{V} a^3} \ ,$$

where a prime shall denote derivation with respect to ϕ here. This yields

$$\ddot{\phi} + 3\frac{\dot{a}}{a}\dot{\phi} + \ell V'(\phi) = 0 \ , \tag{6.9}$$

which is equivalent to the conservation equation

$$\dot{\rho} + 3H(\rho + p) = 0 \ . \tag{6.10}$$

In the following sections, I will also use the Hubble parameter $H = \frac{\dot{a}}{a}$ to keep expressions short and simple.[2]

6.2 Wheeler–DeWitt equation for cosmology

Quantization now consists in the promotion of scale factor, scalar field and their momenta to operators. These operators shall act on the wave function $\psi(a, \phi)$ in the following way

$$\begin{aligned} \hat{p}_a \psi(a,\phi) &= -\mathrm{i}\hbar \frac{\partial \psi}{\partial a} \ , \quad \hat{a}\psi(a,\phi) = a \cdot \psi(a,\phi) \ , \\ \hat{p}_\phi \psi(a,\phi) &= -\mathrm{i}\hbar \frac{\partial \psi}{\partial \phi} \ , \quad \hat{\phi}\psi(a,\phi) = \phi \cdot \psi(a,\phi) \ . \end{aligned}$$

Here we assume that ψ is sufficiently differentiable, but nothing else. The Hamiltonian constraint is then turned into the Wheeler–DeWitt equation. There is some factor-ordering ambiguity in the gravitational kinetic term. Parametrizing this ambiguity by arbitrary real numbers r, m, the Wheeler–DeWitt equation becomes

$$\frac{\hbar^2}{2}\left[\frac{a^{r+m-1}}{2C}\frac{\partial}{\partial a}a^{-r}\frac{\partial}{\partial a}a^{-m} - \frac{\ell}{\mathcal{V}a^3}\frac{\partial^2}{\partial \phi^2}\right]\psi(a,\phi) + V(a,\phi)\psi(a,\phi) = 0 \ ,$$

where

[2] The letter H denotes several different quantities in this thesis. In order to distinguish the Hubble parameter from the Hamiltonian constraint, I use roman, H, and calligraphic, $\mathcal{H}$, H for the Hamiltonian and simply H for the Hubble parameter.

$$V(a,\phi) = -\mathcal{K}a + \frac{\lambda}{3}a^3 + a^3\mathcal{V}V(\phi) \ .$$

Choosing the Laplace–Beltrami factor ordering, we obtain

$$\frac{\hbar^2}{2}\left[\frac{1}{2C}\left(\frac{1}{a^2}\frac{\partial}{\partial a} + \frac{1}{a}\frac{\partial^2}{\partial a^2}\right) - \frac{\ell}{a^3}\frac{\partial^2}{\partial\phi^2}\right]\psi(a,\phi) + V(a,\phi)\psi(a,\phi) = 0 \ . \tag{6.11}$$

6.2.1 Regularization of the Wheeler–DeWitt equation

As described in Chapter 5, Section 5.4.3, a regularization of the Wheeler–DeWitt equation at $a = 0$ can be obtained through the introduction of a new variable α which moves the singularity out to infinity,

$$\alpha = \ln\left(\frac{a}{a_{\rm ref}}\right) \ ,$$

where $a_{\rm ref}$ is some constant with length dimension.[3] With this new variable, (6.11) becomes

$$\frac{\hbar^2}{2}a_{\rm ref}^{-3}e^{-3\alpha}\left[\frac{1}{2C}\frac{\partial^2}{\partial\alpha^2} - \frac{\ell}{\mathcal{V}}\frac{\partial^2}{\partial\phi^2}\right]\psi(\alpha,\phi) + V_{a_{\rm ref}}(\alpha,\phi)\psi(\alpha,\phi) = 0 \ .$$

After multiplication with the non-zero quantity $e^{3\alpha}$, we obtain

$$\frac{\hbar^2}{2a_{\rm ref}^3}\left[\frac{1}{2C}\frac{\partial^2}{\partial\alpha^2} - \frac{\ell}{\mathcal{V}}\frac{\partial^2}{\partial\phi^2}\right]\psi(\alpha,\phi) + e^{3\alpha}V_{a_{\rm ref}}(\alpha,\phi)\psi(\alpha,\phi) = 0 \ , \tag{6.12}$$

where

$$V_{a_{\rm ref}}(\alpha,\phi)\psi(\alpha,\phi) = -\mathcal{K}a_{\rm ref}e^{\alpha} + \frac{\lambda}{3}a_{\rm ref}^3e^{3\alpha} + a_{\rm ref}^3\mathcal{V}e^{3\alpha}V(\phi) \ .$$

Choosing now units such that $\mathcal{V} = 1$ and $C = 1$, i.e. $\kappa^2 = 6$, and $a_{\rm ref} = 1$, the Wheeler–DeWitt equation becomes

$$\frac{\hbar^2}{2}\left[\frac{\partial^2}{\partial\alpha^2} - \ell\frac{\partial^2}{\partial\phi^2}\right]\psi(\alpha,\phi) + e^{3\alpha}V_1(\alpha,\phi)\psi(\alpha,\phi) = 0 \ . \tag{6.13}$$

Note that the Wheeler–DeWitt equation is a hyperbolic equation in the case of a standard scalar field ($\ell = 1$). This structure implies that the scale factor should serve as evolution parameter. Initial conditions should be prescribed

[3] In Chapter 5, Section 5.4.1, we used the letter X to denote this variable. The choice of α here follows the conventional quantum cosmology literature.

on surfaces of constant α. For the phantom field ($\ell = -1$), the equation is no longer hyperbolic but elliptic. If other types of matter are present in addition to the phantom field, the equation becomes ultra-hyperbolic.

In the following two chapters, $\mathcal{V} = 1$ throughout. The constants $a_{\rm ref}$ and κ^2 are kept.

Chapter 7

Quantum phantom cosmology

Phantom dark energy is still an experimentally allowed candidate for dark energy. A characteristic feature of classical cosmological models with this sort of dark energy is the occurrence of a new type of singularity, the so-called big-rip singularity. In contrast to the familiar big-bang singularity, the big rip 'takes place' in the macroscopic universe. The question therefore arises what happens to such a singularity in quantum cosmology? More pointedly, can we expect quantum gravitational effects in the macroscopic universe?

7.1 Phantom dark energy — motivation and general properties

In 1998, supernovae 1a (SNe1a) data suggested for the first time that the Universe might undergo an accelerated expansion, [50]. This was confirmed by a second group in the following year, [51]. From the evolution equation (6.8), it can be read off that the matter component has to satisfy $\rho + 3p < 0$ in order to drive such an accelerated expansion. This implies a violation of the strong energy condition. Introducing the equation of state parameter $w = \frac{p}{\rho}$, acceleration requires $w < -\frac{1}{3}$. The unknown type of matter driving the acceleration of our Universe is called *dark energy*.

What comes to mind first, is the cosmological constant Λ with $w = -1$.[1] This has since then been the most promising and most conservative candidate for dark energy. The serious problem this proposal suffers from, is that no one is able to explain what this cosmological constant may be. The suggestion is that it may just be the vacuum energy density. But for this, the

[1] This follows also from (6.8).

measured value of Λ is far too high. Nonetheless, the so-called ΛCDM-model with dark energy given by an unexplained cosmological-constant term, Λ, and a Cold Dark Matter component, is the current standard model of cosmology.

In 2002, Caldwell suggested to look beyond the $w = -1$ barrier, [55]. He showed that dark energy with $w < -1$, but $\rho > 0$, is indeed consistent with the supernovae data of Riess et al. and Perlmutter et al. [50, 51], and the cosmic microwave background (CMB) data available at that time.[2] He called such an energy component *phantom* and the corresponding equation of state $w < -1$ *super-negative*. Such a super-negative w implies that not only the strong but also the weak and dominant energy conditions are violated. This means that observers exist that measure a negative energy density of the phantom fluid and a phantom energy flow faster than light. From a *theoretical* point of view, this is clearly an odd suggestion.

I first want to discuss the cosmological evolution driven by such dark energy, before exploring the possible ways in which such matter could be realized on the fundamental level.

First, note that (6.10) implies that the energy density grows as $\rho \sim a^{3|1+w|}$ with the scale factor. Phantom thus dominates *late* over matter. As soon *as* it dominates, the Universe accelerates more rapidly than with cosmological constant. This epoch of phantom dominance starts the later the more negative w. The age, and correspondingly the horizon distance, are larger than in the Λ case. But this effect is less and less important, the more negative w becomes — for a fixed value of the matter density. Caldwell moreover calculated characteristic, measurable quantities for a phantom model with $w = -\frac{3}{2}$ and compared it to quintessence and cosmological constant characteristics.

First of all, the volume-redshift relationship in a phantom universe differs of course from that of ΛCDM. The phantom predicts a larger differential number of objects per redshift interval. This leads to more strong gravitationally lensed quasars than ΛCDM. A comparison of the magnitude-redshift relationship with the *binned* data[3] of [50, 51] shows that the phantom with $w = -\frac{3}{2}$ is in best accord with data — *apart from one single data point at redshift larger than one*. But even for this point, the phantom fit is better than the one obtained for Λ. Using the constraints from SNe1a data on the matter density and w, phantom is also in the admissible range. A comparison with the CMB-anisotropy spectrum shows that phantom diminishes the lower multipoles due to a weaker late-time integrated Sachs-Wolffe effect. For low multipoles it is thus *not* in best accord with data. The acoustic peaks, on the other hand, get shifted to higher multipoles, thus fitting

[2] Caldwell used the compilation of CMB data provided by Wang, Tegmark and Zaldarriaga including data from BOOMERaNG, DASI, Maxima and CBI, [52].

[3] Binning was carried out in [56].

the data *better* than the ΛCDM model. Due to a larger amplitude of the growth suppression factor, the perturbations persist longer and stronger in the phantom models.

Most significantly, a phantom-driven universe is finite. Let the time of phantom dominance be denoted by t_p, then

$$a(t) = a(t_p)\left[-w + (1+w)\frac{t}{t_p}\right]^{\frac{2}{3}(1+w)}$$

for $t > t_p$. Clearly, the scale factor diverges after a finite time. For a sufficiently large matter density, this divergence arises well after the present epoch. This divergence of the scale factor after a finite time provides a singular end of the universe called the *big-rip singularity*. It is the growing energy density of a fluid with negative pressure whose repulsion 'rips' the universe apart. In a follow-up of the first paper, Caldwell et al. calculated the time at which different structures in the universe are torn apart for a fixed value of w, [57]. Calculation of the perturbation spectrum caused by phantom energy makes it necessary to introduce a field-theoretical description of the phantom fluid. In order to keep the energy dark and produce the desired equation of state, Caldwell used a scalar field with reversed sign of the kinetic energy term. This just corresponds to the choice $\ell = -1$ in the formulas of the previous Chapter 6. It follows from (6.9) that the field will run *up* and not down the potential. As a justification for such a weird kind of matter, Caldwell cites supergravity theories and higher-derivative theories of gravity.

Caldwell's idea was taken up in the following years. The High-z Supernovae Search Team included in their data analysis the probability contours for w over the matter density. They found that $-1.48 < w < -0.72$ at a 95% confidence level, [53]. In comparison to the Riess and Perlmutter data of 1998/1999, these new samples contain also SNe1a at redshifts z larger than $z = 1$. Another analysis combined the data from the Sloan Digital Sky Survey (SDSS) with the data of the Wilkinson Microwave Anisotropy Probe (WMAP) of 2004, [54].[4] They used a six-parameter model but also allowed w to be a free parameter in one of their evaluations. Using just WMAP temperature and polarization information, they found $w > -1$. But including SDSS data, they obtained $w = -1.05\ ^{+0.13}_{-0.14}$ which one might interpret as a slight indication for phantom dark energy.

New data, including more high-redshift SNe1a (which was the weak region of the phantom model already in Caldwell's first paper), however, are not able to rule out but clearly disfavour the phantom model. The SDSS Supernovae Legacy currently provides the largest and qualitatively best supernovae data set. They found $w = -0.969\ ^{+0.059}_{-0.063}\ (stat)\ ^{+0.063}_{-0.060}\ (sys)$, [58].

[4] During the SDSS-I, power spectra of over 200 000 galaxies were recorded over five years from 2000 until 2005.

The restrictions on the dark energy equation of state were also determined by the WMAP Science Team using the five-year data. They give as limit $-1.14 < w < -0.88$, [49]. These are the most recent constraints on the dark energy equation of state, both published in October 2008.

So, at the time when the following research was carried out, the phantom model was an *experimentally* motivated model. Despite its physical flaws, like the violation of energy conditions, it was found to be in accord with the data and therefore received much interest. In a rather short amount of time, data were put forward that disfavoured this model, bringing physical common sense and experimental data together again. As a side, I may learn that it is sometimes wiser to view experimental data with theoretical prejudice.

However, there are also reasons from the *quantum* cosmological side to study such a phantom universe.

7.2 Quantizing the big-rip singularity

As explained above, phantom dark energy produces a new type of cosmological singularity at which scale factor, energy density and pressure diverge after a finite amount of time. This big-rip singularity thus occurs at *large* scale factor and thus in the *macroscopic* universe. Nonetheless energy density and pressure traverse the Planck scale. At this scale, we assume that quantum gravity effects become important. The motivation standing behind the quantisation of phantom cosmologies is the study of quantum cosmological models with singularity in the macroscopic universe. Are there quantum gravity effects at large scale factor and what is the possible impact of quantum gravity on the new singularity? In particular, we want to see whether this macroscopic big-rip singularity is avoided in the quantum cosmological model.

Moreover, the introduction of a phantom field changes the nature of the Wheeler–DeWitt equation. If no other fields but the phantom are present, it becomes elliptic. This influences the type of boundary conditions to be imposed on the wave function.

7.3 Classical cosmological models with phantom dark energy

In the following, three models will be discussed. The first of these is just a toy model exhibiting some 'phantom features'. The next two are genuine phantom models. The first contains a phantom field with exponential potential, thus yielding a universe ending with a big rip. The second model contains a negative cosmological constant in addition to a phantom field with hyperbolic-cosine potential. The conditions in the latter model are chosen

such that the universe evolves *from big rip to big rip*. None of the models is chosen with the aim to fit cosmological data but to provide solvable and interesting models on the *quantum* level. This means that they should end or start their evolution with a big-rip singularity. For all models, ℓ will be kept as a parameter, allowing the comparison of ordinary scalar field matter with phantom.

There are two steps which are carried out on the classical level with foresight on the quantum analysis. First of all, as pointed out already in the previous section, the phantom field will not be described by a perfect fluid Lagrangian but by a fundamental field. As a substitute for this will serve a scalar field with reversed sign of the kinetic energy term. Secondly, the classical trajectories will ultimately be given in configuration space, i.e. in the (a, ϕ)-plane. This is the space on which the quantum cosmological wave function is defined — the minisuperspace of this model. The classical trajectory *in configuration space* is the only link we have between the classical model and the quantum theory.

7.3.1 Cosmological toy model with vanishing phantom potential

In this section we shall consider a simple model with field potential $V(\phi) = 0$ and cosmological constant $\Lambda = 0$. This leads to an equation of state for stiff matter, $p = \rho$, $w = 1$. Moreover, in such a case the energy density $\rho < 0$, and thus this model does not seem to represent dark energy which is usually assumed to have *positive* energy density. However, it captures interesting 'phantom features', since it violates all energy conditions, and it has the merit that it is easily manageable.

It follows from the remarks after (6.7) that for vanishing potential and $\ell = -1$, the curvature index k has to be chosen negative, $k = -1$, in order that a classical solution exists at all.

Since ϕ is a cyclic variable, p_ϕ is constant, so from (6.2) one has $\dot{\phi}^2 = P_\phi^2/a^6$ with a constant P_ϕ. From (6.6) together with (6.3) one then has

$$\frac{\mathrm{d}\phi}{\mathrm{d}a} = \pm \frac{P_\phi}{a\sqrt{a^4 + \ell\frac{\kappa^2}{6}P_\phi^2}} \ ,$$

which for $\ell = -1$ and the choice $P_\phi > 0$ integrates to

$$\phi(a) = \pm\sqrt{\frac{3}{2}}\frac{1}{\kappa}\arccos\frac{P_\phi\kappa}{\sqrt{6}a^2} \ . \tag{7.1}$$

The classical trajectory (7.1) has a minimum value of the scale factor, $a_{\min} = \sqrt{\frac{P_\phi\kappa}{\sqrt{6}}}$, and reaches infinite values of a at finite values of $\phi = \pm\pi/4$. In this sense, it resembles a big-rip solution. However, with respect to t, the scale factor reaches infinity only at $t = \pm\infty$ and, moreover, $\rho \propto a^{-6}$ which is

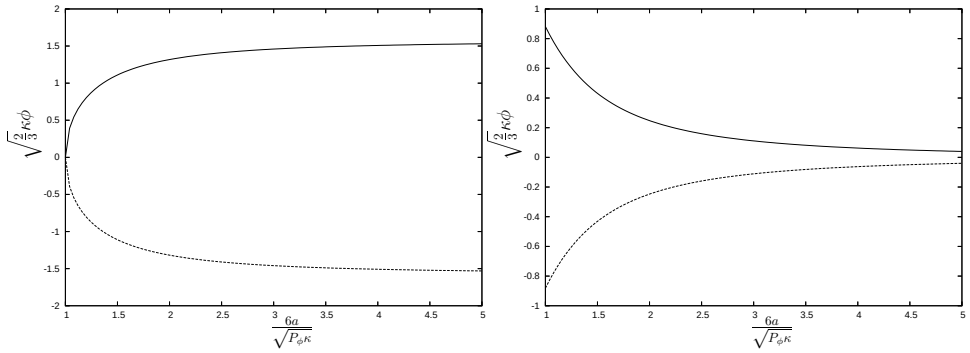

Figure 7.1: The classical trajectory in configuration space for the toy model with vanishing scalar field potential and vanishing cosmological constant. The diagram on the left hand side shows the trajectory for the phantom field model. On the right-hand side the trajectory for the $\ell = 1$ scalar field model is plotted.

the density scaling appropriate to a stiff fluid. Nonetheless, in configuration space the trajectory has some features of a big rip, and this is why this toy model is of interest.

For an ordinary scalar field ($\ell = 1$) and for $k = -1$, one gets instead of (7.1),

$$\phi(a) = \pm\sqrt{\frac{3}{2}}\frac{1}{\kappa}\mathrm{arcsinh}\frac{P_\phi\kappa}{\sqrt{6}a^2} \ . \tag{7.2}$$

There is no turning point. Equation (7.2) just describes two branches for which $a \to \infty$ if $\phi \to 0$, and $a \to 0$ if $\phi \to \pm\infty$. The two solutions (7.1) and (7.2) are depicted in Figure 7.1.

7.3.2 Cosmological model with phantom dark energy

A model with phantom equation of state and a *true* big-rip singularity appears if the dark-energy potential is chosen exponential,

$$V(\phi) = V_0 e^{-\gamma\kappa\phi} \ , \tag{7.3}$$

with $\gamma > 0$ and $\Lambda = 0$. Interest in this type of scalar field potentials in cosmology arose when it became clear that the classical model has an attractor solution with scalar-field domination, [61]. This alleviates the fine-tuning problem of the initial energy of the scalar field. Such an attractor exists not only in the case of a conventional scalar field, but also for the phantom field [59, 60]. At this point, cosmologists quote that exponential potentials for scalar fields arise in the context of Kaluza–Klein theories, higher-derivative gravity in higher dimensions, higher-order gravity, supergravity and superstring theories, see [62] for an overview.

In the following, we shall consider the case of a flat universe, $k = 0$. From the discussion after (6.7) one sees immediately that for this choice of parameters, neither the ordinary scalar field nor the phantom field model possesses classically forbidden regions.

To find a solution to the equations of motion which actually produces phantom behaviour, the system of equations (6.8) and (6.9) is most conveniently transformed into an autonomous system. This system evolves under the Friedmann equation (6.6) as constraint. Solutions with constant equation of state parameter w are fix points of this system. The autonomous system is obtained for the new variables

$$x = \frac{\kappa}{\sqrt{6}}\frac{\dot{\phi}}{H} = \frac{\kappa}{\sqrt{3}}\frac{\mathrm{d}\phi}{\mathrm{d}\alpha} , \qquad y = \frac{\kappa}{\sqrt{3}}\frac{\sqrt{V}}{H} ,$$

where a dot is used to denote differentiation with respect to t. These are considered as functions of $\alpha = \ln\left(\frac{a}{a_{\mathrm{ref}}}\right)$. The constraint (6.6) is then given by

$$1 = \ell x^2 + y^2 ,$$

and the evolution equations (6.8) and (6.9) are equivalent to

$$\begin{aligned} x' &= -3x + 3\ell x^3 + \sqrt{\frac{3}{2}}\ell\gamma y^2 , \\ y' &= xy\left(3x - \sqrt{\frac{3}{2}}\ell\gamma\right) , \end{aligned}$$

where primes denote derivation with respect to the variable α. The equation of state of the matter field is now

$$w = \frac{\ell x^2 - y^2}{\ell x^2 + y^2} \tag{7.4}$$

and we see that it is constant for fix points of the above system. There are only two fix points for non-trivial potential. These are given by

$$x = \ell\frac{\gamma}{\sqrt{6}} , \qquad y = \pm\sqrt{1 - \ell\frac{\gamma^2}{6}} . \tag{7.5}$$

For $y > 0$, the universe is expanding, for $y < 0$ contracting — this depends on the initial value $H_0 = H(t_0)$ of the Hubble parameter. We consider here the case of an expanding universe. For $\ell = -1$ and arbitrary values of γ, as well as for $\ell = +1$ and $\gamma < \sqrt{6}$, the above is an *attractor* solution of the system.

From (7.5), we obtain the equation for ϕ

$$\phi(\alpha) = \ell\frac{\gamma}{\kappa}(\alpha - \alpha_0) + \phi_0 \ , \tag{7.6}$$

where $\alpha_0 = \ln\left(\frac{a_0}{a_{\rm ref}}\right)$ is the initial value of α. We choose $\phi_0 = \ell\frac{\gamma}{\kappa}\alpha_0$. With the knowledge of (7.6), the equation for y can be integrated to yield

$$\frac{a(t)}{a_0} = \left[1 + \ell\frac{\gamma^2 H_0}{2}(t - t_0)\right]^{\ell\frac{2}{\gamma^2}} \ . \tag{7.7}$$

The corresponding equation for $\phi(t)$ is then

$$\phi\,(t) = \frac{2}{\gamma\kappa}\ln\left[1 + \ell\frac{\gamma^2 H_0}{2}\,(t - t_0)\right] . \tag{7.8}$$

For this attractor solution, the 'kinetic energy' is given by

$$E_{\rm kin} \equiv \frac{\kappa^2}{6}\left(\frac{d\phi}{d\alpha}\right)^2 = \frac{\gamma^2}{6} \ ,$$

and thus constant. Therefore, also the 'potential energy' of the scalar field is constant,

$$E_{\rm pot} \equiv \frac{\kappa^2 V}{3H^2} = 1 - \frac{\ell\gamma^2}{6} \ .$$

Inserting the fix-point values into the equation of state (7.4), one immediately sees that

$$w = -1 + \ell\frac{\gamma}{6} \ ,$$

which, for $\ell = -1$, obviously describes a phantom field, whereas the scalar field with $\ell = 1$ covers the range $w > -1$. Accordingly, the energy density scales as $\rho = \rho_0\left(\frac{a}{a_0}\right)^{-\ell\gamma^2}$. As expected, this yields a big-rip singularity for $\ell = -1$. In the limit $t \to t_1 \equiv t_0 - 2\ell/(\gamma^2 H_0)$ the energy density and the scale factor diverge. For $t \to \infty$, a and ρ vanish.

This is opposed to the $\ell = 1$ model. In the limit $t \to t_1$, a goes to zero and ρ diverges, yielding a big-bang singularity, whilst for $t \to \infty$, a diverges and ρ goes to zero.

7.3.3 Cosmological model evolving from big rip to big rip

We now introduce a negative cosmological constant in addition to the phantom field. The attraction of the negative Λ counteracts the repulsion of the phantom. This allows a model which evolves symmetrically between two big rips. Again, the universe is assumed to be flat, $k = 0$, but $\Lambda < 0$. Now

we assume a constant equation of state from the outset. In this case, the energy conservation equation gives

$$\rho = \rho_0 \left(\frac{a}{a_0}\right)^{-3(w+1)} ,$$

and $p_\Lambda = -\rho_\Lambda = -\frac{\Lambda}{\kappa^2}$ for the cosmological constant holds. This can be used to solve the system (6.6) and (6.8) in terms of the scale factor as

$$a(t) = a_0 \left(\frac{\rho_0}{|\rho_\Lambda|}\right)^{\frac{1}{2D}} [\sin X]^{\frac{1}{D}} , \tag{7.9}$$

where

$$X = \frac{|D|}{\sqrt{3}}\sqrt{|\Lambda|}(t - t_0) + \arcsin\left(\sqrt{\frac{|\rho_\Lambda|}{\rho_0}}\right) ,$$

and $D = \frac{3}{2}(1 + w)$. Using the short-hand D, we can rewrite (6.4) in the form

$$V(\phi(t)) = \frac{\ell}{2}\frac{3 - D}{D}\dot{\phi}^2(t) , \tag{7.10}$$

which allows to write the energy density in (6.3) as

$$\rho = \frac{3\ell}{2D}\dot{\phi}^2 = \frac{3}{\kappa^2}\left(\frac{\dot{a}^2}{a^2} - \frac{\Lambda}{3}\right) . \tag{7.11}$$

With all these assumptions we are able to calculate the evolution of the scalar field as

$$\phi(t) = \pm\frac{\sqrt{2}}{\kappa}\frac{\sqrt{\ell D}}{D}\ln\left[\tan\left(\frac{X}{2}\right)\right] , \tag{7.12}$$

For simplicity, we choose t_0 such that $a(t_0 = 0) = 0$. Then X in (7.12) and in (7.9) simplifies to $X = \frac{|D|}{\sqrt{3}}\sqrt{|\Lambda|}t$. For later reference, let us furthermore introduce the short-hand $A = a_0^D\sqrt{\frac{\rho_0}{|\rho_\Lambda|}}$. Note that $D/\ell = |D| > 0$.

For the model with ordinary scalar field, $D > 0$. The evolution of the universe based on (7.9) begins with a big bang at $t = 0$, reaches a maximum $a_{\max} = A^{1/D}$, and terminates with a big crunch at $t = \pi$.

For the phantom case, $D = -|D| < 0$. The evolution starts with a big rip at $t = 0$, reaches a minimum $a_{\min} = A^{-1/|D|}$, and terminates with a big rip at $t = \pi$. Thus we have a *symmetric evolution of the scale factor from big rip to big rip*. The corresponding trajectory therefore has a turning point. This is of special interest in the quantum theory.

We use (7.9) to eliminate the classical time coordinate in (7.12) and obtain the trajectory in configuration space

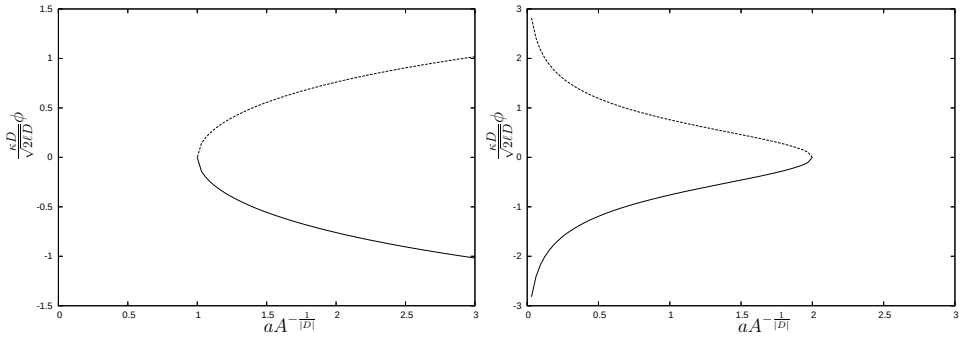

Figure 7.2: The classical trajectories in configuration space for the models with cosh-potential and negative cosmological constant. On the left-hand side, the trajectory for the phantom field model is shown. The classical trajectory for the scalar field model $\ell = 1$ is shown on the right-hand side. The similarity to the classical trajectories in the toy model in Section 7.3.1 is obvious.

$$\phi(a) = \pm\frac{\sqrt{2}}{\kappa}\frac{\sqrt{\ell D}}{D}\ln\left(\frac{\theta}{1+\sqrt{1-\theta^2}}\right) , \tag{7.13}$$

where $\theta = a^D A^{-1}$. This trajectory has two branches. For $\ell = -1$ each of them extends to infinity, that is, $\phi \to \pm\infty$, for $a \to \infty$ and reaches a minimum $\phi(a) = 0$, for $a_{\min} = A^{-1/|D|}$. For $\ell = 1$ one recognizes the presence of the maximum $a_{\max}$. The trajectories in configuration space are depicted in Figure 7.2.

From (7.12) and (7.10), one can reconstruct the potential of the scalar field,

$$V(\phi) = V_0 \cosh^2\left(\sqrt{\frac{3}{\ell D}}D\phi\right) , \tag{7.14}$$

where $V_0 = \frac{(D-3)}{18}\Lambda$. To make the expression simple, I will use the shorthand $F^{-1} = \ell\sqrt{3|D|}$. Note that for $\ell = 1$ the potential is positive only for $D < 3$ (i.e., $w < 1$). This restriction is similar to the restriction $\lambda < \sqrt{6}$ in Section 7.3.2.

7.4 Quantum Cosmological models with phantom dark energy

We now subject the previously presented models to quantization as presented in Chapter 6.

7.4.1 Quantum cosmological toy model with vanishing phantom potential

We start from the Wheeler–DeWitt equation with Laplace–Beltrami factor ordering, (6.11), where for the toy model $V(\phi) = 0$, $\Lambda = 0$ and $k = -1$. Multiplication by a^3 (or, more correctly, regularizing, multiplying be $e^{3\alpha}$ and then transforming back to a), we get the Wheeler–DeWitt equation

$$\frac{\hbar^2}{2}\left[\frac{\kappa^2}{6}\left(a\frac{\partial}{\partial a}+a^2\frac{\partial^2}{\partial a^2}\right)-\ell\frac{\partial^2}{\partial\phi^2}\right]\psi(a,\phi)+\frac{3}{\kappa^2}a^4\psi(a,\phi)=0\ . \quad (7.15)$$

A solution to this equation is found from the separation ansatz,

$$\psi_n(a,\phi)=C_n(a)\varphi_n(\phi)\ . \quad (7.16)$$

This yields

$$\varphi_n(\phi)=e^{-\frac{\mathrm{i}}{\hbar}n\phi}\ . \quad (7.17)$$

We choose the oscillating solution, because real exponentials would yield exponentially growing solutions as $\phi\to\pm\infty$. The equation for the gravitational part is then given by

$$a^2\ddot{C}_n+a\dot{C}_n+\frac{36}{\hbar^2\kappa^4}(a^4+\ell\tilde{n}^2)C_n=0\ , \quad (7.18)$$

where $\tilde{n}^2=\frac{\kappa^2}{3}n^2$. Solutions of this equation are Bessel functions $Z_{\sqrt{-\ell}\frac{3n}{\hbar\kappa}}(\frac{3a^2}{\hbar\kappa})$. However, we have to impose the boundary condition that $\psi(a,\phi)\stackrel{a\to0}{\longrightarrow}0$ in order to reflect the behaviour of the classical trajectories (7.1) which have a minimum with respect to a in configuration space. We therefore have to choose the Bessel function $J_{\sqrt{-\ell}\frac{3n}{\hbar\kappa}}(\frac{3a^2}{\hbar\kappa})$ with $n>0$.

Let us first discuss the phantom case, $\ell=-1$. The connection to the classical solution (7.1) should be performed through a formal WKB-limit '$\hbar\to0$'. We thus have to look for an asymptotic expansion of $J_\mu(z)$ where both the argument and the index are large. We use relation

$$J_\nu(\nu z)=\left(\frac{4\zeta}{1-z^2}\right)^{1/4}\left(\frac{\mathrm{Ai}(\nu^{2/3}\zeta)}{\nu^{1/3}}+\frac{\exp(-\frac{2}{3}\nu\zeta^{3/2})}{1+\nu^{1/6}|\zeta|^{1/4}}\,\mathcal{O}\left(\frac{1}{\nu^{4/3}}\right)\right)\ , \quad (7.19)$$

see [64] and set $\nu=\frac{3n}{\hbar\kappa}$, $z=\frac{a^2}{n}$. The choice for ζ depends on whether $z^2\geq1$ or $z^2<1$,

$$\frac{2}{3}\left(-\zeta\right)^{\frac{3}{2}} = \sqrt{z^2-1} - \arccos\left(\frac{1}{z}\right) \qquad \text{for} \quad z^2 \geq 1 \ , \quad (7.20)$$

$$\frac{2}{3}\zeta^{\frac{3}{2}} = \ln\left(\frac{1+\sqrt{1-z^2}}{z}\right) - \sqrt{1-z^2} \quad \text{for} \quad z^2 < 1 \ , \quad (7.21)$$

see [64]. Now expand the Airy function for large argument $|\tilde{z}|$,

$$\mathrm{Ai}\left(\tilde{z}\right) \sim \frac{1}{2}\tilde{z}^{-\frac{1}{4}}\exp\left(-\frac{2}{3}\tilde{z}^{\frac{3}{2}}\right)\sum_{m=0}^{\infty}(-1)^m c_m\left(\frac{2}{3}\tilde{z}^{\frac{3}{2}}\right)^{-m} \qquad \text{for} \quad |\arg\tilde{z}| < \pi \ , \quad (7.22)$$

$$\mathrm{Ai}\left(-\tilde{z}\right) \sim \pi^{-\frac{1}{2}}\tilde{z}^{-\frac{1}{4}}\left[\sin\left(\frac{2}{3}\tilde{z}^{\frac{3}{2}}+\frac{\pi}{4}\right)\sum_{m=0}^{\infty}(-1)^m c_{2m}\left(\frac{2}{3}\tilde{z}^{\frac{3}{2}}\right)^{-2m} - \cos\left(\frac{2}{3}\tilde{z}^{\frac{3}{2}}+\frac{\pi}{4}\right)\sum_{m=0}^{\infty}(-1)^m c_{2m+1}\left(\frac{2}{3}\tilde{z}^{\frac{3}{2}}\right)^{-2m-1}\right] \qquad \text{for} \quad |\arg\tilde{z}| < \frac{2}{3}\pi \ , \quad (7.23)$$

see [65]. Thus for $z^2 \geq 1$ which corresponds to $\frac{a^4}{n^2} \geq 1$, the argument of the Airy function is negative, $\zeta \leq 0$, and (7.23) applies. The phase associated with the gravitational part is therefore

$$\Theta_n = \nu\left(\sqrt{z^2-1} - \arccos\left(\frac{1}{z}\right)\right) + \frac{\pi}{4} \ .$$

For $z^2 < 1$, the argument of the Airy function is positive, $\zeta > 0$, and the first of the above equations, (7.22), applies. But this yields no phase. The Airy function decays exponentially. This is perfectly consistent as the region $z^2 < 1$ corresponds to $\frac{a^4}{n^2} < 1$ — which is just the classically forbidden region.

In the classically *allowed* region, however, the total phase is given by

$$S_n \equiv \Theta_n \pm \frac{n}{\hbar}\phi \ . \quad (7.24)$$

The extremum of this phase should yield the classical trajectory. The requirement $\partial S_n/\partial n = 0$ at $n = \bar{n}$ leads to

$$\phi(a) = \pm\frac{1}{2}\arccos\frac{\kappa\bar{n}}{\sqrt{6}a^2} \ .$$

Comparing this to the classical trajectory, (7.1), it is clear that $\bar{n}$ has to be identified with P_ϕ.

One can also easily check that S_n is a solution of the Hamilton–Jacobi equation arising from (6.5) through the substitutions $p_a \to \partial S_n/\partial a$ and $p_\phi \to \partial S_n/\partial\phi$.

In the case of the conventional scalar field, one gets a change of sign for the n^2-term in (7.18). The solutions for $C_n(a)$ are then the Bessel functions $J_{i\nu}(\frac{3a^2}{\hbar\kappa})$ and $J_{-i\nu}(\frac{3a^2}{\hbar\kappa})$. Since there are no classically forbidden regions, both solutions seem to be allowed. The classical solution (7.2) follows in the formal limit '$\hbar \to 0$' from the principle of constructive interference.

Since (6.11) is hyperbolic for $\ell = 1$, one is free to impose boundary conditions at constant a, that is, one can either impose one packet or two packets there, depending on whether one wants one branch of the classical solution to be represented or both.

In the phantom case discussed above, the Wheeler–DeWitt equation is elliptic. One imposes there only the boundary condition that ψ goes to zero at $a \to 0$ and that it is at most oscillating at the other boundaries. This fixes the solution to be $J_\nu(\frac{3a^2}{\hbar\kappa})$ or a superposition thereof. Explicitly, one would consider the following superposition for the construction of a wave packet,

$$\psi(a,\phi) = \int_0^\infty \mathrm{d}n\ A(n) e^{-\mathrm{i}\frac{n}{\hbar}\phi} J_{\frac{3n}{\hbar\kappa}}\left(\frac{3a^2}{\hbar\kappa}\right)\ , \tag{7.25}$$

where $A(n)$ is a function of n that is peaked around a particular value $\bar{n}$, e.g. a Gaussian. One would not expect the packet to exhibit dispersion near the minimum of the classical trajectory, since the phase of the Bessel function does not vary rapidly with respect to n. We shall, however, expect the occurrence of dispersion at *large* values of a.

7.4.2 Quantum cosmological model with phantom dark energy

For non-zero, exponential potential as in Section 7.3.2, the Wheeler–DeWitt equation can actually be solved *exactly*. This is obtained through a transformation to new variables. We start from (6.12) with $\Lambda = 0$ and $k = 0$,

$$\frac{\hbar^2}{2}\left[\frac{\kappa^2}{6}\frac{\partial^2}{\partial\alpha^2} - \ell\frac{\partial^2}{\partial\phi^2}\right]\psi(\alpha,\phi) + a_{\mathrm{ref}}^6 V_0 e^{6\alpha-\gamma\kappa\phi}\psi(\alpha,\phi) = 0\ .$$

Exact solution to the Wheeler–DeWitt equation The new variables are chosen such that the potential in front of ψ cancels. This is done in two steps. First, we transform to light-cone like coordinates

$$z_1 = \frac{\sqrt{6}}{\kappa}\alpha + \sqrt{\ell}\phi \ , \qquad z_2 = \frac{\sqrt{6}}{\kappa}\alpha - \sqrt{\ell}\phi \ .$$

For $\ell = 1$, these are just the characteristics of the Wheeler–DeWitt equation. The equation now takes the form

$$\hbar^2 \frac{\partial^2 \psi}{\partial z_1 \partial z_2} + f\left(z_1, z_2\right)\psi = 0 \ ,$$

from which a transformation to new variables can be made such that $f\left(z_1, z_2\right)$ is cancelled. This corresponds to a transformation to variables

$$\begin{aligned} u_\ell(\alpha,\phi) &= \frac{\kappa}{3}\sqrt{\frac{V_0}{3}}a_{\rm ref}^3 \frac{e^{3\alpha - \frac{\gamma}{2}\kappa\phi}}{1-\ell\left(\frac{\gamma}{\sqrt{6}}\right)^2}\left(\cosh(X) + \frac{1}{\sqrt{\ell}}\frac{\gamma}{\sqrt{6}}\sinh(X)\right) \ , \\ v_\ell(\alpha,\phi) &= \frac{\kappa}{3}\sqrt{\frac{V_0}{3}}a_{\rm ref}^3 \frac{e^{3\alpha - \frac{\gamma}{2}\kappa\phi}}{1-\ell\left(\frac{\gamma}{\sqrt{6}}\right)^2}\left(\frac{1}{\sqrt{\ell}}\sinh(X) + \ell\frac{\gamma}{\sqrt{6}}\cosh(X)\right) \ , \end{aligned}$$

where $X = \sqrt{\ell}(\sqrt{\frac{3}{2}}\kappa\phi - \ell\frac{\gamma\sqrt{6}}{2}\alpha)$. For both, the phantom and the ordinary field, u_ℓ and v_ℓ are real. Note that whereas v_1, u_{-1} and v_{-1} cover the entire real line, $u_1 > 0$ is restricted to the positive real axis due to the requirement $\gamma < \sqrt{6}$.

The Wheeler–DeWitt equation in these variables takes the simple form

$$\hbar^2\left(\frac{\partial^2\psi}{\partial u_\ell^2} - \ell\frac{\partial^2\psi}{\partial v_\ell^2}\right) + \psi = 0 \ .$$

Making a WKB-ansatz, $\psi(u_\ell, v_\ell) = Ce^{\pm\frac{\rm i}{\hbar}S}$, one obtains at lowest order the Hamilton–Jacobi equation

$$\left(\frac{\partial S_0}{\partial u_\ell}\right)^2 - \ell\left(\frac{\partial S_0}{\partial v_\ell}\right)^2 = 1 \ . \tag{7.26}$$

This is solved via a separation ansatz by $S_{0n} = nu_\ell - \sqrt{\ell(n^2-1)}v_\ell$. Of course, the Hamilton–Jacobi equation is also solved by actions carrying different signs in front of u_ℓ and v_ℓ. These are obtained from the one chosen above by rotations in the (u_ℓ, v_ℓ)-plane. For $\ell = -1$, all solutions can be mapped onto each other in this way. This is an obvious consequence of the rotational symmetry of equation (7.26) for $\ell = -1$. As $u_1 > 0$ for the conventional scalar field, here only *two* solutions can be mapped onto each other. So that for $\ell = 1$, we have two solutions given by

$$S_{0n}^{(1)} = nu_1 - \sqrt{(n^2-1)}v_1 \ , \qquad S_{0n}^{(2)} = -nu_1 - \sqrt{(n^2-1)}v_1 \ .$$

Plugging this lowest-order ansatz into the Wheeler–DeWitt equation, one finds that the equation is already satisfied *exactly*.

Recovery of classical trajectories From the classical action $S^{(1)}_{0n}$, the equations of motion are obtained via $\frac{\partial S^{(1)}_{0n}}{\partial n}|_{n=\bar{n}} = 0$. This yields

$$\phi(\alpha) = \ell\frac{\gamma}{\kappa}\alpha + \sqrt{\frac{2}{3}}\frac{1}{\kappa}\text{arctanh}\left(c^{(1)}_{\bar{n},\ell}\right) \ ,$$

where

$$c^{(1)}_{\bar{n},\ell} = \frac{\ell\gamma\bar{n} - \sqrt{6\ell(\bar{n}^2-1)}}{\gamma\sqrt{\ell(\bar{n}^2-1)} - \bar{n}\sqrt{6}} \ .$$

The same holds, of course, also for $S^{(2)}_{0n}$ where $c^{(1)}_{\bar{n},\ell}$ is replaced by

$$c^{(2)}_{\bar{n},\ell} = \frac{\gamma\bar{n} + \sqrt{6(\bar{n}^2-1)}}{\gamma\sqrt{(\bar{n}^2-1)} + \bar{n}\sqrt{6}} \ .$$

For the choice

$$\bar{n}^2 = 1/E_{\text{pot}} = \left(1 - \frac{\ell\gamma^2}{6}\right)^{-1} \ ,$$

$c^{(1)}_{\bar{n},\ell} = 0$, $c^{(2)}_{\bar{n},\ell} = 0$ and one obtains the classical trajectory

$$\phi(\alpha) = \ell\frac{\gamma}{\kappa}\alpha \ ,$$

cf. (7.6).

Construction of wave packets For the phantom model, we get the following *exact* wave-packet solution to the Wheeler–DeWitt equation,

$$\psi(u_{-1}, v_{-1}) = \int dn\, A(n)\left(C_1 e^{\frac{\mathrm{i}}{\hbar}S^{(1)}_{0n}} + C_2 e^{-\frac{\mathrm{i}}{\hbar}S^{(1)}_{0n}}\right) \ .$$

For the scalar field, on the other hand, four terms contribute,

$$\psi(u_{-1}, v_{-1}) = \int dn\, A(n)\left(C_1 e^{\frac{\mathrm{i}}{\hbar}S^{(1)}_{0n}} + C_2 e^{-\frac{\mathrm{i}}{\hbar}S^{(1)}_{0n}} + C_3 e^{\frac{\mathrm{i}}{\hbar}S^{(2)}_{0n}} + C_4 e^{-\frac{\mathrm{i}}{\hbar}S^{(2)}_{0n}}\right) \ . \tag{7.27}$$

By construction, the classical trajectories can be recovered from these equations through the principle of constructive interference. We choose for the amplitude a Gaussian with width σ centred around $\bar{n}$,

$$A(n) = \frac{e^{-\frac{(n-\bar{n})^2}{2\sigma^2\hbar^2}}}{(\sqrt{\pi}\sigma\hbar)^{1/2}} \ .$$

Let us start with the calculation for the wave packet for the phantom model. Taking $C_1 = C_2$ for definiteness, one obtains wave packets of the form

$$\psi(u_{-1}, v_{-1}) \approx C_1 \pi^{1/4} \sqrt{\frac{2\sigma\hbar}{1 - \mathrm{i}\sigma^2\hbar S_0''}} \exp\left(\frac{iS_0}{\hbar} - \frac{S_0'^2}{2(\sigma^{-2} - \mathrm{i}\hbar S_0'')}\right) + \text{c.c} \ , \tag{7.28}$$

where a Taylor expansion of $S_{0n}^{(1)}$ has been carried out around $\bar{n}$ and primes denote derivatives with respect to n. Terms of order $(n-\bar{n})^3$ in the exponent have been neglected. This can be done if the Gaussian is strongly peaked around $\bar{n}$, that is, if σ is sufficiently small. For simplicity, in this expression $S_0 = S_{0n}^{(1)}(\bar{n})$. Since $S_{0n}'^{(1)}(\bar{n}) = 0$ defines the classical trajectory, the packet is peaked around it.

For the conventional scalar field, choose $C_1 = C_2$ and $C_3 = C_4$. However, we leave the relation between C_1 and C_3 open for the time being. Then one obtains *two* packets of the form (7.28), one of them with $S_0 = S_{0n}^{(1)}(\bar{n})$ and the second one with $S_0 = S_{0n}^{(2)}(\bar{n})$.

The wave packet for the phantom-field model is depicted in Appendix F, Figure F.4. It can be clearly seen that the wave packet follows the classical trajectory

$$u_{-1} = -\frac{\bar{n}}{\sqrt{1 - \bar{n}^2}} v_{-1} \ . \tag{7.29}$$

The wave packet for the scalar field model is depicted in Appendix F, Figure F.5. In this plot, the entire (u_1, v_1)-lane is shown — but the (α, ϕ)-plane just corresponds to one quadrant of it. All quadrants are equivalent to each other. The wave packet is seen to follow the classical trajectory.

For both, the conventional scalar field and the phantom field, the wave packet *spreads* as $v_\ell^2 \to \infty$. This can be seen from (7.28), since the term proportional to $[S_{0n}''(\bar{n})]^2$ in the width of the Gaussian increases without limit,

$$S_{0n}''(\bar{n}) = \frac{v_\ell}{(\ell(\bar{n}^2 - 1))^{\frac{3}{2}}} \ .$$

This is even more obvious from the absolute square of the wave packet (neglecting for simplicity the complex conjugate part in (7.28)),

$$|\psi(u_\ell, v_\ell)|^2 \approx |C_1|^2 \sqrt{\pi} \frac{2\sigma\hbar}{\sqrt{1 + \sigma^4\hbar^2 (S_0'')^2}} \exp\left(-\frac{S_0'^2}{\sigma^{-2} + \sigma^2\hbar^2 (S_0'')^2}\right) \ .$$

The spreading occurs due to the non-trivial dispersion relation, that is, due to the fact that S_{0n} depends non-linearly on n. The semiclassical approximation is thus not valid throughout configuration space.

Behaviour at the classical singularities For the phantom field we have $u_{-1} \to -\infty$, $v_{-1} \to \infty$ when we approach the big-rip singularity. As the wave packet spreads upon approach of that region, *this singularity lies in a genuine quantum region*. Note that for $\ell = -1$, u_ℓ, v_ℓ are defined in terms of sine and cosine. Since these are between -1 and 1 for all α and ϕ, we can write

$$v_\ell^2 \sim e^{6\alpha - \gamma\kappa\phi} \equiv e^{6\alpha} V(\phi) \ .$$

Therefore, the occurrence of the non-trivial potential is responsible for the dispersion.

The big-rip singularity is thus *smoothed out* — the wave packets disperse. We loose connection to classical spacetime. Classical spacetime dissolves. Correspondingly, we can no longer use the approximate time parameter given by WKB-time. Time and classical evolution come to an end, and one is just left with a stationary quantum state. This is a *quantum gravity effect at large scales*. Hitherto such a case has only been encountered near the turning point of a classically recollapsing universe, as a consequence of the demand that the wave function go to zero for large scale factor, [63].

Turning now to the conventional scalar field model, we recall that due to the fact that $u_1 > 0$, here two inequivalent actions exist. Apart from the wave packet constructed from the function $S_{0n}^{(1)} = nu_1 - \sqrt{n^2-1}v_1$, one gets a second wave packet constructed from $S_{0n}^{(2)} = -nu_1 - \sqrt{n^2-1}v_1$. Moreover, the *entire* (α, ϕ)-plane is mapped into only one *quarter* of the (u_1, v_1)-plane. One would therefore require the wave packet to vanish on the boundary of the physical region, i.e. on $(u_1, 0)$ and $(0, v_1)$. The only solution satisfying this requirement is naturally the trivial one. To get a *non-trivial* solution, one has to lessen the boundary condition and require $\psi = 0$ either *only on* $u_1 = 0$ or *only on* $v_1 = 0$. Note that $u_1 = 0$ corresponds to $\alpha \to -\infty$ and v_1 vanishes for $\alpha \to -\infty$ and for $\phi(\alpha) = \frac{\gamma}{\kappa}\alpha + \text{arctanh}\left(-\frac{\gamma}{\sqrt{6}}\right)$. But as the latter corresponds to the classical trajectory for specific initial conditions, it is not consistent to require $\psi = 0$ for $v_1 = 0$ — this would make it impossible to find wave packets that follow this specific classical path. Therefore we just require vanishing on the $u_1 = 0$ line. But this means that we require $\psi \to 0$ as $\alpha \to -\infty$. *This is in fact just DeWitt's boundary proposal* (recall that $\alpha \to -\infty$ as $a \to 0$). Note that v_1 vanishes also for $\alpha \to -\infty$. So, in fact, we require that ψ vanish at the origin of the (u_1, v_1)-plane. This necessitates the choice $C_1 = -C_3$ in (7.27).

The implementation of this condition results in a wave packet which vanishes at the big-bang singularity, $\psi \to 0$ as $\alpha \to -\infty$, and spreads for large α. The big-bang singularity does therefore not exist in the quantum theory. In Appendix F, Figure F.5 the wave packet is shown for different

choices of C_1 and C_3. From Appendix F, Figure F.6 can be seen that the choice of C_1 and C_3 satisfies the condition of vanishing ψ at the origin of the (u_1, v_1)-plane.

The fact that the wave packet does not vanish at both, the $u_1 = 0$ and $v_1 = 0$ line, is due to the non-normalizability of the wave packet in both α and ϕ, which in turn has its origin in the fact that the classical trajectory has no turning point.

In the phantom field model, no such restriction occurs due to the fact that the entire (u_{-1}, v_{-1})-plane represents the entire (α, ϕ)-plane. There is no big-bang singularity in the phantom model and consequently, no condition $\psi = 0$ as $a \to 0$ *arises*.[5] In this model, at least, it seems that DeWitt's boundary proposal arises *naturally* — through the choice of coordinates. It only arises in the scalar-field model which *on the classical level* exhibits a big bang. For the phantom-field model, on the other hand, the variables are chosen in the same way as for the scalar-field case. However, here DeWitt's boundary proposal does *not* arise as a natural condition. This is in accord with the fact that the phantom model does not start with a big-bang singularity on the classical level.

7.4.3 Quantum cosmological model with phantom dark energy and cosmological constant

To obtain a classical trajectory with turning point in configuration space, we introduced a negative cosmological constant and a potential of the form $V(\phi) = V_0 \cosh^2(\phi/F)$, cf. (7.14), in Section 7.3.3. The Wheeler–DeWitt equation for this model is given by

$$\frac{\hbar^2}{2}\left(\frac{\kappa^2}{6}\frac{\partial^2}{\partial\alpha^2} - \ell\frac{\partial^2}{\partial\phi^2}\right)\psi(\alpha,\phi) + a_{\text{ref}}^6 e^{6\alpha}\left(V_0\cosh^2\left(\frac{\phi}{F}\right) + \frac{\Lambda}{\kappa^2}\right)\psi(\alpha,\phi) = 0\ . \tag{7.30}$$

To find a solution to this equation, I divided the ϕ-axis into two regions. One region is the regime of large $|\phi|$, i.e. $V(\phi) > \frac{|\Lambda|}{\kappa^2}$. This is the region were the phantom field dominates over Λ. In this region lie the classical big-rip singularities. Moreover, I approximated the matter potential in this region for large $|\phi|$.

The second region is the region of small $|\phi|$, $V(\phi) < \frac{|\Lambda|}{\kappa^2}$. Here, the attractive force of Λ dominates over the phantom repulsion. This is the region of the classical turning point. Furthermore, I approximated the matter potential for small $|\phi|$.

Due to this splitting, however, the impact of the classical turning point on the wave function at the big-rip singularity can only be felt upon matching

[5] I want to emphasize again that we do not put in the condition of vanishing ψ by hand, but that it is enforced upon us through the choice of coordinates in which the Wheeler–DeWitt equation becomes exactly solvable.

the solutions of the respective regions. Such a matching has not been carried out.

Vicinity of the big-rip singularity

We thus assume that the Λ-term in (7.30) is negligible in comparison to the phantom potential. The classical singularities lie in a region of large $|\phi|$. Thus in order to study the quantum behaviour there, we furthermore approximate the potential for large $|\phi|$,

$$\tilde{V}(\phi) \approx \frac{a_{\rm ref}^6 V_0}{4} e^{\pm\frac{2\phi}{F}} \ ,$$

where in the following the upper sign refers to positive ϕ, and the lower sign to negative ϕ. This makes the problem very similar to the one of Section 7.4.2. The Wheeler–DeWitt equation is here simplified by a transformation on the variables

$$\begin{aligned} u_\ell(\alpha,\phi) &= \sqrt{\frac{V_0}{3}}\frac{\kappa}{\sqrt{6}} a_{\rm ref}^3 \frac{e^{3\alpha\pm\frac{1}{F}\phi}}{1-\ell\left(\frac{2}{F\kappa}\right)^2}\left(\cosh(X) \mp \frac{1}{\sqrt{\ell}}\sqrt{\frac{2}{3}}\frac{1}{F\kappa}\sinh(X)\right) \ , \\ v_\ell(\alpha,\phi) &= \sqrt{\frac{V_0}{3}}\frac{\kappa}{\sqrt{6}} a_{\rm ref}^3 \frac{e^{3\alpha\pm\frac{1}{F}\phi}}{1-\ell\left(\frac{2}{F\kappa}\right)^2}\left(\frac{1}{\sqrt{\ell}}\sinh(X) \mp \ell\sqrt{\frac{2}{3}}\frac{1}{F\kappa}\cosh(X)\right) \ , \end{aligned}$$

where $X = \sqrt{\ell}\left(\sqrt{\frac{3}{2}}\kappa\phi \pm \ell\frac{\sqrt{6}}{F\kappa}\alpha\right)$. In these variables, we recover the form

$$\hbar^2\left(\frac{\partial^2\psi}{\partial u_\ell^2} - \ell\frac{\partial^2\psi}{\partial v_\ell^2}\right) + \psi = 0 \ .$$

Again, one obtains a solution from a WKB-ansatz. The Hamilton–Jacobi equation is again given by (7.26). This equivalence is, of course, only formal, since u_ℓ and v_ℓ are defined differently. It is again solved by $S_{0n}^{(1)} = nu_\ell - \sqrt{\ell\,(n^2-1)}v_\ell$ and additionally by $S_{0n}^{(2)} = -nu_\ell - \sqrt{\ell\,(n^2-1)}v_\ell$ for $\ell = 1$. The remarks of Section 7.4.2 concerning choice of solution apply here as well.

The equations of motion obtained for $\frac{\partial S_{0n}}{\partial n}|_{n=\bar{n}} = 0$ are

$$\phi(\alpha) = \mp\ell\frac{2}{\kappa^2}\frac{1}{F}\alpha + \operatorname{arctanh}\left(c_{\bar{n},\ell}\right) \ ,$$

with some constant $c_{\bar{n},\ell}$ depending on $\bar{n}$ and ℓ. Introducing D again, (7.14), we obtain

$$\phi(\alpha) = \mp\frac{2\sqrt{3}}{\kappa^2}\sqrt{\ell D}\alpha + \operatorname{arctanh}\left(c_{\bar{n},\ell}\right) \ .$$

This solution coincides approximately with the classical solutions (7.13) which can be seen from the following calculations.

If one approximates (7.13) for $\ell = -1$ for *large* a, one gets ($\pm$ label the different branches of the classical solution)

$$\phi_{\pm}(\alpha) = \pm\frac{2\sqrt{3}}{\kappa^2}\sqrt{\ell D}\alpha \pm F\ln(2A) \ ,$$

where $\alpha \geq 0$. Therefore, the limit of large positive ϕ is obtained on the ϕ_+–branch, while the limit for large negative ϕ is reached on the ϕ_-–branch.

On the other hand, an approximation of (7.13) for *small* a in the case $\ell = 1$ yields

$$\phi_{\pm}(\alpha) = \pm\frac{2\sqrt{3}}{\kappa^2}\sqrt{\ell D}\alpha \mp F\ln(2A) \ ,$$

where $\alpha \leq 0$. Due to this, the limit of large positive ϕ is obtained on the ϕ_-–branch, and for large negative ϕ on the ϕ_+–branch. Thus the solution to the approximated Hamilton–Jacobi equation (7.26) coincides with the approximation of equation (7.13). Of course, a special choice for $\bar{n}$ has to be made to fix the onset.[6]

With the help of the classical action S_{0n}, the approximate Wheeler–DeWitt equation can be solved. Again, the WKB-ansatz satisfies the equation exactly. The wave packet is of the same form as in Section 7.4.2, with a different definition of u_ℓ and v_ℓ and another choice of the centre of the Gaussian, $\bar{n}$. As in the case of vanishing cosmological constant, the wave packet spreads for $v_\ell^2 \to \infty$. The big-rip singularity in these variables occurs at $v_{-1}^2 \to \infty$, $u_{-1} \to \infty$. Thus, again, the singularity is hidden in a quantum regime and the semiclassical approximation is not valid throughout configuration space.

Due to the restriction $D < 3$ for the $\ell = 1$ model, the same remarks concerning the range of the coordinates as in Section 7.4.2 apply here. So

[6]The fact that for $\ell = -1$ large ϕ correspond to *large* a, and for $\ell = 1$ large ϕ correspond to *small* a is due to 'phantom-scalar field duality'. This denotes the following property of the Wheeler–DeWitt equation for $k = 0$. The Wheeler–DeWitt equation of a phantom field can be transformed in the Wheeler–DeWitt equation for an ordinary scalar field through

$$a \to \frac{1}{\bar{a}} \ , \qquad \phi \to -\mathrm{i}\bar{\phi} \ .$$

Under this transformation, the Wheeler–DeWitt equation with phantom field,

$$\left(\frac{\hbar^2}{2}\frac{\kappa^2}{6}a\frac{\partial}{\partial a}a\frac{\partial}{\partial a} - \ell\frac{\hbar^2}{2}\frac{\partial^2}{\partial\phi^2} + a^6\left(V(\phi) + \frac{\Lambda}{\kappa^2}\right)\right)\psi(a,\phi) = 0 \ ,$$

takes the form of a Wheeler–DeWitt equation with ordinary scalar field but altered and *complex* potential

$$\left(\frac{\hbar^2}{2}\frac{\kappa^2}{6}\bar{a}\frac{\partial}{\partial\bar{a}}\bar{a}\frac{\partial}{\partial\bar{a}} + \ell\frac{\hbar^2}{2}\frac{\partial^2}{\partial\bar{\phi}^2} + \frac{1}{\bar{a}^6}\left(V(i\bar{\phi}) + \frac{\Lambda}{\kappa^2}\right)\right)\psi(\bar{a},\bar{\phi}) = 0 \ .$$

The transformation for ϕ is thus just a Wick rotation.

at the big bang, $\psi \to 0$. This condition is enforced through the choice of coordinates. Again, this arises only at the big bang and only in the model that on the classical level actually exhibits a big-bang singularity.

Vicinity of the classical turning point

We assume that $V(\phi) \ll \frac{|\Lambda|}{\kappa^2}$. But this implies that $E_{\text{pot}} = \frac{\kappa^2}{3}\left(V(\phi) - \frac{|\Lambda|}{\kappa^2}\right) < 0$. For the phantom case, this is the classically forbidden region. This will be of relevance in the choice of solutions from the general one.

In the vicinity of the classical turning point, $\phi \to 0$, i.e. $V(\phi) \to 4V_0$. The Wheeler–DeWitt equation in this realm is the given by

$$\frac{\hbar^2}{2}\left(\frac{\kappa^2}{6}\frac{\partial^2}{\partial\alpha^2} - \ell\frac{\partial^2}{\partial\phi^2}\right)\psi(\alpha,\phi) + a_{\text{ref}}^6 e^{6\alpha}\left(4V_0 + \frac{\Lambda}{\kappa^2}\right)\psi(\alpha,\phi) = 0 \ .$$

Using the short-hand $B = a_{\text{ref}}|V_0 + \frac{\Lambda}{\kappa^2}|$, the equation becomes

$$\frac{\hbar^2}{2}\left(\frac{\kappa^2}{6}\frac{\partial^2}{\partial\alpha^2} - \ell\frac{\partial^2}{\partial\phi^2}\right)\psi(\alpha,\phi) - Be^{6\alpha}\psi(\alpha,\phi) = 0 \ ,$$

and thus very similar as (7.15). It is consequently solved by the same ansatz

$$\psi_n(\alpha,\phi) = C_n(\alpha)\varphi_n(\phi) \ .$$

For convenience, we define the separation constant somewhat differently from (7.17). Then we arrive at

$$\varphi_n(\phi) = e^{-\frac{n}{\sqrt{\ell}\hbar}\phi} \ .$$

We choose n such that we get an *oscillating* solution for the ordinary scalar field, $\varphi_n(\phi) = e^{-\mathrm{i}\frac{n}{\hbar}\phi}$. For the phantom field, however, we choose the one that depends *exponentially* on ϕ, namely $\varphi_n(\phi) = e^{\pm\frac{n}{\hbar}\phi}$, where $\pm$ are chosen such that φ_n increases exponentially with increasing $|\phi|$. In order to obtain this, we have to choose a purely imaginary n in both cases. We replace n by $\mathrm{i}n$. The gravitational part then has to satisfy

$$\ddot{C}_n - \frac{12B}{\hbar^2\kappa^2}\left(e^{6\alpha} - \tilde{n}^2\right)C_n = 0 \ ,$$

where $\tilde{n}^2 = \frac{n^2}{2B}$. The general solution to this equation is then, as in Section 7.4.1, given in terms of Bessel functions

$$\psi(\alpha,\phi) = c_1 \mathrm{J}_{\mathrm{i}\nu}\left(\mathrm{i}\frac{2}{\kappa\hbar}\sqrt{\frac{B}{3}}e^{3\alpha}\right) + c_2\mathrm{Y}_{\mathrm{i}\nu}\left(\mathrm{i}\frac{2}{\kappa\hbar}\sqrt{\frac{B}{3}}e^{3\alpha}\right) \ ,$$

where $\nu = \sqrt{\frac{2}{3}}\frac{n}{\hbar\kappa}$. For the phantom field, the classical turning point lies at a minimum of the scale factor, whereas it occurs at maximum scale factor

for the ordinary scalar field. This knowledge should be used to exclude one of the two Bessel functions. To this end, we need the asymptotics of Bessel functions for large and small imaginary argument and imaginary index.

Recovery of the classical trajectories is done in principle in full analogy to Section 7.4.1. But here, we do not expect to recover the *exact* trajectory: we just used an approximation of the correct Wheeler–DeWitt equation. Of course, for the phantom field, no classical trajectory exists in this realm.

To obtain a full picture, we would have to match the two solutions, the one obtained in the vicinity of the respective singularities and the one that holds in the vicinity of the classical turning point, along the $V(\phi) = \frac{|\Lambda|}{\kappa^2}$–line. Such a matching has not been carried out due to the complicated dependence of u_ℓ and v_ℓ on α and ϕ. In my opinion it is also questionable whether such a matching *could* be carried out. More than just solving the Wheeler–DeWitt equation in two regions, I further approximated the potential in each region.

7.5 Conclusions

In the preceding chapter, I showed that standard quantum cosmology *can* be applied to scenarios with phantom matter. The main motivation for the quantization of these phantom cosmologies were the novel features with regard to the structure of the Wheeler–DeWitt equation (elliptic or ultra-hyperbolic instead of hyperbolic) as well as the presence of a new type of singularity. This big-rip singularity occurs at macroscopic scale factor and thus provides, in fact, the major question behind this project: do quantum effects occur in the macroscopic universe?

For three models, I discussed classical trajectories in configuration space. The first model was just a toy model without true phantom energy. In the second model, however, phantom dark energy was mimicked by a phantom field with exponential potential. The last model was chosen such that the classical trajectory has a turning point in configuration space. Moreover, it contained singularities at both ends of its evolution.

For the first two models, an exact solution to the Wheeler–DeWitt equation was given. In the third case, it was necessary to approximate the field potential in the vicinity of the singularity. Carrying out approximations in a differential equation is a difficult task, because one never knows whether the neglected terms really remain small after a solution was found. I showed, however, that in all three cases the classical trajectory can be recovered from the quantum solution through the principle of constructive interference. For the last model, this was only possible approximately.

Using these solutions, I constructed wave packets following the classical trajectory in configuration space. In a last step, I discussed the influence of quantum effects at the singularity.

In the approach to the big-rip singularity, wave packets disperse. This signals the break-down of the semi-classical approximation and thus of space-time. In this sense, the big-rip singularity is resolved on the quantum level. It would however be interesting to study how an inside observer would perceive the approach to this singularity. For this it would be necessary to find out at which time-scale quantum effects set in.

For each phantom model, I also discussed the corresponding one containing an ordinary scalar field. Instead of a big-rip singularity, these exhibit the well-known big bang at $a = 0$. In accord with DeWitt's proposal, the structure of the equation imposes conditions which then imply that the wave function vanish at the big bang. In this way, this singularity is also removed from the theory. Without the boundary condition the wave packet would just have approached the region $\alpha \to -\infty$ without spreading; this lack of dispersion is a result of the fact that the Wheeler–DeWitt equation takes the form of a free wave equation in this limit.

Note that there are two different mechanisms resolving the big-rip and the big-bang singularity. Whereas the first is hidden in a genuine quantum regime, the latter is removed through a vanishing wave function.

References: This work was published in [66]. The model presented in Section 7.3.3 was proposed by Mariusz P. Dabrowski.

Chapter 8

Quantum Cosmology with big-brake singularity

A second singularity that occurs at large scale factor is the so-called big-brake singularity. In contrast to the big rip, it arises at finite value of the scale factor. The big-brake singularity is a generic feature of cosmological models with anti-Chaplygin gas. In the present chapter, we will carry out a classical and a quantum cosmological analysis of such a big-brake singularity.

The project presented in this chapter was a direct follow-up of the project presented in the previous chapter. The motivation behind it is similar: the study of models with singularities in the macroscopic universe in the framework of quantum cosmology. After having studied *one* such singularity, why should one be interested in a *second* one?

The major reason for doing so is that not all singularities are alike. The singularity that is the subject of *this* chapter, is, for reasons that will become clear later, called *big-brake singularity*. It exhibits features very differently from the big rip. First of all, it can be induced by an ordinary scalar field. Furthermore, recall that at the big-rip singularity, pressure, energy density and scale factor diverge after finite time. The big-brake singularity now occurs at *finite* scale factor. The singularity in this model thus no longer lies on the boundary of configuration space, as do the big rip ($a \to \infty$) and the big bang ($a = 0$). No boundary conditions can be imposed in the region of the big brake. Moreover, as it is a singularity that lies somewhere in configuration space, one can study evolution *through* the singularity. The situation is then analogous to the big-bang singularity in loop quantum cosmology. There, due to the existence of negative scale factors, the big bang lies in the middle of configuration space.

The big brake belongs to a wider class of cosmological singularities called soft, quiescent, or sudden. These singularities occur at finite value of the

scale factor and its time derivative and hence of the Hubble parameter. But the first or higher derivatives of the Hubble parameter are divergent. This implies that some curvature invariants diverge. At the big brake, for example, the first derivative of the Hubble parameter diverges, $\dot{H} \to -\infty$, while $H \to 0$. Due to the fact that ${}^{(3)}R \sim \dot{H} + 2H^2$, we see that the curvature of three-space has to diverge at this singularity. More precisely, what happens is the following. The acceleration becomes more and more negative, $\ddot{a} \to -\infty$, thus causing the universe to slow down and finally come to a halt, $\dot{a} \to 0$. The universe then ends its evolution at a finite radius $a = a_\star$. It is this behaviour that gave the singularity its name.

The big-brake was first considered in [67] and later discussed in detail in [68]. It can arise in tachyonic cosmological models [69] with a particular potential. But the big brake can also occur in a much simpler model. A universe filled with a perfect fluid obeying the equation of state $p = A/\rho$, where A is a positive constant, ends its evolution with a big-brake singularity. This is, again, a rather odd equation of state. It was first considered in [70] in the context of wiggly strings.[1] But it can also be viewed as a generalization of the so-called Chaplygin equation of state. The corresponding Chaplygin gas obeys the equation $p = -A/\rho$. It was introduced into cosmology to unify the dark sectors of the universe, [71]. How this can be achieved can be seen immediately from the energy conservation equation, (6.10). The Chaplygin gas behaves like dust at early times and like a cosmological constant at late times.

Because of the reversed sign in the equation of state, we call the fluid with $p = +A/\rho$ used here *anti*-Chaplygin gas. Of course, it is again not the purpose of the model to fit current universe observations. The idea is merely to produce a sufficiently simple but nonetheless interesting model for quantum cosmology.

8.1 Classical cosmology with big-brake singularity

The model is set up in a flat Friedmann–Robertson–Walker universe. This is filled with an anti-Chaplygin gas. The first task is now to derive the corresponding scalar field potential that might produce such an anti-Chaplygin equation of state. To this end, we first of all calculate the evolution of the energy density with scale factor. This follows in the usual way from (6.10) and the equation of state,

$$\rho(a) = \sqrt{\frac{B}{a^6} - A}\,, \tag{8.1}$$

where $B > 0$ is some integration constant of dimension mass squared. The constant A has dimension of mass squared over length to the sixth power.

[1] These are cosmic strings with small-scale wiggles imposed on their dynamics.

We have chosen the solution with $\rho \geq 0$. Note that ρ is well defined only for $a < a_\star \equiv (B/A)^{1/6}$, cf. Figure 8.2. As $a_\star$ is approached, the density goes to zero.

Using the result (8.1), one obtains from (6.6)

$$\int_a^{a_\star} \frac{\mathrm{d}\tilde{a}}{\left(\frac{B}{\tilde{a}^2} - A\tilde{a}^4\right)^{\frac{1}{4}}} = \frac{\kappa}{\sqrt{3}}\,(t_0 - t) \ , \tag{8.2}$$

where $a(t_0) = a_\star$ is the big-brake singularity and $a(0) = 0$ the big bang. In order to calculate this integral, we substitute $z = (B/a^6 - A)^{1/4}$, with $0 \leq z \leq \infty$. Then (8.2) becomes

$$\int_0^z \mathrm{d}\tilde{z}\ \frac{\tilde{z}^2}{\tilde{z}^4 + A} = \frac{\kappa\sqrt{3}}{2}(t_0 - t) \ . \tag{8.3}$$

The integral on the left-hand side can be found in [72]. For (8.3) one then gets[2]

$$\begin{aligned} & \frac{1}{4A^{1/4}\sqrt{2}} \left(\ln \frac{z^2 - A^{1/4} z\sqrt{2} + A^{1/2}}{z^2 + A^{1/4} z\sqrt{2} + A^{1/2}} \right. \\ & \left. \quad +2 \arctan \frac{A^{1/4} z\sqrt{2}}{A^{1/2} - z^2} + \pi\theta(z^2 - A^{1/2}) \right) \\ = \quad & \frac{\kappa}{\sqrt{3}}(t_0 - t) \ . \end{aligned}$$

We have added the Heaviside θ-function in order to make the arctan-function continuous at the point $z^2 = A^{1/2}$.

For the total time that elapses from big bang to big brake one then finds

$$t_0 = \frac{2}{3}\sqrt{C} \int_0^\infty \mathrm{d}z\ \frac{z^2}{z^4 + A} = \frac{\pi}{\sqrt{6}\kappa A^{1/4}} \ . \tag{8.4}$$

A simple *approximate* solution can be found in the vicinity of $a_\star$. To this end, we write $a = a_\star - \Delta a$, which simplifies the above integral to

$$\int_0^{\Delta a} \mathrm{d}\Delta a \frac{1}{a_\star (6\Delta a)^{\frac{1}{4}}} = \frac{\kappa}{\sqrt{3}}\,(t - t_0) \ ,$$

yielding

$$\Delta a(t) = [c_1(t_0 - t)]^{\frac{4}{3}} \ .$$

So we find for the scale factor and its derivatives

[2]The derivation of this expression is due to Alexander Y. Kamenshchik.

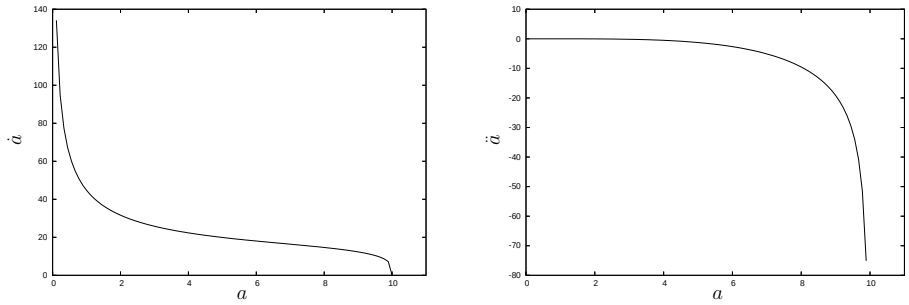

Figure 8.1: These plots show the dependence of the derivatives of the scale factor on the scale factor itself. It can be clearly seen how the velocity, see left-hand side, diminishes due to an infinite deceleration, see right-hand side.

$$a(t_0) = a_\star, \quad \dot{a}(t_0) = 0, \quad \ddot{a}(t_0) = -\infty \ .$$

At t_0, the evolution of the scale factor comes to a halt. Its 'speed' is zero due to an infinite negative acceleration. It is this peculiar feature that gave the singularity its name.

The first and second time derivatives of the scale factor in terms of the scale factor itself are given by simple expressions. To this end, note that (8.2) can be differentiated with respect to a, thus connecting $\dot{a}(t)$ with the scale factor according to

$$\frac{\mathrm{d}a}{\mathrm{d}t} = \frac{\kappa}{\sqrt{3}} a \left(\frac{B}{a^6} - A \right)^{\frac{1}{4}} \ , \tag{8.5}$$

cf. Figure 8.1. Obviously, as $a \to a_\star$, $\dot{a} \to 0$. Differentiating again with respect to time, one finds

$$\frac{\mathrm{d}^2 a}{\mathrm{d}t^2} = \frac{\kappa^2}{3} a \left(\frac{B}{a^6} - A \right)^{\frac{1}{2}} \left[1 - \frac{B}{4a^6} \left(\frac{B}{a^6} - A \right)^{-1} \right] \ , \tag{8.6}$$

showing that $\ddot{a}(t) \to -\infty$ as $a \to a_\star$, cf. Figure 8.1.

What remains to be found, is an equation for ϕ. As we are interested in the quantum model, the solution in configuration space, $\phi(a)$, suffices. This is obtained from

$$\dot{\phi}^2 = \rho + p \ ,$$

using the equation of state and the Friedmann equation (6.6). The exact solution is

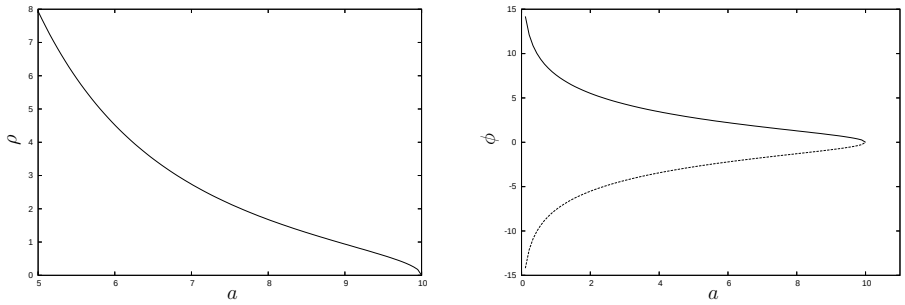

Figure 8.2: The evolution of the energy density ρ of the scalar field with scale factor a is depicted on the left-hand side. The classical trajectory in configuration space is shown on the right-hand side. This is the only information about the classical trajectory that we can access on the quantum level. Time t is absent in quantum cosmology.

$$\phi_{\mp}(a) = \mp\frac{1}{\sqrt{3\kappa^2}}\mathrm{artanh}\left(\sqrt{1-\frac{Aa^6}{B}}\right) , \tag{8.7}$$

cf. Figure 8.2. This is only consistent if the potential is chosen to be

$$V(\phi) = V_0\left(\sinh\left(\sqrt{3\kappa^2}|\phi|\right) - \frac{1}{\sinh\left(\sqrt{3\kappa^2}|\phi|\right)}\right) . \tag{8.8}$$

Given the trajectories $\phi(a)$ and $a(t)$ — the latter in explicit form only in the vicinity of the singularity — the classical model is thus fully described. Note that $V_0 = \sqrt{A/4}$. From (8.4) we find for the total lifetime of this model universe

$$t_0 \approx 7\times 10^2 \frac{1}{\sqrt{V_0\left[\frac{\mathrm{g}}{\mathrm{cm}^3}\right]}}\ \mathrm{s} .$$

This lifetime is much bigger than the current age of *our* Universe if

$$V_0 \ll 2.6\times 10^{-30}\ \frac{\mathrm{g}}{\mathrm{cm}^3} ,$$

which is a reasonable result because the critical value of V_0 just corresponds to the scale of the observed dark-energy density.

8.2 Quantum cosmology of the big-brake model

We use the regularized Wheeler–DeWitt equation (6.12) with potential (8.8), derived in the previous Chapter 6,

$$\frac{\hbar^2}{2}\left(\frac{\kappa^2}{6}\frac{\partial^2}{\partial\alpha^2}-\frac{\partial^2}{\partial\phi^2}\right)\psi(\alpha,\phi)$$
$$+a_{\mathrm{ref}}^6 V_0 e^{6\alpha}\left(\sinh\left(\sqrt{3\kappa^2}|\phi|\right)-\frac{1}{\sinh\left(\sqrt{3\kappa^2}|\phi|\right)}\right)\psi(\alpha,\phi)=0\ . \quad (8.9)$$

As we are interested in the behaviour of (8.9) in the vicinity of the big-brake singularity, where ϕ is small, we approximate the potential there. We find

$$\frac{\hbar^2}{2}\left(\frac{\kappa^2}{6}\frac{\partial^2}{\partial\alpha^2}-\frac{\partial^2}{\partial\phi^2}\right)\psi(\alpha,\phi)-\frac{\tilde{V}_0}{|\phi|}e^{6\alpha}\psi(\alpha,\phi)=0\ , \quad (8.10)$$

where $\tilde{V}_0 = V_0 a_{\mathrm{ref}}^6/\sqrt{3\kappa^2}$.

8.2.1 Born–Oppenheimer approximation to the Wheeler–DeWitt equation

In order to solve this, already approximated, equation, I used another approximation: the Born–Oppenheimer approximation. This scheme is explained in detail in Appendix D, Section D.1. First, make the ansatz $\psi(\alpha,\phi)=\sum_n C_n(\alpha)$
$\varphi_n(\alpha,\phi)$, where $\varphi_n(\alpha,\phi)$ is a solution of

$$-\left(\frac{\hbar^2}{2}\frac{\partial^2}{\partial\phi^2}+\frac{\tilde{V}_0}{|\phi|}e^{6\alpha}\right)\varphi_n(\alpha,\phi)=E_n(\alpha)\varphi_n(\alpha,\phi)\ . \quad (8.11)$$

This is the radial part of the time-independent Schrödinger equation for a particle in a Coulomb potential with $l=0$ and the wave function $r\varphi_n$, where $|\phi|$ takes the place of the radial coordinate r. Thus, the normalizable solutions are given by

$$\varphi_n(x_n)=N_n x_n e^{-\frac{x_n}{2}}\mathrm{L}_{n-1}^1(x_n)\ , \quad (8.12)$$

where $x_n = 2\sqrt{-\frac{2E_n(\alpha)}{\hbar^2}}|\phi|$, and $\mathrm{L}_{n-1}^1(x_n)$ denote the associated Laguerre polynomials. $N_n = n^{-\frac{3}{2}}$ is the normalization factor and $n\in\mathbb{N}$.

The choice of the normalizable solution to (8.11) is enforced through the condition on the wave function imposed for large $|\phi|$, cf. Section 8.3.4. From it one can infer that the *exact* solution obtained from the Wheeler–DeWitt equation with *full* potential has a discrete spectrum. Coincidence with the behaviour at small $|\phi|$ is thus only achieved if the normalizable solution (8.12) is selected, since the non-normalizable solutions have a continuous spectrum.

Note that $\varphi_n(x_n)\to 0$ for $|\phi|\to 0$, since $\mathrm{L}_{n-1}^1(0)=n$.

To simplify notation, we introduce $Z(\alpha) \equiv \hbar^2/V_\alpha$ and $V_\alpha \equiv \tilde{V}_0 e^{6\alpha}$. Then, $x_n = 2|\phi|/Z(\alpha)n$. The functions $\varphi_n(x_n)$ are orthogonal such that[3]

$$\int \mathrm{d}\phi \; \varphi_n(x_n)\varphi_l(x_l) = Z(\alpha)\delta_{nl} \; . \tag{8.13}$$

The energy eigenvalues are

$$E_n(\alpha) = -\frac{V_\alpha^2}{2\hbar^2 n^2} \; . \tag{8.14}$$

Inserting this ansatz in (8.10), it is necessary to carry out a number of approximations to arrive at a solvable equation for C_n. After a Born–Oppenheimer approximation an expansion in κ is made. The details of this approximation are given in Appendix D. The resulting equation for $C_n(\alpha)$ becomes

$$\ddot{C}_n(\alpha) - \frac{6{V_\alpha}^2}{\hbar^4 n^2 \kappa^2} C_n(\alpha) = 0 \; , \tag{8.15}$$

where dots denote derivatives with respect to α. Thus C_n is given by

$$C_n(\alpha) = c_1 \mathrm{I}_0\left(\frac{1}{\sqrt{6}}\frac{V_\alpha}{\hbar^2 n\kappa}\right) + c_2 \mathrm{K}_0\left(\frac{1}{\sqrt{6}}\frac{V_\alpha}{\hbar^2 n\kappa}\right) \; , \tag{8.16}$$

where I_0, K_0 denote modified Bessel functions of first and second kind, respectively. As a boundary condition, we require that the solution should vanish in the classically forbidden region, $a > a_\star$. Therefore, $c_1 = 0$ and only the MacDonald function K_0 remains as solution.

On the level of the Born–Oppenheimer approximation, the complete solution is therefore given by

$$\psi\,(\alpha,\phi) = \sum_{n=1}^{\infty} A_n N_n \mathrm{K}_0\left(\frac{1}{\sqrt{6}}\frac{V_\alpha}{\hbar^2 n\kappa}\right)\left(2\frac{V_\alpha}{n}|\phi|\right) e^{-\frac{V_\alpha}{n|\phi|}} \mathrm{L}^1_{n-1}\left(2\frac{V_\alpha}{n}|\phi|\right) \; . \tag{8.17}$$

8.2.2 Derivation of classical equations of motion from the principle of constructive interference

To derive a phase from this expression, we approximate (8.11) and (8.15) further by a WKB-approximation. Making the ansatz $\varphi_n(\alpha,\phi) = e^{\frac{i}{\hbar}S^\phi_{n0}(\alpha,\phi)}$ in (8.11), $C_n(\alpha) = e^{\frac{i}{\hbar}S^\alpha_{n0}(\alpha)}$ in (8.15), one obtains to zeroth order in $\hbar$ the Hamilton–Jacobi equation for the ϕ- and α-part, respectively. Integration yields for $S^\phi_{n0}(\alpha,\phi)$

[3]The validity of this relation is clear from the property of the φ_n being eigenfunctions of a Hermitian operator; its direct verification is discussed in [77].

$$S^{\phi}_{n0}(\alpha,\phi) = \hbar k\left[\arcsin\left(1-\frac{V_\alpha|\phi|}{\hbar^2 n^2}\right)-\frac{\pi}{2}\right] \\ -\sqrt{2V_\alpha|\phi|}\sqrt{1-\frac{V_\alpha|\phi|}{2\hbar^2n^2}}-\frac{\pi}{4}\ , \tag{8.18}$$

in which the Langer boundary condition at the α-dependent turning point $\phi_{\rm turn}\,(\alpha) = 2\hbar^2n^2/V_\alpha$ has been employed. From (8.15), no phase results. This coincides with the limit $\hbar \to\ 0$ in (8.16), as $\lim_{x\to\infty} \mathrm{K}_0\,(x) \approx \sqrt{\frac{\pi}{2x}}e^{-x}$. So $S^{\phi}_{n0}(\alpha,\phi)$ constitutes the entire phase.

The classical equations of motion should follow from the phase through the principle of constructive interference,

$$\begin{aligned}\frac{\partial S^{\phi}_{n0}}{\partial n}|_{n=\bar n} &= \hbar\left[\arcsin\left(1-\frac{V_\alpha|\phi|}{\hbar^2n^2}\right)-\frac{\pi}{2}\right]+\frac{\sqrt{2V_\alpha|\phi|}}{n}\sqrt{1-\frac{V_\alpha|\phi|}{2\hbar^2n^2}} \\ &\stackrel{!}{=} 0\ .\end{aligned} \tag{8.19}$$

Here, $\bar n = \sqrt{\frac{\tilde V_0}{\sqrt{3\kappa^2}}}\frac{a_\star^3}{\hbar}$. This constant arises under the conditions that, first, n and so also $\bar n$ have to be dimensionless, and that, secondly, the only constants of the model are V_0 (or $\tilde V_0$), $a_\star$ (or A and B), $\hbar$ and κ.[4] With this choice, (8.19) simplifies to

$$\frac{\partial S^{\phi}_{n0}}{\partial n}|_{n=\bar n} = \hbar\left[-\arccos\left(1-\left(\frac{a}{a_\star}\right)^6|\phi|\right)+\left(\frac{a}{a_\star}\right)^3\sqrt{2|\phi|-\left(\frac{a}{a_\star}\right)^6\phi^2}\right]\ .$$

For the classical trajectory, (8.7), this is

$$\begin{aligned}\frac{\partial S^{\phi}_{n0}}{\partial n}|_{n=\bar n} &= \hbar\left[-\arccos\left(1-\frac{|\phi|}{\cosh^2\left(\sqrt{3\kappa^2}|\phi|\right)}\right)\right. \\ &+ \left.\frac{\sqrt{|\phi|}}{\cosh\left(\sqrt{3\kappa^2}|\phi|\right)}\sqrt{2-\frac{|\phi|}{\cosh^2\left(\sqrt{3\kappa^2}|\phi|\right)}}\right]\ .\end{aligned} \tag{8.20}$$

But the classical equation of motion was derived using the *full* potential. The quantum theory, on the other hand, uses an approximation to the original potential which is valid up to order $\mathcal{O}\left(|\phi|^{\frac{3}{2}}\right)$ for small ϕ. Applying the same approximation to (8.20), one finds

[4] There is of course also the reference length $a_{\rm ref}$, but this is no characteristic quantity of the model.

$$\frac{\partial S^{\phi}_{n0}}{\partial n}|_{n=\bar{n}} = \hbar\, \mathcal{O}\Big(|\phi|^{\frac{3}{2}}\Big) \ ,$$

and so the classical solution (8.7) satisfies the condition for constructive interference with the above choice for $\bar{n}$ for small ϕ, which is consistent with the approximation of the potential in (8.10).

8.2.3 Singularity avoidance

As explained in Chapter 4, two criteria are used here to account for singularity avoidance: first, the vanishing of the wave function and second, the dispersion of wave packets along the classical trajectory in the vicinity of the singularity. I will therefore study both, the vanishing of the wave function at the origin as well as wave packets along the classical trajectory. But I will also comment on the implication of the use of expectation values in the context of singularity avoidance.

8.2.4 Behaviour at the classical singularity

Wave packets constructed from the solutions of (8.10) are of the general form

$$\psi(\alpha,\phi) = \sum_{n=1}^{\infty} A_n C_n(\alpha)\varphi_n(\alpha,\phi) \ .$$

We can choose initial conditions on a hypersurface $\alpha = \alpha_0$. Here, it suffices to fix the values $\psi(\alpha_0,\phi)$ and $\frac{\partial\psi(\alpha,\phi)}{\partial\alpha}|_{\alpha=\alpha_0}$. As for the chosen normalizable solution (8.12), $\varphi_n(\alpha,\phi)$ vanishes at $\phi = 0$ for all n and α, the wave packet is zero there. This is, of course, independent of the initial conditions. But the classical singularity occurs at $\phi = 0$. So out of these solutions, no wave packet can be constructed which does *not* vanish at the classical singularity. This suffices to count as singularity resolution according to the criteria I admitted. Note that the big brake is thus resolved in a different way than the big rip: the latter was hidden in a genuine quantum region, but the wave function was non-zero there.

As already stated in Chapter 4, there are also other criteria which are used to prove singularity avoidance. These make use of an *internal-time* variable, which just means that one of the configuration space variables is chosen as *evolution parameter*. This is done throughout configuration space or only for a portion of it, depending on whether the respective variable changes monotonously with *classical* time t. I do not support this point of view, but I nonetheless want to show how it would work in this context, but see Chapter 9 for more details on the internal-time idea.

Taking α as internal time variable, one can calculate the probability distribution,

$$|\psi|^2(\alpha_0,\phi) = \sum_{l,n} A_n A_l C_n(\alpha_0) C_l(\alpha_0) \varphi_l(\alpha_0,\phi)\varphi_n(\alpha_0,\phi) \ ,$$

for each 'instant of time' α_0. It is obvious that $|\psi|^2(\alpha_0,0) = 0$ at $\phi = 0$. This is a consequence of the choice of (8.12). Assuming that the internal time α can be treated in the same way as classical time t, one can then calculate expectation values on surfaces of constant α. These are obtained through an integration over ϕ with the measure provided through the matter Hamiltonian, cf. (8.13). From this follows that

$$\langle\psi(\alpha,\phi)|\psi(\alpha,\phi)\rangle = \sum_n A_n^2 C_n^2(\alpha) Z(\alpha) \sim \sum_n A_n^2 \frac{\left[\mathrm{K}_0(e^{6\alpha})\right]^2}{e^{6\alpha}} \ .$$

The larger α becomes, the smaller the norm of ψ. Note that the norm is not conserved in internal time α.

Before calculating now the expectation value for $|\phi|$, recall that the avoidance of the singularity of the Coulomb potential in ordinary quantum mechanics is caused by a lowest bound on the energy due to quantization. This again leads to a minimal radius for the 'trajectory' of the electron.

Analogously to the Coulomb potential in ordinary quantum mechanics, the energy of the matter component in our model is also bounded from below. The minimal energy, given by (8.14) for $n = 1$, corresponds to a minimal 'radius', that is, to a minimal value for $|\phi|$. This is given by

$$\begin{aligned} \langle|\phi_n|\rangle(\alpha) &= [C_n(\alpha)]^2 \frac{3}{2} [Z(\alpha)]^2 n^2 \\ &= \left[K_0\left(\frac{1}{\sqrt{6}}\frac{V_\alpha}{\hbar^2 n\kappa}\right)\right]^2 \frac{3\hbar^4}{2V_\alpha^2} n^2 \ , \end{aligned} \tag{8.21}$$

for $n = 1$. The classical singularity lies at $\alpha = \alpha_\star$. In this case the minimal energy is given by

$$E_1(\alpha_\star) = -\frac{V_{\alpha_\star}^2}{2\hbar^2} \ ,$$

and the expectation value for $|\phi|$ is consequently given by $\langle|\phi_1|\rangle(\alpha_\star)$. The boundedness of the energy here prevents the scalar field to evolve to the singularity, $|\phi| = 0$, in complete analogy to the Coulomb case.

Note that for $\alpha \to \infty$, the energy is no longer bounded. In this case $\langle|\phi_1|\rangle \to 0$, cf. (8.21). Of course, one should keep in mind that the expectation value in quantum cosmology, as defined above, has no interpretation in terms of measurement results as it has in conventional quantum theory.

Construction of wave packets

Apart from the avoidance of the singularity, we want to study semi-classical and quantum regimes of the model. To this end, we construct semi-classical wave packets and study their behaviour. Especially, we are interested in the regions of configuration space where these packets spread — if they spread at all.

We want $\psi(\alpha_0, \phi)$ to be a Gaussian centred at ϕ_0 with width $\sqrt{\frac{Z_0}{2}}$, where $Z_0 \equiv Z(\alpha_0)$. The centre ϕ_0 should be the value of the classical trajectory at α_0. Note that we have two classical solutions, ϕ_+ and ϕ_-, see (8.7).[5] So in fact, we have to construct two Gaussians, one centred at ϕ_0, the other at $-\phi_0$ and superpose both. Write therefore

$$\psi(\alpha_0, \phi) = \psi_-(\alpha_0, \phi) + c_1 \psi_+(\alpha_0, \phi) \ ,$$

where ψ_+ denotes the part of the wave packet being centred around ϕ_0 and ψ_- the part centred around $-\phi_0$ at initial 'time' α_0.

The calculation of the wave packet will employ only the WKB solution of (8.15). With suitable initial conditions, it reads

$$C_n(\alpha) = \frac{e^{3\alpha_0}}{e^{3\alpha}} \exp\left[-\frac{1}{6} \frac{\tilde{V}_0}{\sqrt{2\hbar^2 n^2}} \sqrt{\frac{6}{\kappa^2}} \left(e^{6\alpha} - e^{6\alpha_0} \right) \right] . \tag{8.22}$$

Introducing $\tau \equiv e^{6\alpha}$ (and denoting $\tau_0 \equiv e^{6\alpha_0}$),

$$C_n(\tau) = \left(\frac{\tau_0}{\tau} \right)^{\frac{1}{2}} \exp\left[-\frac{1}{6} \frac{\tilde{V}_0}{\sqrt{2\hbar^2 n^2}} \sqrt{\frac{6}{\kappa^2}} \left(\tau - \tau_0 \right) \right] .$$

Start with the ψ_+–part of the wave packet. We here find the requirement

$$\psi_+(\alpha_0, \phi) = \sum_{n=1}^{\infty} A_n^+ \varphi_n(\alpha_0, \phi) \stackrel{!}{=} e^{-\frac{(\phi - \phi_0)^2}{Z_0}} \ .$$

Decomposing the Gaussian into the $\varphi_n(\alpha_0, \phi)$, one obtains for the coefficients the somewhat lengthy expression

$$\begin{aligned} A_n^+ = \quad & \tfrac{N_n}{n} \exp\left[-\tfrac{\phi_0^2}{Z_0} + \tfrac{1}{2Z_0} \left(\tfrac{1}{2n} - \phi_0 \right)^2 \right] \times \\ & \textstyle\sum_{m=0}^{n-1} (-1)^m (m+1) \tfrac{(n!)}{(n-m-1)!(m+1)!} \times \\ & \left(\sqrt{\tfrac{2}{Z_0}} \tfrac{1}{n} \right)^m \mathrm{D}_{-(m+2)} \left[\sqrt{\tfrac{2}{Z_0}} \left(\tfrac{1}{2n} - \phi_0 \right) \right] \ , \end{aligned}$$

[5] The case with two Gaussians is the most general one. One may, of course, wish to choose only one Gaussian in order to represent only one branch of the classical solutions by a wave packet.

where $D_m(x)$ denote parabolic cylinder functions. Note that this expansion in φ_n cannot be performed at $\phi = 0$. Here, $\varphi_n(\alpha, \phi = 0) = 0$ for all n as remarked above.

The amplitude of ψ_- is obtained in a similar way, or by just substituting $-\phi_0$ for ϕ_0. The solution is

$$\begin{aligned} A_n^- = \quad & \tfrac{N_n}{n} \exp\left[-\tfrac{\phi_0^2}{Z_0} + \tfrac{1}{2Z_0}\left(\tfrac{1}{2n} + \phi_0\right)^2\right] \times \\ & \textstyle\sum_{m=0}^{n-1} (-1)^m (m+1) \tfrac{(n!)}{(n-m-1)!(m+1)!} \times \\ & \left(\sqrt{\tfrac{2}{Z_0}} \tfrac{1}{n}\right)^m \mathrm{D}_{-(m+2)}\left[\sqrt{\tfrac{2}{Z_0}}\left(\tfrac{1}{2n} + \phi_0\right)\right] \ . \end{aligned}$$

So the wave packet is given by

$$\psi(\alpha, \phi) = \sum_{n=1}^{\infty} \left[A_n^+ + c_1 A_n^-\right] C_n(\alpha) \varphi_n(\alpha, \phi) \ . \tag{8.23}$$

The total probability for the wave packet is calculated via

$$\begin{aligned} \int \mathrm{d}\phi \, |\psi|^2 = \quad & \tfrac{\tau_0 \hbar^2}{\tilde{V}_0} \tfrac{1}{\tau^2} \textstyle\sum_{n=1}^{\infty} \left[A_n^+ + c_1 A_n^-\right]^2 \times \\ & \exp\left(-\tfrac{1}{3} \tfrac{\tilde{V}_0}{\sqrt{2\hbar^2 n^2}} \sqrt{\tfrac{6}{\kappa^2}} (\tau - \tau_0)\right) \ . \end{aligned}$$

Probability is thus not conserved with respect to internal 'time' τ. Choose the normalization of the wave packet such that at α_0 it is normalized to unity, $\int \mathrm{d}\phi \, |\psi|^2 = 1$. Then,

$$\psi(\alpha, \phi) = \frac{1}{C_{\mathrm{norm}}} \sum_{n=1}^{\infty} \left[A_n^+ + c_1 A_n^-\right] C_n(\alpha) \varphi_n(\alpha, \phi) \ , \tag{8.24}$$

where the normalization factor is given by

$$C_{\mathrm{norm}} \equiv \sqrt{\frac{\hbar^2}{\tilde{V}_0 \tau_0} \sum_{n=1}^{\infty} \left[A_n^+ + c_1 A_n^-\right]^2} \ .$$

A plot of the wave packet is shown in Figure F.3. Let me make some technical remarks concerning the plot of the wave packet before discussing the features of the packet itself. The wave packet is given by an infinite series. In any numerical evaluation this series has to be truncated at some point. As becomes obvious from Figure F.1, no new information is obtained if more than 20 terms are taken into account. Therefore, the series evaluated for the plot terminates at $n = 20$.

We recognize that the wave function is peaked around the two branches of the classical trajectory in configuration space, but goes to zero if the

region of the classical big-brake singularity, $a \to a_\star$, is approached. In this sense the classical singularity is avoided in the quantum theory. This is a consequence of the choice of the normalizable solution (8.12), which vanishes at $\phi = 0$ — the region of the big-brake singularity. Moreover, it can be seen directly in Figure F.3 or read off from Figure F.2 that the packet *disperses* long before the classical singularity is approached.

8.3 Remarks on the big-bang singularity

8.3.1 Solution to the Wheeler–DeWitt equation

So far, only the *big-brake* singularity of the model was considered. But the model possesses a second singularity. Namely, its evolution starts with a big bang: as $a \to 0$, one has $|\phi| \to \infty$. Thus one can approximate the potential by an exponential in the vicinity of this singularity — and such an equation was already solved in Section 7.4.2. The calculation is analogous to the one presented there. Namely, we start from the Wheeler–DeWitt equation with approximated potential

$$\frac{\hbar^2}{2}\left(\frac{\kappa^2}{6}\frac{\partial^2}{\partial\alpha^2} - \frac{\partial^2}{\partial\phi^2}\right)\psi(\alpha,\phi) + \frac{\tilde{V}_0}{2}e^{6\alpha+\sqrt{3}\kappa|\phi|}\psi(\alpha,\phi) = 0 \ .$$

Introducing $z_1 = \frac{\sqrt{6}}{\kappa}\alpha + |\phi|$, $z_2 = \frac{\sqrt{6}}{\kappa}\alpha - |\phi|$, one arrives at

$$\hbar^2 \frac{\partial^2}{\partial z_1 \partial z_2}\psi(z_1,z_2) = f(z_1,z_2)\psi(z_1,z_2) \ .$$

Choosing again coordinates such that $f(z_1,z_2)$ cancels, one arrives at

$$\hbar^2\left(\frac{\partial^2}{\partial u^2} - \frac{\partial^2}{\partial v^2}\right)\psi(u,v) + \psi(u,v) = 0 \ , \tag{8.25}$$

where

$$u(\alpha,\phi) = \sqrt{2V_0}\frac{a_{\mathrm{ref}}^3}{9}e^{3\alpha\pm\frac{\sqrt{3}\kappa}{2}|\phi|}\left[\cosh X \mp \frac{1}{\sqrt{2}}\sinh X\right] \ ,$$
$$v(\alpha,\phi) = \sqrt{2V_0}\frac{a_{\mathrm{ref}}^3}{9}e^{3\alpha\pm\frac{\sqrt{3}\kappa}{2}|\phi|}\left[\sinh X \mp \frac{1}{\sqrt{2}}\cosh X\right] \ ,$$

and $X = \sqrt{\frac{3}{2}}\kappa\left(\phi \pm \frac{\sqrt{3}}{\kappa}\alpha\right)$. A solution to this equation can be found from the WKB–ansatz $\psi = \int \mathrm{d}n A(n) e^{\pm\frac{i}{\hbar}S_{0n}}$. Inserting this ansatz into (8.25) yields the Hamilton–Jacobi equation of which an exact solution is given by $S_{0n}^{(1)} = nu - \sqrt{n^2-1}v$. Of course, the Hamilton–Jacobi equation is also solved by actions with different signs in front of u and v. These are obtained from the one chosen above through rotations in the (u,v)-plane.

As $u > 0$, only two solutions can be mapped onto each other and $S_{0n}^{(2)} = -nu - \sqrt{n^2-1}v$ provides a second solution.

8.3.2 Recovery of classical trajectories

The classical trajectory in the vicinity of the big bang is recovered using the principle of constructive interference $\frac{\partial S_{0n}}{\partial n}|_{n=\bar{n}} = 0$. For $\bar{n} = \sqrt{2}$ one finds $\phi(\alpha) = \pm\frac{\sqrt{3}}{\kappa}\alpha$. This is just the classical trajectory obtained from (8.7) in the limit $|\phi| \gg 1$ with initial condition $B = \frac{A}{4}$ and fixed A.

8.3.3 Construction of wave packets

We get the following exact wave-packet solution to the Wheeler–DeWitt equation:

$$\begin{aligned}\psi(u,v) &= \int \mathrm{d}n\, A(n)\left(C_1 e^{\frac{i}{\hbar}S_{0n}^{(1)}} + C_2 e^{-\frac{i}{\hbar}S_{0n}^{(1)}}\right) \\ &\quad + \int dn\, A(n)\left(C_3 e^{\frac{i}{\hbar}S_{0n}^{(2)}} + C_4 e^{-\frac{i}{\hbar}S_{0n}^{(2)}}\right) .\end{aligned}$$

By construction, the classical trajectories can be recovered from this equation through the principle of constructive interference. Choosing as amplitude a Gaussian with width σ centred around $\bar{n}$,

$$A(n) = \frac{e^{-\frac{(n-\bar{n})^2}{2\sigma^2\hbar^2}}}{(\sqrt{\pi}\sigma\hbar)^{1/2}} ,$$

and taking $C_1 = C_2$, $C_3 = C_4$ for definiteness, one obtains two contributions of the form

$$\psi(u,v) \approx C_i \pi^{1/4}\sqrt{\frac{2\sigma\hbar}{1 - i\sigma^2\hbar S_0''}}\exp\left(\frac{iS_0}{\hbar} - \frac{S_0'^2}{2(\sigma^{-2} - i\hbar S_0'')}\right) + \text{c.c.} , \quad (8.26)$$

where $i = 1$ for $S_0 = S_{0n}^{(1)}$ and $i = 3$ for $S_0 = S_{0n}^{(2)}$ and primes denote derivatives with respect to n. All this is completely analogous to the scalar field with exponential potential discussed in Section 7.4.2. As there, a Taylor expansion of S_{0n} has been carried out around $\bar{n}$. Terms of the order $(n-\bar{n})^3$ in the exponent have been neglected which is admissible if the Gaussian is strongly peaked around $\bar{n}$, that is, if σ is sufficiently small. Since $S_{0n}^{\prime(i)}(\bar{n}) = 0$ gives the classical trajectory, the packet is peaked around it.

8.3.4 Singularity avoidance

Here, we can argue as in Section 7.4.2. Due to the fact that $u > 0$, two inequivalent actions exist. Apart from the wave packet constructed from $S_{0n}^{(1)} = nu - \sqrt{n^2-1}\,v$, one gets a second wave packet constructed from $S_{0n}^{(2)} = -nu - \sqrt{n^2-1}\,v$. Moreover, the entire (α, ϕ) plane is mapped only into a quarter of the (u, v) plane. One would therefore require the wave packet to vanish on the boundary of the physical region. To get a solution which does not vanish on the classical trajectory, we require $\psi = 0$ at $u = 0$. But u vanishes for $\alpha \to -\infty$, i.e. at the big bang. We thus simply require that the wave packet vanishes at the big bang. This requirement is forced upon us through the coordinates in which the Wheeler–DeWitt equation is exactly solvable. As $v = 0$ for $\alpha \to -\infty$ as well, the requirement that $\psi \to 0$ as $\alpha \to -\infty$ amounts to requiring that the wave function vanishes at the origin of the (u, v) plane. The implementation of this condition results in a wave packet which vanishes at the big-bang singularity, $\psi \to 0$ as $\alpha \to -\infty$, and spreads for large α. Thus, again, we find that the big-bang singularity is resolved through a vanishing wave function.

But if we take $\alpha \to -\infty$ along the classical trajectory, we move into a region were $|\phi| \to \infty$. Our wave packet follows the classical trajectory and thus we find that $\psi \to 0$ for large $|\phi|$.

The condition implied for large $|\phi|$ in the vicinity of the big bang thus implies and justifies the normalization condition imposed in the derivation of the solution to the Wheeler–DeWitt equation in the vicinity of the big brake, cf. (8.11). We thus impose basically two conditions on the wave function. The first one is that $\psi \to 0$ when $|\phi| \to \infty$, resulting in a normalization condition for the approximate solution in the vicinity of the big-brake singularity and the elimination of the big-bang singularity. The second condition is to require $\psi \to 0$ as $\alpha \to \infty$ to ensure the existence of wave packets that follow the classical trajectory.

Note that because we require $\psi \to 0$ as $\alpha \to \infty$ and as $\alpha \to -\infty$, the wave function is fixed at both ends. Upon matching the solution for large $|\phi|$ to the one obtained for small $|\phi|$, quantization conditions should arise, cf. Figure 8.3.

8.4 Relation to Loop Quantum Cosmology

Loop quantum gravity arose as an alternative proposal to quantum general relativity, the main difference being the choice of fundamental variables, entailing a representation of operators inequivalent to the Schrödinger representation. The symmetry reduced 'minisuperspace' model of loop quantum

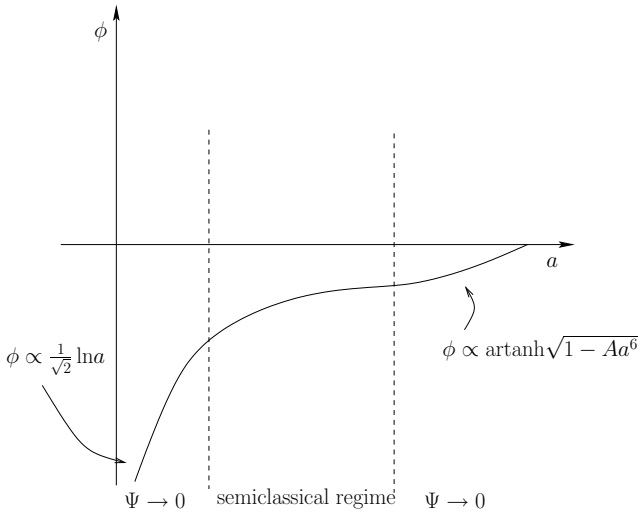

Figure 8.3: The different regions of the wave packets for one of the branches of the classical trajectory are shown. For small and large scale factor, the universe undergoes a quantum phase.

gravity goes under the name of loop quantum cosmology.[6] As loop quantum gravity is also a canonical quantization scheme, the governing equation in the cosmological context is the Hamiltonian constraint. On the quantum level, the universe is likewise described via a wave function on configuration space which has to be the solution of the *quantized* Hamiltonian constraint.

The difference between both approaches lies in the way this equation is quantized. In loop quantum cosmology, one uses a so-called polymer representation instead of the conventional Schrödinger representation. This is done in analogy to the full theory. If this procedure is carried out in a naive way, it leads to a difference equation in steps of a smallest length μ_0. In geometrodynamics, one arrives at a differential equation, the Wheeler–DeWitt equation. In the continuum limit, $\mu_0 \to 0$, suitable conditions on the higher derivatives of the wave function implied, the loop quantum cosmological difference equation fades into the Wheeler–DeWitt equation [73].

The question here is under which conditions the continuum limit is justified. It is justified if the discreteness of spacetime is negligible compared to the length scales occurring in the model. For large scale factor, $a \gg \mu_0$, one can argue that the limit $\mu_0 \to 0$ is a sensible approximation. Thus singularity avoidance for large-scale singularities as, for example, the big rip

[6] I will use the name loop quantum cosmology here exclusively for the loop quantized Friedmann–Robertson–Walker models, simply for the sake of brevity.

or big brake, in loop quantum cosmology reduces to singularity avoidance induced by the Wheeler–DeWitt equation.

At the time when this research was carried out, several versions of the Hamiltonian constraint in loop quantum cosmology were published. Ashtekar *et al.* [75, 76] extended the ansatz using μ_0, replacing it by $\bar{\mu}$, which is a function of the densitized triad operator $\hat{p}$. The equation is then a difference equation in eigenvalues v of the volume operator, and the Wheeler–DeWitt equation follows in the continuum limit for large volume. The factor ordering of the Wheeler–DeWitt equation then does depend on the factor ordering chosen for the difference equation. In [73] and [75, 76] different factor orderings have been chosen.

The two difference equations, in μ_0 or $\bar{\mu}$, can be understood in a broader context as implementing different actions of the full Hamiltonian constraint. They are thus just two special cases of a wider class of constraints that might arise, the actual form of which should in principle be determined by the full Hamiltonian constraint, [74]. Whereas in the first case, the coordinate edge length of a holonomy is fixed and does not depend on the scale factor, in the second case it does. This can be interpreted as an implementation of the fact that in the full theory, the Hamiltonian constraint, whatever its exact form may be, creates vertices — in addition to changing the edge labels of the existing edges. As new vertices are created, the edge lengths decrease. The altered dynamics using $\bar{\mu}$ then corresponds to a lattice in which the number of vertices grows linearly with volume.

In the following, both factor orderings, [73] and [75, 76], will be considered.

8.4.1 Non-covariant factor ordering

The Wheeler–DeWitt equation emerging in the continuum limit of the difference equation employed in [73] is

$$\frac{\hbar^2}{2}\left[\frac{\kappa^2}{6}a^2\frac{\partial^2\psi}{\partial a^2} - \frac{\partial^2\psi}{\partial\phi^2}\right] - a^6\frac{\tilde{V}_0}{|\phi|}\psi = 0 \ ,$$

which differs from (8.10) by the choice of factor ordering. Making the ansatz $\psi(a,\phi) = \sum_n A(n)C_n(a)\varphi_n(a,\phi)$ and requiring $\varphi_n(a,\phi)$ to be a solution of

$$\left(\frac{\hbar^2}{2}\frac{\partial^2}{\partial\phi^2} + a^6\frac{\tilde{V}_0}{|\phi|}\right)\varphi_n(a,\phi) = -E_n(a)\varphi_n(a,\phi) \ ,$$

one finds as before the solution

$$\varphi_n(x_n) = N_n x_n e^{-\frac{x_n}{2}} L^1_{n-1}(x_n) \ ,$$

where $x_n = 2\sqrt{-\frac{2E_n(a)}{\hbar^2}}|\phi|$ and $E_n(a) = -\frac{1}{2\hbar^2 n^2}\tilde{V_0}^2 a^{12}$, $N_n = n^{-\frac{3}{2}}$. Then the equation for $C_n(a)$ is given by

$$\frac{\mathrm{d}^2 C_n(a)}{\mathrm{d}a^2} - \frac{6\tilde{V_0}^2}{\hbar^4 n^2 \kappa^2} a^{10} C_n(a) = 0 \ ,$$

which is solved by

$$C_n(a) = c_1 \sqrt{a}\, J_{\frac{1}{12}} \left(\frac{1}{6} \sqrt{-\frac{6\tilde{V_0}^2}{\hbar^4 n^2 \kappa^2}} a^6 \right) + c_2 \sqrt{a}\, Y_{\frac{1}{12}} \left(\frac{1}{6} \sqrt{-\frac{6\tilde{V_0}^2}{\hbar^4 n^2 \kappa^2}} a^6 \right) \ .$$

The complete solution has an analogous form to the quantum geometrodynamical formulation in Section 8.2.1. The decisive result is that, because only the factor ordering of the *gravitational part* has been changed compared to (8.10), the solution for $\varphi_n(\phi, a)$ handles the singularity avoidance in this framework as well.

8.4.2 Covariant factor ordering

The factor ordering in the more recent papers [75, 76] yields the Laplace–Beltrami factor ordering for the Wheeler–DeWitt equation in the continuum limit. As this is the factor ordering we employed throughout this chapter anyway, the results of the previous sections carry over to the loop quantum cosmology analysis without alteration.

Note, though, that a consistent loop quantization requires a polymer representation of the matter fields as well. This would require a Bohr compactification of ϕ which may bound the approximate potential $V(\phi) = \frac{\tilde{V_0}}{|\phi|}$ from above. As the vanishing of the wave function at $\phi = 0$ is related to the divergence of the potential at this point, it is not clear whether the previous results would survive in the polymer representation. Namely, it is imaginable that the regularity condition and thus the ensuing result that $\varphi_n(\phi = 0, \alpha) = 0$ becomes redundant.

Qualitatively, the 'loop quantization' of ϕ would result in a difference equation in a and ϕ. In the vicinity of the big bang, this equation can be approximated by a differential equation in ϕ and a difference equation in a. In the vicinity of the big brake, however, the variable a can be considered continuous and obeys a differential equation. But now ϕ is still described by a difference equation.

8.5 Conclusion

I studied a Friedmann–Robertson–Walker model with a scalar field obeying an *anti-Chaplygin* equation of state. This model classically ends with a big-brake singularity. The singularity stands out because of its negatively diverging second derivative of the scale factor. This works as an infinitely strong 'brake', forcing the derivative of the scale factor to go to zero.

Upon quantizing this model in the quantum geometrodynamical framework, we are led to the Wheeler–DeWitt equation. It can be solved in the vicinity of the big-brake singularity. A separation ansatz yields a Schrödinger-type equation for the hydrogen atom for ϕ which here plays the role of the radius in the quantum mechanical equation. Normalizable solutions to this equation vanish at $\phi = 0$, which corresponds to the singularity. Thus, independent of the choice of initial conditions, whatever wave packet is constructed out of these solutions, it is condemned to vanish at the singularity. Therefore we can conclude that in this model as well, the large scale, soft, future singularity is removed from the quantum theory.

This singularity resolution is different from the one encountered for the big-rip singularity. It is twofold. First of all, the wave function itself vanishes at the big brake. This feature depends crucially on the fact that the *normalizable* solutions for the matter-dependent part of the wave function have been chosen. This choice is justified through the study of the same model in the vicinity of its second singularity, the big bang. The fact that a solution could be found exactly in this region is due to the special form of the potential. But more than that, the wave packet spreads as the big-brake singularity is approached.

The choice of variables enforces a boundary condition which causes the wave function to vanish at the big bang. This singularity is thus also eliminated in the quantum theory. The boundary condition at the big bang is, again, just DeWitt's boundary proposal. It is this condition which justifies the choice of normalizable solutions in the vicinity of the big-brake singularity. So in this sense, vanishing of the wave function at the big brake and at the big bang are connected. Better, the vanishing at the big-brake singularity is a consequence of the vanishing at the big bang.

The imposition of boundary conditions on both ends of the evolution, near the big bang and near the big brake, should imply some kind of quantization rule upon matching the wave packets in both regimes. Such a matching has not been carried out.

References: This work was published in [48].

Chapter 9

Effective constraints for quantum systems

The last chapter is devoted to the application of a generalized effective formulation of quantum theory to constrained systems. This formulation is based on quantum phase space coordinatized by expectation values of fundamental operators and moments of the wave function. As a generalization of the effective action, such a formulation is of interest in its own right. Moreover, the hope is to gain some insight into the internal-time approach.

9.1 Motivation

9.1.1 From the quantum General Relativity side

Canonical general relativity obeys a constrained Hamiltonian dynamics. The arisal of constraints signals that one is dealing with more variables then are actually necessary to describe the physical properties of the system. At some point, one has to get rid of these superfluous variables — either before or after quantization. Canonical general relativity is characterized by four constraints at each point in Σ. The Hamiltonian constraint generates evolution from one hypersurface to the adjacent one. The diffeomorphism constraints generate diffeomorphisms in Σ. The corresponding *quantum* diffeomorphism constraints can easily be interpreted. They insure that the wave function does not depend on the three-*metric* but only the three-*geometry*. So it seems that is does not matter whether we solve this constraint on the classical level and quantize on the space of three-geometries (superspace) or whether we quantize on the space of all three-metrics and reduce the metric

dependence through the quantum diffeomorphism constraints.[1]

The issue is somewhat more subtle in the case of the Hamiltonian constraint. On the classical level, it encodes time evolution.[2] So we expect that on the quantum level it does the same. The problem is just that no foliation and thus *no time exists on the quantum level*. In fact, the quantum Hamiltonian constraint, i.e. the Wheeler–DeWitt equation, is a second-order, functional differential equation in which all fields occur on an equal footing.

The fact that the quantum Hamiltonian has to encode time in some way but does not refer to any time variable, is known as *the problem of time*. At its heart lies the observation that the three-metric contains information about time in an inextractable way. This follows by a simple counting argument. The six degrees of freedom of the three-metric are reduced to three through the diffeomorphism constraints. So one of the remaining three has to be unphysical and thus related to time. The question is just — which one?

There are several attitudes towards this problem. The first decision one has to make is whether one wants to eliminate this superfluous variable already on the classical level or keep it on the quantum level. In Isham's terminology, we have to decide whether we want to identify *time before quantization* or *time after quantization*. Note that identifying time before quantization would yield a theory with classical time parameter and a Hamiltonian which is no longer constrained to vanish but generates evolution just in that parameter. On the quantum level, we would thus obtain a Schrödinger equation. This may seem very appealing because we are back to something we know. But, as I pointed out in Chapter 3, we should not expect to recover too many familiar features when we quantize gravity. Moreover, we do not know *which* of the three degrees of freedom we should identify with time. But each choice produces a unitarily inequivalent quantum theory. In addition to that, there are indications that time may not be separable from three-geometry, [82]. And, most convincingly, the constraints of general relativity cannot be solved globally, [29], implying that such a time variable cannot be defined globally.

For all these reasons, we want to include the 'time' variable into the set of quantum operators. So we follow the *time after quantization* approach. This yields the already known Wheeler–DeWitt equation.

Even though it is clear now that we are going to identify time after quantization, it is less clear how this identification should be carried out. In the previous chapters, the Wheeler–DeWitt equation was treated as a

[1] But, of course, it is mathematically simpler to quantize on the space of all three-metrics and make the reduction on the quantum level by simply changing the domain of dependence of the wave function to superspace. We want to emphasize that this is done only formally.

[2] Time here refers to the foliation parameter.

timeless equation. The concept of time was entirely reserved for the semi-classical realm. Here, *time* referred to the foliation time recovered from the Wheeler–DeWitt equation in the WKB-limit. There is just one concept of time in this interpretation and this is the well-known classical foliation parameter.

However, this is not the only way in which one can identify time on the quantum level. Namely, one can try to identify a time parameter through its function. For example, the Klein–Gordon form of the Wheeler–DeWitt equation suggests to interpret the determinant of the three-metric or three-volume as time parameter. Such a time variable built up from the configuration space variables after quantization is called *internal time*. Note, however, that this concept of time is unrelated to the classical concept of time. It refers to the fact that one of the quantized degrees of freedom plays the *role* of a time variable, i.e. has the same function. But it has nothing whatsoever to do with a time-coordinate of an ordinary spacetime. The internal-time variable defines 'equal-time' hypersurfaces in superspace on which inner products are defined through the integration over all remaining degrees of freedom. Therefore different choices of internal time result in unitarily inequivalent theories.

The internal-time approach clearly transcends the framework of ordinary quantum field theory in that it interprets 'equal-time' hypersurfaces in configuration space analogously to spatial hypersurfaces in ordinary spacetime. Most significantly, expectation values of observables obtained through an integration over hypersurfaces of constant internal time are interpreted in terms of measurement results. This is a non-trivial generalization of the interpretational framework of ordinary quantum field theory. The previous, semi-classical time interpretation is clearly the more conservative framework.

Despite these warnings, the internal-time idea provided one major motivation for the project discussed in this chapter. The hope was that a different formulation of the same theory might shed light on the problems of the internal-time approach.

Such an alternative formulation of quantum theory is provided by the geometrical formulation which goes back to Kibble, [79]. The basic observation is that any quantum theory formulated on a Hilbert space can be alternatively formulated on a generally infinite-dimensional phase space, the so-called *quantum phase space*.[3] Schrödinger dynamics is then replaced by Hamiltonian dynamics. The central quantity is no longer the wave function but expectation values of fundamental operators. More specifically, one can choose expectation values of canonical variables and moments of the wave function as coordinates on this quantum phase space. The *quantum Poisson bracket* between any two functions on this phase space is, up to a prefactor,

[3]This quantum phase space is finite dimensional for systems with finite-dimensional Hilbert space, e.g. spin systems.

the expectation value of the commutator of the same function with operators as arguments. Dynamics is then given by a *quantum Hamiltonian* in the form of Hamilton's equations of motion — but employing the quantum Poisson bracket, of course. This is the already known set-up of the geometrical formulation of quantum mechanics — or quantum field theory for that matter. Assume now that moreover a set of constraints on quantum phase space has been found which is equivalent to Dirac's constraint quantization condition.

Why should such a setup be a promising starting point to study the internal-time approach?

Because a quantum constrained system in the geometrical formulation is described by an infinite-dimensional phase space equipped with a number of constraints. Thus Dirac's constraint quantization condition can be dealt with in the language of classical, constrained systems — and therefore we can draw on the familiar concepts and methods.

Choosing one variable as internal time corresponds to the elimination of the corresponding canonical pair on quantum phase space and all degrees of freedom associated with it. But this can then be done in full analogy to a classical, constrained system.

After such an elimination has been carried out, we can analyse the structure of the resulting system. That means we can check whether the remaining operators can be formulated as self-adjoint ones on a Hilbert space. This can then be done by simply requiring that all remaining variables on quantum phase space be real.[4]

Mathematically, this sounds very straight and simple, not least because everything is formulated in the well-known language of classical constrained systems. Theories arising for different choices of internal time and their relations can be studied. Probing the resulting theory for a Hilbert space structure seems also pretty easy — because we do not have to write down the inner product explicitly but simply implement *reality conditions*.

The only crux is that quantum phase space is generally infinite dimensional. Practically, it may turn out very difficult to solve a system of coupled, even though *ordinary*, differential equations. Approximation schemes will generally be needed to reduce this system to a *finite* number of coupled equations. Such approximations can correspond to perturbations of an arbitrary state, or even an arbitrary class of states. These can be semi-classical states, but also coherent or squeezed states, for example.

Moreover, there exist systems for which the infinitely many equations decouple. These are called *solvable systems*. They can be used as starting point of a perturbation series. This is then analogous to the perturbation around free field theories usually employed for interacting ones in quantum

[4]Recall that quantum variables are just expectation values and that the expectation value of any phase space operator function can be expanded in terms of them.

field theory.

Therefore the generalization of the effective equation scheme is a first step in the study of the internal-time approach.

9.1.2 From the quantum theory side

One may also approach the present chapter from a completely different, but much more general perspective.

The geometrical formulation of quantum theory serves as the starting point for the derivation of so-called *generalized effective equations*. Effective equations describe quantum motion in an approximate manner through equations of the classical type.

A well-known and successful example of such a framework is the *effective action*. It is derived from the path-integral in quantum field theory (or quantum mechanics for that matter) through a perturbation around the mean field. In the absence of an external source, the mean field is just the vacuum expectation value of the field itself. Otherwise the mean field corresponds to the matrix element of the field operator between two different in and out vacua. The assumption is that quantum fluctuations around the mean field are small. At the end of a series of mathematical tricks then stands an effective action which equals the classical action plus quantum corrections,

$$\Gamma_{\text{eff}}[\langle\phi\rangle] = S[\langle\phi\rangle] + \sum_k \hbar^k \Gamma_{k-\text{loop}}[\langle\phi\rangle] \ ,$$

where $\langle\phi\rangle$ is the mean field, S the classical and Γ_{eff} the effective action. $\Gamma_{k-\text{loop}}$ correspond to k-loop corrections. From the $\hbar$-expansion we read off that the formalism works only in the semi-classical regime.

Furthermore, the mean field as matrix element between in and out vacua ties the effective action to scattering-like setups.

Moreover, the effective action is tied to the covariant framework. It is interesting in its own right to look, firstly, for a *generalization* of this effective scheme which, secondly, applies to Hamiltonian formulations.[5] Also, the effective action uses an expansion around the vacuum in the case of a free theory. But it may be desirable to expand around some other state but the vacuum. Most importantly in this context, for general relativity the vacuum is not known.

Speaking of a generalization implies, of course, that the original scheme is comprised in the general one.

Such a generalization of the effective action scheme is provided by a formulation that goes under the somewhat lengthy name of *generalized ef-*

[5] One may invoke that the WKB-approximation is the canonical scheme corresponding to the effective action. But still, the WKB-approximation is tied to the semi-classical regime.

fective equations. These can be derived from the geometrical formulation outlined above through suitable expansions around *arbitrary states* — not only semi-classical ones.

So, from a more general perspective, the present chapter is concerned with the question of how this generalization of the effective-action scheme might be formulated for *constrained systems*. And even though this thesis deals with quantum cosmology, I want to see this as a major motivation behind the work presented below. Moreover, a consistent formulation of constraints in the generalized effective-equation framework has to be *developed* before any questions on the internal-time approach can be answered — or even be posed.

9.2 Geometrical formulation

The basic structures of the geometric formulation of quantum theory are easily understood. Let the Hilbert space of the theory be denoted by $\mathfrak{H}$ with Hermitian inner product $\langle \cdot, \cdot \rangle$. Now regard $\mathfrak{H}$ as a *real* vector space, equipped with complex structure J which just represents multiplication by the imaginary unit, i. Correspondingly, split the Hermitian inner product into its real and imaginary parts

$$\langle \Psi | \Phi \rangle = \frac{1}{2\hbar} G(\Psi, \Phi) + \frac{\mathrm{i}}{2\hbar} \Omega(\Psi, \Phi) \ ,$$

where the prefactor is chosen out of convenience and Ψ, $Phi \in \mathfrak{H}$. Then G defines a positive inner product and Ω a symplectic form. Both are non-degenerate. Canonically identifying the tangent space at each point of $\mathfrak{H}$ with $\mathfrak{H}$ itself, Ω extends to a non-degenerate, closed two-form. It follows immediately that every Hilbert space can be interpreted as a phase space. Now define for every self-adjoint operator $\hat{F}$ the Schrödinger vector field

$$Y_{\hat{F}}(\Psi) = -\frac{1}{\hbar} J \hat{F} \Psi \ .$$

Furthermore define the expectation value function

$$F: \qquad \mathfrak{H} \longrightarrow \mathbb{R} \ , \qquad F(\Psi) = \langle \Psi | \hat{F} \Psi \rangle \ .$$

The Hamiltonian vector field X_F generated by F then coincides with the Schrödinger vector field $Y_{\hat{F}}$. From this follows immediately that the Schrödinger equation in this formulation is just Hamilton's equation,

$$\dot{\Psi} = Y_{\hat{H}}(\Psi) = X_H(\Psi) \ .$$

Moreover, the commutator defines a quantum Poisson bracket for the expectation values via

$$\{F, K\}_\Omega = \Omega(X_F, X_K) = \frac{1}{\mathrm{i}\hbar}\langle\Psi|[\hat{F}, \hat{K}]|\Psi\rangle \ . \tag{9.1}$$

But, of course, the space of physical states is not the Hilbert space $\mathfrak{H}$, but the space of *rays* in the Hilbert space. Wave functions which differ only by a complex factor are physically equivalent. This fact is taken care of if one restricts attention to those elements of $\mathfrak{H}$ which have unit norm. The resulting space is a projective Hilbert space denoted by $\mathfrak{P}$. The above definitions carry over to the new, projective Hilbert space if one considers $\mathfrak{P}$ as the reduced phase space arising under the first-class constraint

$$C = \langle\Psi|\Psi\rangle - 1 \ .$$

The resulting reduced phase space is called *quantum phase space*. This phase space is, generally, infinite dimensional.[6]

9.3 Generalized effective equations — the setting

The geometrical formulation gave the idea for the effective framework. But we will see that the framework presented in this section can be seen as completely independent from the geometrical formulation.

The effective-equation formalism is set up in quantum phase space. We equip this space with coordinates suited for our purpose. These coordinates are given by the expectation values of fundamental operators, in the following simply $\hat{p}$ and $\hat{q}$. Expectation values of higher powers of these operators are included in the following way

$$G^{a,b} = \left\langle (\hat{p} - \langle\hat{p}\rangle)^a (\hat{q} - \langle\hat{q}\rangle)^b \right\rangle_{\mathrm{Weyl}} \ , \tag{9.2}$$

where $a, b \in \mathbb{N}$ such that $a + b \geq 2$ and the totally symmetric ordering is used, indicated by the index $\langle\cdot,\cdot\rangle_{\mathrm{Weyl}}$.

For $a+b = 2$, this provides fluctuations $(\Delta q)^2 = G^{0,2} = G^{qq}$ and $(\Delta p)^2 = G^{2,0} = G^{pp}$ as well as the correlation $G^{1,1} = G^{qp}$. Generally, $G^{a,b}$ are just the moments of the wave function. I will also speak of *quantum variables* when I refer to these moments, because these are just the quantities that incorporate the quantum properties. Whereas one can think of $\langle\hat{p}\rangle$ and $\langle\hat{q}\rangle$ as the peak position of a wave packet which, in a semi-classical state, simply follows the classical trajectory, quantum variables describe deviations from this classical path. As long as they are negligible, the wave packet describes an approximate classical path. But as soon as higher moments become important, the wave packet spreads and its motion is no longer well described by a classical path.

[6] Finite dimensional quantum phase spaces arise, e.g., for spin systems.

So instead of making the detour via the geometrical formulation, I could equally well have started with the statement that we describe a state by its moments rather than a wave function in a certain Hilbert space representation. Such a setting has the immediate advantage that the description is manifestly representation independent and deals directly with quantities of physical interest, namely expectation values and fluctuations.

Before continuing, let me spend some words on notation. In the following, I will denote expectation values simply by 'un-hatting' the corresponding operator, e.g. q for $\langle\hat{q}\rangle$. Also, I will often use p and q as indices instead of numbers on low moments, as already introduced above, e.g. $G^{0,2} = G^{qq}$. The largest system that is going to be considered here, has two degrees of freedom. The notation for quantum variables of such a system is just a straight generalization of (9.2), namely

$$G^{a,b}_{c,d} \equiv \langle(\hat{p}-p)^a(\hat{q}-q)^b(\hat{p}_1-p_1)^c(\hat{q}_1-q_1)^d\rangle_{\rm Weyl} \ ,$$

for a system with canonical pairs $(q,p;q_1,p_1)$. For these as well, the more intuitive notation using a direct listing of canonical variables, as in $G^q_{p_1} = G^{0,1}_{1,0}$ will be used. The *order* M of a quantum variable is defined as the total power of contributing operators, i.e. $M = a+b+c+d$ in the case of the two-dimensional system. I hope the notation is sufficiently intuitive for the reader to follow easily.

The symplectic structure is obtained from (9.1). But for simplicity, I will drop the index Ω which denotes that we are actually dealing with the *quantum* Poisson bracket. This can be safely done as no reference is made to the classical Poisson bracket anywhere in this chapter.

For functions of expectation values p and q, the quantum Poisson bracket just yields the same result as would the *classical* Poisson bracket for *classical* phase space functions of the same form. Expectation values have vanishing Poisson brackets with the moments of the wave function.[7] The Poisson bracket between quantum variables is, on the other hand, fairly complicated,

$$\begin{aligned}\{G^{a,b},G^{c,d}\} = &\sum_{r,s=0}^{\infty}(-\tfrac{1}{4}\hbar^2)^{r+s}\times\\ &\sum_{j,k}\binom{a}{j}\binom{b}{k}\binom{c}{k}\binom{d}{j}G^{a+c-j-k,b+d-j-k}(\delta_{j,2r+1}\delta_{k,2s}-\delta_{j,2r}\delta_{k,2s+1})\\ &-adG^{a-1,b}G^{c,d-1}+bcG^{a,b-1}G^{c-1,d}\ ,\end{aligned} \tag{9.3}$$

where the summation range of j and k is given by $0 \le j \le \min(a,d)$ and $0 \le k \le \min(b,c)$, respectively. Therefore, low order moments are easier

[7] Actually, it is this feature which motivated the choice of coordinates in the beginning, [81].

calculated via expectation values of commutators. In the original paper by Bojowald and Skirzewski, [81], the general Poisson bracket between quantum variables for systems with an *arbitrary* number of degrees of freedom is also given.

Now we still lack the dynamics. The Schrödinger equation is replaced by the Hamiltonian equations of motion on quantum phase space. These are determined by the quantum Hamiltonian which is just the expectation value of the Hamiltonian operator. Carrying out a Taylor expansion around the expectation values, it can be expressed in terms of our coordinates as

$$\begin{aligned} H_Q(q,p,G^{a,b}) &= \langle H(\hat{q},\hat{p})\rangle_{\rm Weyl} = \langle H(q+(\hat{q}-q),p+(\hat{p}-p))\rangle_{\rm Weyl} \\ &= H(q,p)+\sum_{a=0}^{\infty}\sum_{b=0}^{\infty}\frac{1}{a!b!}\frac{\partial^{a+b}H(q,p)}{\partial p^a\partial q^b}G^{a,b}\ , \end{aligned} \tag{9.4}$$

where we understand $G^{a,b}=0$ if $a+b<2$ and $H(q,p)$ is the classical Hamiltonian evaluated on expectation values. As written explicitly, we assume the Hamiltonian to be Weyl ordered. This is done for definiteness. Any other ordering can be reduced to Weyl ordering by adding re-ordering terms and vice versa. Note that (9.4) is *in form* already very close to the effective action — it also consists of a classical contribution plus quantum corrections.

Having a Hamiltonian and Poisson brackets, one can compute Hamiltonian equations of motion

$$\dot{q}=\{q,H_Q\}\ ,\qquad \dot{p}=\{p,H_Q\}\ ,\qquad \dot{G}^{a,b}=\{G^{a,b},H_Q\}\ .$$

Of course, these equations generalize to any system with more than one degree of freedom. This ends our reformulation of the quantum system.

So, in sum, we traded the Schrödinger equation for an *infinite* set of ordinary differential equations. Even worse, in general, these equations are all coupled to each other. But such a bad deal can actually turn into a good stroke of business.

First and foremost, there are systems for which the equations of motion for expectation values and higher moments decouple. In this case, the system of equations is *solvable*. This is the case for the harmonic oscillator, for example. I will follow the terminology of [81] and call such systems *solvable systems*. One can expand around such systems. In this way, one gets a chance to solve also *non-solvable* models. The anharmonic oscillator can, for example, be considered as a perturbation of the harmonic oscillator. In this sense, solvable systems provide a basis for perturbation theory to analyse more general systems — just like free quantum field theory provides a solvable basis for interacting ones.

Moreover, semi-classical and some other regimes, depending on the system under study, allow one to truncate and decouple the equations consistently. In this way, one arrives at a *finite* set of ordinary differential equations.

Lastly, I want to comment briefly on the relation of this scheme to the effective action. Using the example of the anharmonic oscillator, it was shown that the equations of motion obtained from this effective-equation scheme coincide with those derived from the effective action — for a special choice of parameters. These parameters arise as integration constants from the Hamiltonian equations of motion. In the generalized effective-equation formalism, these are arbitrary — corresponding to an arbitrary initial state. But for a specific choice, one recovers effective action results. In this way, one sees directly that generalized effective equations allow a wider class to expand around than the effective-action scheme.

9.4 Effective constraints

So far for the *unconstrained* system. But what I actually promised in the introduction, was a description for systems subject to constraints. To develop such a formulation, it seems sensible to start with the easiest case, namely *one* constraint. In the following, C will always refer to the classical constraint.

The characteristic property of reparametrization-invariant theories is that this one constraint is given by the Hamiltonian, $C = H = 0$. We will also consider this special case. Then we can further divide these systems into *artificially* reparametrization-invariant and *truly* reparametrization-invariant systems. Into the latter category fall all diffeomorphism-invariant theories. One can distinguish the two classes on the canonical level in a simple way: artificially reparametrization-invariant theories have at least one momentum that occurs only *linearly* whereas for truly reparametrization-invariant systems all momenta enter *quadratically* in the Hamiltonian.

In this chapter, we will only make the first steps into the direction of an effective formalism for constrained systems. We will only deal with artificially reparametrization-invariant systems.

On the quantum level, assuming Dirac's quantization condition, the constraint translates into $\hat{C}|\Psi\rangle = 0$.

The main task in this section will be to reformulate Dirac's condition on quantum phase space.

9.4.1 Form of quantum constraints

What should we expect from such a reformulation? First of all, the resulting set of conditions should neither be stronger nor weaker than Dirac's

condition. This is the case if it eliminates exactly one degree of freedom. That means, through the set of effective constraints, expectation values of *one* canonical pair have to be eliminated. But more than that, all quantum variables *depending* on this canonical pair have to be removed. There are infinitely many of these and therefore we expect infinitely many effective constraints.

On the other hand, we know that $\hat{C}|\Psi\rangle = 0$. This means that all expectation values of the form

$$C^{(n)} := \langle \hat{C}^n \rangle = 0 \ , \tag{9.5}$$

$$C^{(n)}_{f(q,p)} := \langle f(\hat{q},\hat{p})\hat{C}^n \rangle = 0 \ , \tag{9.6}$$

for positive integer n and arbitrary phase space functions $f(q,p)$, will vanish. This is the *general* set of quantum constraints. The expectation value $C_Q = \langle \hat{C} \rangle$ will be called *principal constraint*. Note that for each constraint an expansion as in (9.4) can be carried out.

All concepts used in the theory of constrained systems can be carried over to the quantum constraint system. The vanishing of quantum constraints defines the *quantum constraint surface*. Quantum constraints whose Poisson brackets with all other constraints vanish on this surface are first-class constraints, the remaining ones are second class. Transformations generated by first-class quantum constraints are gauge transformations. A second-class function varies along some of these flows and is therefore gauge dependent. For the sake of brevity, I will often drop 'quantum' when no confusion can arise.

The quantum constraints satisfy the commutation relations

$$[\hat{f}\hat{C}^n, \hat{g}\hat{C}^m] = [\hat{f},\hat{g}]\hat{C}^{n+m} + \hat{f}[\hat{C}^n,\hat{g}]\hat{C}^m + \hat{g}[\hat{f},\hat{C}^m]\hat{C}^n \ , \tag{9.7}$$

whose expectation value in any physical state vanishes. Consequently, these constraints have vanishing quantum Poisson brackets on the constraint surface, thus providing a weakly commuting set,

$$\{C^{(n)}_f, C^{(m)}_g\} = \frac{1}{i\hbar}\langle [\hat{f}\hat{C}^n, \hat{g}\hat{C}^m] \rangle \approx 0 \ .$$

Therefore, the set of quantum constraints is first class.

Each quantum constraint thus removes variables in two ways. First of all, by simply solving the quantum constraint for one variable, the latter is eliminated from the theory. But in addition to that, first-class constraints generate gauge transformations on a phase space. So on quantum phase space, quantum constraints generate gauge transformations.

They transform states into physically equivalent states through

$$G^{a,b}_{c,d} \qquad \longrightarrow \qquad \tilde{G}^{a,b}_{c,d} = G^{a,b}_{c,d} + \epsilon_0 \{C^{(n)}_f, G^{a,b}_{c,d}\} \ ,$$

and similarly for expectation values.

One may not have expected that gauge flows generated by quantum constraints on the quantum phase space play an important role. In the usual Dirac treatment, only a constraint equation is written for states, but no gauge flow on the Hilbert space needs to be factored out. In fact, the gauge flow which one could define by $\exp(\mathrm{i}t\hat{C})|\psi\rangle$ for a self-adjoint $\hat{C}$ is trivial on physical states which solve the constraint equation, $\hat{C}|\psi\rangle = 0$. I.e. it does not change these states. On the other hand, recall the parametrized, free particle reviewed in Chapter 3. There, it was emphasized that $\hat{C}^\dagger$ generates the gauge flow on quantum states. However, such a procedure is never carried out in quantum general relativity.

How can we understand that the gauge flow becomes non-trivial in the phase-space formulation? Note that the gauge flow could only be trivial if we would assume self-adjoint constraint operators. Because only then $\langle\psi|\hat{C} = 0$ would follow from $\hat{C}|\psi\rangle = 0$ and the flow on the expectation values, $\langle\psi|\exp(-\mathrm{i}t\hat{C})\hat{A}\exp(\mathrm{i}t\hat{C})|\psi\rangle$ vanish. But constraint operators are generally not self-adjoint, recall Chapter 3 and [25].

The expectation values and moments we are dealing with when imposing quantum constraints thus have to form a much wider manifold than the Hilbert space setting would allow. Here, not only constraint equations but also gauge flows on the constraint surface are crucial.

From these very general considerations, we inferred the general structure of quantum constraints. But so far, $f(\hat{q},\hat{p})$ is any function of the operators. We thus have not only infinitely many, but even *uncountably* many constraints. Not all of them can be independent. So what we need, is a prescription that tells us which f we should choose.

The resulting infinite set should still be first class. Moreover, we have to insure that it removes the infinite number of variables on quantum phase space associated with one classical degree of freedom.

9.4.2 Iteration procedure for quantum constraints

Also, for practical purposes, one would like to keep the number of allowed functions minimal while keeping the system complete. Then, however, the set of quantum constraints is not guaranteed to be closed for any restricted choice of phase-space functions f used in their definition.

If $C_f^{(n)}$ and $C_g^{(m)}$ are quantum constraints, closure requires the presence of $C_{[f,g]}^{(n+m)}$ (for $n \geq 2$), $C_{f[C^n,g]}^{(m)}$ and $C_{g[C^m,f]}^{(n)}$ as additional constraints according to (9.7). In this way, a construction procedure for a closed set of quantum constraints is prescribed. It has to be iterated until a closed system is obtained. This iteration does not necessarily terminate after a finite number of steps.

But although many independent constraints have to be considered for

a *complete* system, most of them will involve quantum variables of a high order. To a given order in the moments, it is thus sufficient to consider only a finite number of constraints. Such truncations and approximations will be discussed by examples in Section 9.5.

9.4.3 Reality conditions

Note that the definition of constraints in (9.6) uses expectation values of non-symmetric operators, thus implying complex valued constraint functions. But this is no problem as we are working with *unconstrained* states. Only *after* the constraints have been solved, all expectation values should be real. Because solving the constraints corresponds to the transition from the *kinematical Hilbert space*, or *auxiliary Hilbert space*, to the physical Hilbert space.[8] Whereas the *kinematical* Hilbert space is just an auxiliary structure used to set up the theory, define operators etc., the *physical* Hilbert space contains physically admissible states and thus carries the standard interpretation. On the kinematical Hilbert space, operators *can* have imaginary eigenvalues — because we are still dealing with unphysical states and with an unphysical inner product. On the physical Hilbert space, however, expectation values should be real, as here they correspond to measurement results.

The transition from kinematical to physical Hilbert space then consists simply in the implementation of reality conditions for the physical variables.

So there is no problem in defining quantum constraints as complex functions. Moreover, in constrained theories it is even required to work with constraint operators which are not self-adjoint, cf. Chapter 3 again, and thus complex valued constraints have to be expected in general.

Furthermore, note that we *cannot* order symmetrically in (9.6). This would give rise to terms where some $\hat{q}$ or $\hat{p}$ appear to the right while others remain to the left. Thus the expression would not vanish for physical states and therefore not correspond to a constraint.

Still, one could avoid the question of reality of the constraints altogether by using quantum constraints defined as $G^{C^n f(q,p)} = \langle \hat{C}^n \widehat{f(p,q)} \rangle_{\mathrm{Weyl}}$ such as $G^{C^n q}$ and $G^{C^n p}$ with the symmetric ordering used as in (9.2). Here, the symmetric ordering contained in the definition of quantum variables must leave $\hat{C}$ intact as a possibly composite operator, i.e. we have for instance $G^{C,p} = \frac{1}{2}\langle \hat{C}\hat{p}+\hat{p}\hat{C}\rangle - Cp$ independently of the functional form of $\hat{C}$ in terms of $\hat{q}$ and $\hat{p}$. Otherwise it would not be guaranteed that the expectation value vanishes on physical states. We could not include variables with higher powers of q and p, such as $G^{C^n pp}$ as constraints because there would be terms in the totally symmetric ordering (such as $\hat{p}\hat{C}^n\hat{p}$) not annihilating a

[8] I find the name *kinematical Hilbert space* somewhat misleading because it suggests the existence of a 'dynamical Hilbert space' as opposite. However, it is this convention that dominates the literature and so I will succumb to convention.

physical state. But, e.g., $G^{\hat{C}\hat{p}^2}$ understood as $\frac{1}{2}\langle\hat{C}\hat{p}^2+\hat{p}^2\hat{C}\rangle - Cp^2$ would be allowed. The use of such symmetrically ordered variables would imply real quantum constraints.

However, this procedure cannot be reconciled with the requirement that the constraints be first class. We have, for instance,

$$\begin{aligned}\{G^{C^n f(q,p)}, G^{C^m g(q,p)}\} &= \frac{1}{4\mathrm{i}\hbar}\langle[\hat{C}^n\hat{f}+\hat{f}\hat{C}^n, \hat{C}^m\hat{g}+\hat{g}\hat{C}^m]\rangle \\ &\quad -\frac{g}{2\mathrm{i}\hbar}\langle[\hat{C}^n\hat{f}+\hat{f}\hat{C}^n, \hat{C}^m]\rangle - \frac{C^m}{2i\hbar}\langle[\hat{C}^n\hat{f}+\hat{f}\hat{C}^n, \hat{g}]\rangle \\ &\quad -\frac{f}{2\mathrm{i}\hbar}\langle[\hat{C}^n, \hat{C}^m\hat{g}+\hat{g}\hat{C}^m]\rangle - \frac{C^n}{2\mathrm{i}\hbar}\langle[\hat{f}, \hat{C}^m\hat{g}+\hat{g}\hat{C}^m]\rangle \\ &\quad +\{C^n f, C^m g\}\,.\end{aligned}$$

The first commutator contains, apart from several terms which vanish when the expectation value is taken in a physical state, also the two terms $[\hat{C}^n,\hat{g}]\hat{C}^m\hat{f}$ and $\hat{f}\hat{C}^m[\hat{C}^n,\hat{g}]$. Their expectation value only vanishes if $\hat{f}$ or $\hat{g}$ commute with $\hat{C}$. This would require quantum perennials to be known and used in the quantum constraints, which in general would be too restrictive and impractical. A further possibility using Weyl-ordered constraints of a specific form was discussed in [80], but seems to be less practical in concrete examples.

So, in general, there seems neither reason nor way to escape complex quantum variables on the kinematical level. We will even see that complex-valued quantum variables are actually helpful to ensure consistency.

9.4.4 Linear constraint operator: Number of effective constraints

We now have a prescription that, for a given phase-space function f, generates quantum constraints through an iteration procedure. The question remains, how f shall be chosen.

We will show here that for the general class of *linear* constraints, f can be chosen as polynomial function in the fundamental operators. The results are then *locally* valid for any system subject just to one, single constraint.

The number of degrees of freedom shall be finite, say $N+1$ and the canonical operators are $(\hat{q}^i,\hat{p}_i)_{i=1,\ldots,N+1}$ with the usual commutation relations $[\hat{q}^i,\hat{p}_j]=i\hbar\,\delta^i_j$. The system is subject to the linear constraint $\hat{C}$.

A first step that simplifies the analysis significantly, is to transform on variables in which the constraint is just one of the configuration variables. I.e. we introduce a new canonical pair $(\hat{q},\hat{p})$ which is defined through the requirements

$$\hat{q} = \hat{C} \qquad \text{and} \qquad [\hat{q}, \hat{p}] = \mathrm{i}\hbar \ .$$

Furthermore, the following commutation relations with the remaining variables $(\hat{x}_i)_{i=1,\ldots,2N}$ shall be satisfied

$$[\hat{q}, \hat{x}_i] = [\hat{p}, \hat{x}_i] = 0 \quad , \quad [\hat{x}_i, \hat{x}_j] = i\hbar\,(\delta_{i,j-N} - \delta_{i-N,j})$$

where $\hat{x}_i$ with $i = 1, \ldots, N$ and $i = N+1, \ldots, 2N$ correspond to the configuration and momentum operators, respectively.[9]

So in the following, our quantum system is parametrised by expectation values $q := \langle\hat{q}\rangle$, $p := \langle\hat{p}\rangle$, $x_i := \langle\hat{x}_i\rangle$, $i = 1, \ldots, 2N$ and quantum variables

$$G^{a_1,a_2,\ldots,a_{2N};b,c} = \left\langle (\hat{x}_1 - x_1)^{a_1} \cdots (\hat{x}_{2N} - x_{2N})^{a_{2N}} (\hat{p} - p)^b (\hat{q} - q)^c \right\rangle_{\text{Weyl}} \ . \tag{9.8}$$

If we follow the above prescription, the quantum constraints are given by $C_f = \langle \hat{f}\hat{C} \rangle$ — and *we restrict $\hat{f}$ to be a polynomial operator in the canonical variables*. Now we have to prove that this proposition indeed works.

First of all, it is consistent with $\hat{C}|\psi\rangle = 0$ and the set of operators of the form $\hat{f}\hat{C}$ is closed under taking commutators. As a result, the set of all functions C_f is first class. Note that $C_f^{(n)}$ is automatically included in the above constraints through $C_{f'}$ where $\hat{f}' = \hat{f}\hat{C}^{n-1}$, which is polynomial in the canonical variables as long as $\hat{f}$ is.

What remains to be shown is that the restriction to polynomial functions f is sufficient to reduce quantum phase space properly. To see how the degrees of freedom are reduced, we proceed order by order.

Variables of order M in $N+1$ canonical pairs are defined as in (9.8), with $\sum_i a_i + b + c = M$. The total number of different combinations of this form is the same as the number of ways the positive powers adding up to M can be distributed between $2(N+1)$ terms, that is $\binom{M+2(N+1)-1}{2(N+1)-1}$. So at order M, we have $\binom{M+2(N+1)-1}{2(N+1)-1}$ quantum variables. Out of the total set, those of the form $G^{a_1,a_2,\ldots,a_{2N};b,0}$, $b \neq 0$, and $G^{a_1,a_2,\ldots,a_{2N};0,c}$, $c \neq 0$, should be eliminated through constraints. The variables $G^{a_1,a_2,\ldots,a_{2N};0,0}$, however, shall be unrestricted.

The remaining part of this section is dedicated exactly to this proof.

First, it is convenient to make another change of variables. We note that in order to permute two non-commuting canonical operators in a product we need to add $i\hbar$ times a lower order product. Starting with a completely symmetrized product of order M and iterating the procedure we arrive at

[9] The linear combinations that would satisfy the above relations may be obtained by performing a linear canonical transformation on the operators. Such combinations are not unique, but this fact is not important for the purpose of counting the degrees of freedom.

a sum of unsymmetrized products of order M and lower. This justifies the choice of new variables,

$$F^{a_1,a_2,\dots a_{2N};b,c} := \left\langle (\hat{x}_1)^{a_1} \cdots (\hat{x}_{2N})^{a_{2N}} \hat{p}^b \hat{q}^c \right\rangle \tag{9.9}$$

It is easy to see that there is a one-to-one correspondence between variables (9.8), combined with the expectation values, and (9.9), but the precise mapping is tedious to derive. As it is not needed for the counting of degrees of freedom, we will not bother with it here.

Now our constraints require $F^{a_1,a_2,\dots a_{2N};b,c} \approx 0$ for $c \neq 0$ — because $\hat{q}|\psi\rangle = 0$. Moreover, all of the constraints $C_f = \langle \hat{f}\hat{C}\rangle$ may be written as a combination of the variables $F^{a_1,a_2,\dots a_{2N};b,c}$, $c \neq 0$. However, the constraints cannot restrict $F^{a_1,a_2,\dots a_{2N};b,0}$! But anyway, we have not taken into account the gauge flows yet.

Each C_f generates a flow on the constraint surface. But we decided to study the elimination of degrees of freedom order by order in the moments. At a *fixed* order, however, the flows generated by the constraints are not necessarily independent on the constraint hypersurface. This is due to the fact that quantum phase space reduced to a fixed order of moments is a non-symplectic Poisson manifold.[10]

This degeneracy becomes obvious when we count the degrees of freedom to a given order. To order M, the constraints are accounted for by variables $F^{a_1,a_2,\dots a_{2N};b,c+1}$, where $\sum_i a_i + b + c + 1 = M$. Counting as earlier in the section, there are $\binom{M+2(N+1)-2}{2(N+1)-1}$ such variables. Subtracting the number of constraints from the number of quantum variables of order M, we are left with

$$\begin{aligned} & \binom{M+2(N+1)-1}{2(N+1)-1} - \binom{M+2(N+1)-2}{2(N+1)-1} \\ = & \left(\frac{M+2(N+1)-1}{M+2(N+1)-1-(2N+1)} - 1 \right) \binom{M+2(N+1)-2}{2(N+1)-1} \\ = & \frac{2(N+1)-1}{M} \binom{M+2(N+1)-2}{2(N+1)-1} \end{aligned} \tag{9.10}$$

unrestricted quantum variables. If each constraint generates an independent non-vanishing flow, we should subtract the number of constraints from the result again. We find $\frac{2(N+1)-1-M}{M}\binom{M+2(N+1)-2}{2(N+1)-1}$ physical degrees of freedom at order M. However, this number becomes *negative* once M is large enough, raising the possibility that the system has been over-constrained.

[10] This can be read off from the Poisson bracket (9.3). The Poisson bracket of two quantum variables of order M, $a+b=M$, $c+d=M$, contains quantum variables of order $2M$ and lower.

Fortunately, we can show that our system is not overconstrained, i.e. the gauge flows are not all independent. This is demonstrated by the following argument.

All of the operators $\hat{x}_i$ commute with the original constraint operator $\hat{C} = \hat{q}$, which means that any function of the expectation value of a polynomial in $(\hat{x}_i)_{i=1,\dots,2N}$, here denoted by $g = \langle g[\hat{x}_i]\rangle$, weakly commutes with every constraint

$$\begin{aligned}\{C_f, \langle g[\hat{x}_i]\rangle\} &= \frac{1}{i\hbar}\left\langle\left[\hat{f}\hat{C}, g[\hat{x}_i]\right]\right\rangle = \frac{1}{\mathrm{i}\hbar}\left\langle \hat{f}\left[\hat{C}, g[\hat{x}_i]\right] + \left[\hat{f}, g[\hat{x}_i]\right]\hat{C}\right\rangle \\ &= \frac{1}{\mathrm{i}\hbar}\left\langle\left[\hat{f}, g[\hat{x}_i]\right]\hat{C}\right\rangle \approx 0\ . \end{aligned} \tag{9.11}$$

This means that *the variables $F^{a_1,a_2,\dots a_{2N};0,0}$ are not only unconstrained but also unaffected by the gauge flows.*

They can be used to construct the quantum variables corresponding to precisely N canonical pairs, so that we have *at least* the correct number of physical degrees of freedom.

But, still, we were not able to remove the variables $F^{a_1,a_2,\dots a_{2N};b,0}$, $b \neq 0$. They are unrestricted by the constraints. So they have to be affected by gauge transformations. In fact, we can show that $F^{a_1,a_2,\dots a_{2N};b,0}$, $b \neq 0$ are gauge dependent. This follows from a direct evaluation of

$$\begin{aligned}\{C_f, F^{a_1,a_2,\dots a_{2N};b,0}\} &= \frac{1}{\mathrm{i}\hbar}\left\langle\left[\hat{f}\hat{C}, (\hat{x}_1)^{a_1}\cdots(\hat{x}_{2N})^{a_{2N}}\hat{p}^b\right]\right\rangle \\ &= \frac{1}{\mathrm{i}\hbar}\left\langle\left[\hat{f}, (\hat{x}_1)^{a_1}\cdots(\hat{x}_{2N})^{a_{2N}}\hat{p}^b\right]\hat{C}\right. \\ &\quad \left. + \mathrm{i}\hbar b\hat{f}(\hat{x}_1)^{a_1}\cdots(\hat{x}_{2N})^{a_{2N}}\hat{p}^{b-1}\right\rangle \\ &\approx b\left\langle\hat{f}(\hat{x}_1)^{a_1}\cdots(\hat{x}_{2N})^{a_{2N}}\hat{p}^{b-1}\right\rangle\ . \end{aligned}$$

One may still suspect that a gauge may be selected such that the flows on one of these variables vanish. However, this is not the case. Substituting a constraint such that $\hat{f} = g[\hat{x}_i]\hat{C}^{b-1}$, where $g[x_i]$ is some polynomial in $2N$ variables

$$\left\{C_{gC^{b-1}}, F^{a_1,a_2,\dots a_{2N};b,0}\right\} \approx b\left\langle\ g[\hat{x}_i]\left((\hat{x}_1)^{a_1}\cdots(\hat{x}_{2N})^{a_{2N}}\right)\hat{C}^{b-1}\hat{p}^{b-1}\right\rangle$$

and commuting all the $\hat{C}$ to the right one by one, such that $\hat{C}^{b-1}\hat{p}^{b-1} = (b-1)!(i\hbar)^{b-1} + \cdots$ up to operators of the form $\hat{A}\hat{C}$, we have

$$\left\{C_{gC^{b-1}}, F^{a_1,a_2,\dots a_{2N};b,0}\right\} \approx b!(\mathrm{i}\hbar)^{b-1}\left\langle\ g[\hat{x}_i]\left((\hat{x}_1)^{a_1}\cdots(\hat{x}_{2N})^{a_{2N}}\right)\right\rangle\ . \tag{9.12}$$

Since the right-hand side is a gauge-independent function, (9.12) tells us that it is impossible to get rid of all flows on a given variable $F^{a_1,a_2,\ldots a_{2N};b,0}$ by simply picking a gauge.

In summary, using an alternative set of variables $F^{a_1,a_2,\ldots a_{2N};b,c}$ defined in (9.9) we find that constraints become $F^{a_1,a_2,\ldots a_{2N};b,c} \approx 0$, $c \neq 0$. The variables $F^{a_1,a_2,\ldots a_{2N};b,0}$, $b \neq 0$, are gauge dependent, which leaves the gauge-invariant and unconstrained physical variables $F^{a_1,a_2,\ldots a_{2N};0,0}$. These may then be used to determine the physical quantum variables $G^{a_1,\ldots,a_{2N};0,0}$ defined in (9.8).

Thus, for a linear constraint a correct reduction of the degrees of freedom is achieved by applying constraints of the form $C_f = \langle \hat{f}\hat{C}\rangle$ where f is polynomial in the canonical variables, as can be directly observed order by order in the quantum variables. Locally, our procedure of effective constraints is complete and consistent since any irreducible constraint can locally be chosen as a canonical coordinate. Still, global issues may pose non-trivialities since entire gauge orbits must be factored out when constraints are solved.

Example: A canonical variable as constraint: $\hat{C} = \hat{q}$

Given that the precise implementation of a set of quantum constraints depends on the form of the constrained system, we illustrate typical properties by an example. Further examples can be found in [80] which, for the sake of brevity, are not included here.

The method of solving quantum constrained systems may be outlined as follows: We start by finding the complete first-class set of constraint functions representing Dirac's constraint condition. Setting these to zero defines the constraint surface, with constraint functions generating gauge transformations on it. We construct observables from the *gauge-invariant* functions and recover dynamics, where appropriate, as a gauge transformation of non-observable quantities.

To show how the counting argument of the previous section actually works in practice, we will consider the constraint $\hat{C} = \hat{q}$.

From $C^{(n)} = 0$ we obtain that all quantum variables G^{q^n} are constrained to vanish, in addition to $C_Q = q$ itself. $C_q^{(n)}$ is included as $C^{(n+1)}$. To eliminate moments depending not only on $\hat{q}$ but also on $\hat{p}$, we have to add $C_{p^m}^{(n)} = \langle \hat{p}^m \hat{q}^n\rangle$. If we restrict attention to moments up to second order, it suffices to add $C_p^{(n)} = \langle \hat{p}\hat{q}^n\rangle$. This already produces a closed set of constraints.

In this example, it is feasible to work with the symmetrically ordered quantum variables introduced in Section 9.4.3. Because here we know a quantum variable that is gauge invariant, i.e. commutes with the constraint, namely $\hat{q}$. For instance, quantum variables $G^{C^n q}$ and $G^{C^n p}$ form a closed

set of constraints as may be deduced from (9.8) and the subsequent discussion. Using the Poisson relations (9.3), we verify the first-class nature of the system of constraints

$$\begin{aligned}
\{G^{a,0}, G^{c,0}\} &= 0 && \text{for} \quad b=d=0\ ,\\
\{G^{a,0}, G^{c,1}\} &= a(G^{a+c-1,0} - G^{a-1,0}G^{c,0}) \approx 0 && \text{for} \quad b=0, d=1\ ,\\
\{G^{a,1}, G^{c,1}\} &= (a-c)G^{a+c-1,1} - aG^{a-1,1}G^{c,0} + cG^{a,0}G^{c-1,1} \approx 0 \\
& && \text{for} \quad b=d=1\ .
\end{aligned}$$

To discuss moments up to second order, constraints with at most a single power of p are needed. These constraints are in fact equivalent to constraints given by quantum variables due to

$$\begin{aligned}
G^{q^n} &= \langle(\hat{q}-q)^n\rangle = \sum_{j=0}^{n}\binom{n}{j}(-1)^j q^j\langle\hat{q}^{n-j}\rangle = \sum_{j=0}^{n-1}\binom{n}{j}(-1)^j q^j C^{(n-j)} + (-1)^n q^n\ ,\\
G^{q^n p} &= \frac{1}{n+1}\langle(\hat{q}-q)^n(\hat{p}-p) + (\hat{q}-q)^{n-1}(\hat{p}-p)(\hat{q}-q) + \cdots + (\hat{p}-p)(\hat{q}-q)^n\rangle\\
&= \frac{1}{n+1}\langle(n+1)(\hat{p}-p)(\hat{q}-q)^n + \tfrac{1}{2}\mathrm{i}n(n+1)\hbar(\hat{q}-q)^{n-1}\rangle\\
&= \langle\hat{p}\hat{q}^n\rangle - p\langle(\hat{q}-q)^n\rangle + \sum_{j=1}^{n}\binom{n}{j}(-1)^j q^j\langle\hat{p}\hat{q}^{n-j}\rangle + \frac{1}{2}\mathrm{i}n\hbar\langle(\hat{q}-q)^{n-1}\rangle\\
&= C_p^{(n)} - pG^{q^n} + \sum_{j=1}^{n-1}\binom{n}{j}(-1)^j q^j C_p^{(n-j)} + (-1)^n q^n p + \frac{1}{2}\mathrm{i}\hbar n G^{q^{n-1}}\ .
\end{aligned}$$

Starting from $n=1$, one can iteratively verify that the relations above provide a one-to-one mapping from

$$\left(C^{(n)}, C_p^{(m-1)}\right)_{n,m\in\mathbb{N}} \quad \longrightarrow \quad \left(G^{q^n}, G^{q^m p}\right)_{n,m\in\mathbb{N}}\ .$$

This is thus a specific demonstration of the relation between (9.8) and (9.9) discussed in Section 9.4.4.

Thus, the constraint surface can be analyzed using quantum variables. What about the gauge flow?

For this type of classical constraint, reordering will only lead to either a constant or to terms depending on quantum variables defined without reference to $\hat{p}$. Since these are already included in the set of constraints and a constant does not alter the generated canonical transformations, they can be eliminated when computing the gauge flow.[11]

The gauge flow generated by the quantum constraints up to second order can therefore be computed using symmetrically-ordered quantum variables,

[11] The constant term $\frac{1}{2}\mathrm{i}\hbar$ in $G^{q^n p}$ for $n=1$, however, will play an important role in determining the constraint surface.

such as G^{q^n} and $G^{q^n p}$ for example. For the moments of different orders, we then have the following constraints and gauge transformations.

Expectation values

We just have one constraint, $q = 0$, which generates the gauge transformation $p \mapsto p + \epsilon_1$.

Fluctuations

Here, two constraints arise, $G^{qp} \approx$ const. and $G^{qq} \approx 0$. They generate gauge transformations $G^{pp} \mapsto G^{pp} + 4\epsilon_2 G^{qp}$ and $G^{pp} \mapsto G^{pp}(1 + 2\epsilon_3)$, respectively. As we will see in (9.13) below, G^{qp} is non-zero on the constraint surface, so that G^{pp} can be freely rescaled using gauge transformations.

Higher Moments

At each order, we have constraints $C_{p^m}^{(n-m)}$ with $m < n$. The only quantum variable that is not eliminated through these is G^{p^n}. This has to be removed by gauge transformations, generated for example by G^{q^n}. This confirms the counting of Section 9.4.4.

Thus, to second order, we see that two moments are eliminated by quantum constraints while the remaining one is gauge. In this way, the quantum variables are eliminated completely.

Let us now discuss the reality conditions of quantum variables. The constraint $C_p^{(1)} = \langle \hat{p}\hat{q} \rangle = 0$ implies that

$$G^{qp} = \frac{1}{2}\langle \hat{q}\hat{p} + \hat{p}\hat{q} \rangle - qp = \langle \hat{p}\hat{q} \rangle - qp + \frac{1}{2} i\hbar \approx \frac{1}{2}\mathrm{i}\hbar \tag{9.13}$$

must be imaginary. From the point of view of the kinematical inner product this seems problematic since we are taking the expectation value of a symmetrically ordered product of self-adjoint operators. However, the inner product of the kinematical Hilbert space is only auxiliary, and from our perspective not even necessary to specify. But an imaginary value of some kinematical quantum variable is not only admissible but also has a big advantage. Only with such an imaginary value can the uncertainty relations be satisfied.

For an *unconstrained* system, we have the generalized uncertainty relation

$$G^{qq} G^{pp} - (G^{qp})^2 \geq \frac{\hbar^2}{4}\,. \tag{9.14}$$

So this relation *has* to hold on the physical Hilbert space. More than that, we want it to hold on the kinematical level already. We do so because the

uncertainty relations are useful to construct coherent states and it is often helpful to have them at ones disposal.

If we had worked with *real* quantum constraints, $G^{qq} \approx 0 \approx G^{qp}$ instead of $C^{(2)} \approx 0$, $C_p^{(1)} \approx 0$, the uncertainty relation (9.14) would have been violated. The imaginary value of G^{qp} obtained with our definition of the quantum constraints, on the other hand, allows us to implement the constraints in a way *respecting the standard uncertainty relation*. Choosing $-(G^{qp})^2 = \frac{1}{4}\hbar^2$ from (9.13) just saturates relation (9.14).

So, linear constraints show that quantum constraints can be formulated in a consistent and complete way. This then also entails *local* consistency and completeness of every system with a single constraint since every constraint can be linearized by a canonical transformation. But nonetheless, global issues may be important, especially in quantum theory. Moreover, such a canonical transformation linearizing a constraint induces complicated transformations on quantum variables in which their orders mix. Such a linearization, even though in principle possible, might thus not be a very handy option. We will therefore have to discuss non-linear examples to show the practicality of our procedures. Note that already in the above, simple example, we restricted attention to second-order moments.

9.5 Approximation schemes

The aim of this section will be to devise a general method to consistently approximate quantum constraints. The approximated constraints will be referred to as *effective constraints*. Such a reduction of the full system is in general necessary for practical purposes. The infinite-dimensional system has to be reduced to a certain finite order of quantum variables so that one can actually retrieve some information from it.

The most immediate approximation scheme that may come to mind, is a sharp *truncation*. That means that we simply set all quantum variables above a certain order to zero. It is then necessary to check whether the system of constraints can still be formulated in a consistent way after such a reduction has been carried out. A priori one cannot assume, for instance, that a sharply truncated system of constraints has any non-trivial solution at all. It may turn out that all degrees of freedom are removed by the truncated constraints. Also, it is not clear how many (truncated) constraints have to be taken into account at a certain order of the truncation. In this section, we first consider a linear example and show that it can be consistently truncated. We then turn to the more elaborate and more physical example of the parametrized free, non-relativistic particle. Here, *sharp truncations turn out to be inconsistent*.

This motivated a more careful approximation scheme which removes

degrees of freedom order by order — as is already suggested by the proof in Section 9.4.4.

We will first show that sharp truncations are unreliable as a general tool, before turning to consistent approximations.

9.5.1 Truncations

We will discuss truncations for the simple constraint $C = q$ of Section 9.4.4 and then turn to the relativistically more relevant case of the parametrized, free particle.

Truncated system of constraints for $\hat{C} = \hat{q}$

The system of Section 9.4.4 is governed by a constraint $C = q$, which, on the quantum level, entails the constraint operator $\hat{C} = \hat{q}$. This implies the following constraints on quantum phase space:

$$\begin{aligned}
C^{(n)} &= \langle \hat{C}^n \rangle = C^n + \sum_{j=0}^{n-1} \binom{n-1}{j} C^j G_{0,n-j} \ , \\
C_q^{(n)} &= \langle \hat{q}\hat{C}^n \rangle = C^{(n+1)} \ , \\
C_p^{(n)} &= \langle \hat{p}\hat{C}^n \rangle = pC^n + p\sum_{j=0}^{n-1} \binom{n}{j} C^j G_{0,n-j} \\
&\quad + \sum_{j=0}^{n-1} \binom{n}{j} \frac{C^j}{a_{n-j}} \left(G_{1,n+1} - \mathrm{i}\hbar \frac{(n-j)^2}{(n-j+1)} G_{0,n-j-1} \right) \ ,
\end{aligned}$$

where a_{n-j} are constant coefficients. These are accompanied by similar expressions of higher-polynomial constraints, i.e. $C_{p^m}^{(n)}$, which are more lengthy in explicit form due to the reordering involved in quantum variables.

The lowest power constraint yields $C^{(1)} = C \approx 0$. Inserting this, the higher power constraints reduce to

$$\begin{aligned}
C^{(n)} &\approx G_{0,n} \quad , \quad C_q^{(n)} \approx G_{0,n+1} \ , \\
C_p^{(n)} &\approx pG_{0,n} + \frac{1}{a_n} \left(G_{1,n} - i\hbar \frac{n^2}{(n+1)^2} G_{0,n-1} \right) \ .
\end{aligned}$$

Performing a sharp truncation at N^{th} order, we set $G_{a,b} = 0$ for all $a+b > N$. As non-trivial constraints remain

$$\begin{aligned}
C^{(n)}|_N &\approx G_{0,n} && \text{for all} \quad n \leq N \ , \\
C^{(n)}_p|_N &\approx pG_{0,n} + \frac{1}{a_n}\left(G_{1,n} - \mathrm{i}\hbar\frac{n^2}{(n+1)^2}G_{0,n-1}\right) && \text{for all} \quad n \leq N-1 \ , \\
C^{(N)}_p|_N &\approx pG_{0,N} + \frac{1}{a_N}\left(-i\hbar\frac{N^2}{(N+1)^2}G_{0,N-1}\right) && \text{for} \quad n = N \ .
\end{aligned}$$

Solving the quantum constraints $C^{(n)} \approx 0$ and inserting the solutions into the constraints $C^{(n)}_p$, yields

$$\begin{aligned}
C^{(n)}_p|_N &\approx \frac{1}{a_n}G_{1,n} && \text{for all} \quad n \leq N-1 \ , \\
C^{(n)}_p|_N &\approx 0 && \text{for all} \quad n \geq N.
\end{aligned}$$

Thus we find that for the truncated system, $G_{0,n}$ are eliminated through the constraints $C^{(n)} = 0$, whereas the quantum variables $G_{1,n}$ are eliminated through $C^{(n)}_p = 0$. Higher polynomial constraints can be expanded as

$$\begin{aligned}
C^{(n)}_{p^k} &= \sum_{i=0}^{k}\sum_{j=0}^{n}\binom{k}{i}\binom{n}{j}p^i C^j \langle(\hat{p}-p)^{k-i}(\hat{q}-q)^{n-j}\rangle \\
&\approx \sum_{i=0}^{k}\binom{k}{i}p^i\langle(\hat{p}-p)^{k-i}(\hat{q}-q)^n\rangle = \frac{G_{k,n}}{b_{k,n}} + \cdots
\end{aligned}$$

with some coefficients $b_{k,n}$. Moments of lower order in p are not written explicitly because they can be determined from constraints of smaller k.

So we see that these constraints fix all remaining moments *except* $G_{n,0}$. Moreover, due to the constraint $C^{(1)} = C \approx 0$, expectation values are restricted to the classical constraint hypersurface. No further restrictions on these degrees of freedom arise.

The remaining *unconstrained* $G_{n,0}$ are, on the other hand, pure gauge. They can be changed arbitrarily by a gauge transformation. This again confirms considerations of Section 9.4.4 because the gauge flow of $C^{(n)}_{q^m} = C^{(n+m)}$ is sufficient to remove all gauge without making use of $C^{(n)}_{p^m}$ with $m \neq 0$, where operators not commuting with the constraint would occur.

We therefore conclude that the system can thus be truncated consistently. For a truncation at N^{th} order of a linear classical constraint, constraints up to order N have to be taken into account.

However, the linear case is quite special because we only had to truncate the *system* of constraints, but not *individual* constraints: any effective constraint contains quantum variables of only one, fixed order. This was

just the point of Section 9.4.4. When $\hat{C}$ is linear, we can impose all of the constraints and remove all gauge degrees of freedom in variables up to a given order without invoking higher-order constraints. This is accomplished by treating higher-order constraints as imposing conditions on higher-order quantum variables — possibly in terms of the lower-order unconstrained ones — and noting that, using (9.12), there is no need to refer to constraints containing polynomial terms of order above $F^{a_1,a_2,\ldots,a_{2N};b,0}$ itself, in order to demonstrate that this variable may be rescaled using gauge transformations.

Moreover, the gauge-invariant degrees of freedom that remain, weakly commute with *all* constraints and not just with the constraints up to the order considered, see (9.11). As a result, in the *linear* example of Section 9.4.4, higher-order constraints do not affect the reduction of degrees of freedom for lower orders and so could be disregarded without making any approximations. For *non-linear* constraints, however, orders of moments mix and constraints relevant at low orders can contain moments of *higher* order. It is then more crucial to see how the higher moments could be disregarded consistently, as we will do in what follows.

Truncated system of constraints for the parametrized free, non-relativistic particle

The motion of a free particle of mass M in one dimension is described on phase space with coordinates (p, q). Through the introduction of an arbitrary time parameter τ, time can be turned into an additional degree of freedom. The system is then formulated on the 4-dimensional phase space with coordinates $(t, p_t; q, p)$. The Hamiltonian constraint of the parametrized free, non-relativistic particle is given by

$$C = p_t + \frac{p^2}{2M} \ ,$$

which is constrained to vanish. Because this is the scenario encountered in all reparametrization-invariant theories, this example is of special interest for diffeomorphism-invariant theories as general relativity.

Promoting phase space variables to operators, Dirac constraint quantization yields the quantum constraint

$$\left(\hat{p}_t + \frac{\hat{p}^2}{2M}\right)\Psi = 0 \ . \tag{9.15}$$

In the Schrödinger representation, one arrives at an equation that is formally equivalent to the time-dependent Schrödinger equation[12]

[12] In contrast to the ordinary, time-dependent Schrödinger equation, time is an *operator* in the equation obtained here and not an external parameter. This implies that the Hamiltonian which generates evolution in time, $\hat{\mathcal{H}}_{\rm phys} = \frac{\hat{p}^2}{2M}$, has the same action on physical states as the momentum operator canonically conjugate to time. In contrast to

$$\mathrm{i}\hbar\frac{\partial\Psi(t,q)}{\partial t} = -\frac{\hbar^2}{2M}\frac{\partial^2\Psi(t,q)}{\partial q^2} \ .$$

As is well known, solutions to this equation are given by

$$\Psi(t,q) = \int \mathrm{d}k A(k) e^{\frac{\mathrm{i}}{\hbar}E(k)t + \mathrm{i}kq} \ , \tag{9.16}$$

where $E(k) = \frac{\hbar^2 k^2}{2M}$.

For quantum variables we use, as before, the notation

$$G^{a,b}_{c,d} = \langle(\hat{p}-p)^a(\hat{q}-q)^b(\hat{p}_t-p_t)^c(\hat{t}-t)^d\rangle_{\mathrm{Weyl}} \ .$$

In their general form, the set of constraints on the quantum phase space is given in Appendix E.

Zeroth-order truncation Truncation of the system at zeroth order, i.e. setting all quantum variables to zero, yields $C^{(n)}|_{N=0} = C^n$ together with

$$C^{(n)}_q|_{N=0} = qC^n + \frac{\mathrm{i}\hbar}{2}n\frac{p}{m}C^{n-1} \quad , \quad C^{(n)}_t|_{N=0} = tC^n + \frac{\mathrm{i}\hbar}{2}nC^{n-1}$$

as the required constraints. This truncation is *not* consistent. Inserting the condition $C = 0$ into the expressions for the remaining constraints results in inconsistent equations. For example $C^{(1)}_t|_{N=0} = tC + \frac{1}{2}\mathrm{i}\hbar$, implies $\frac{\mathrm{i}\hbar}{2} = 0$.

The reason for the failure of the sharp truncation seems clear. A truncation at zeroth order can be understood as neglecting all quantum properties of the system. But this is not possible for a free particle. For example, there is no solution in which the spread in both p and q would stay negligible throughout the particle's evolution. There is no wave-packet which would remain tightly peaked throughout and so a description in terms of expectation values alone seems insufficient in this case.

Second-order truncation But even if one takes into account second-order quantum variables, spreads and correlations, an inconsistent system results. The *expanded* constraints can also be found in Appendix E, which we now sharply truncate at second-order moments.

From $C^{(n)}$ only three non-trivial constraints follow

the physical Hamiltonian, which is bounded below and positive semi-definite, the spectrum of the time momentum $\hat{p}_t$ covers the entire real line. On physical solutions, however, only positive 'frequencies' contribute.

$$\begin{aligned} C^{(1)} &= C + \frac{1}{2M} G^{2,0}_{0,0} \,, \\ C^{(2)}|_{N=2} &= C^2 - (6C - 4p_t)\, G^{2,0}_{0,0} + \frac{4p}{2M} G^{1,0}_{1,0} + G^{0,0}_{2,0} \,, \\ C^{(3)}|_{N=2} &= C^3 \,, \end{aligned}$$

upon inserting the constraints successively. Thus for an $N = 2$ order truncation, at $n = 3$, the classical constraint is recovered and must vanish for the truncated system. Then, $C^{(1)} \approx 0$ yields $G^{2,0}_{0,0} \approx 0$ which is too strong for a consistent reduction since one expects the fluctuation G^{pp} to be freely specifiable. It has to remain a physical degree of freedom after solving the constraints, for otherwise no general wave packet as in (9.16) can be posed as an initial condition for the free particle. As we see, *the sharply truncated system is over-constrained*. In particular, the constraint $C^{(3)}$, when truncated to second-order moments, reduces to the classical constraint C^3, which then immediately implies $G^{pp} = 0$ due to $C^{(1)}$.

This observation points already — and luckily — to a resolution of the inconsistency. While $C^{(1)}$ is already of second order even without a truncation, $C^{(3)}$ contains higher-order moments. The truncation is then inconsistent because we are ignoring higher orders next to an expression which we then constrain to be zero. Thus, a more careful approximation scheme must be devised where we do not truncate sharply but ignore higher moments *only* when they appear together with lower moments *not constrained to vanish*. In such a scheme, as discussed in the following section, $C^{(3)}$ would pose a constraint on the higher moments in terms of $C \approx -G^{pp}/2M$, but would not require C or G^{pp} to vanish.

9.5.2 Consistent approximations

We saw in the preceding section that a sharp truncation scheme in which all moments larger than a certain order are set to zero, does not yield sensible results. We have to devise a more careful approximation scheme that takes into account, e.g. products of moments, but also orders of $\hbar$. Most importantly, we want to avoid the situation in which terms are considered as dominant over others and are consequently set to zero, which themselves vanish due to some quantum constraint.

General procedure and moment expansion

The consistent approximation is based on a *moment expansion*. To formalize this *moment expansion*, one replaces each moment

$$G^{a,b}_{c,d} \quad \xrightarrow{\text{by}} \quad \lambda^{a+b+c+d} G^{a,b}_{c,d}$$

and expands in λ. This automatically guarantees that higher-order moments appear at higher orders in the expansion, and that products of moments are of higher order than the moments themselves. Moreover, in order to leave the uncertainty relation unchanged, we have to replace

$$\hbar \qquad \stackrel{\text{by}}{\longrightarrow} \qquad \lambda^2\hbar \ ,$$

which ensures that it is of higher order, too, without performing a specific $\hbar$-expansion. After the λ-expansion has been performed, λ can be set equal to one to reproduce the original terms.

In addition to such an expansion scheme, it is important to establish a certain hierarchy of constraints that determines in which order the constraints have to be considered, i.e. which are to serve as constraints on expectation values, which restrict second-order moments and which have to be considered as constraints on higher-order moments.

This scheme will be demonstrated for the parametrized particle of the previous section, but the general considerations apply to any parametrized, non-relativistic, i.e. *artificially* reparametrization-invariant, system.

Hierarchy of constraints

Variables and constraints were determined in Section 9.5.1. The form of constraints establishes a hierarchy, suggesting to solve $C^{(n)}$ first, then $C_q^{(n)}$, $C_t^{(n)}$, $C_{p_t}^{(n)}$ and $C_p^{(n)}$, and the remaining constraints (E.3) – (E.6) first for $k = 1$, then $k = 2$ etc. Note that for each k in (E.3) – (E.6) the $r = k$ term is the only contribution of a form not appearing at lower orders. The terms occurring in the r-sum are linear combinations of the constraints (E.3) – (E.6) for $k' < k$. Thus, apart from the $r = k$ term, all terms vanish if the lower k constraints are satisfied.

To actually see that this is a sensible way of sorting the constraints, one has to look a bit more into the structure of the constraints.

First notice that the structure of the constraints is such that on the constraint hypersurface $C^{(n)}$, $C_{qp^k}^{(n)}$, $C_q^{(n)}$, $C_{tp^k}^{(n)}$ and $C_t^{(n)}$ contain as lowest order terms expectation values, whereas $C_{pp^k}^{(n)}$, $C_p^{(n)}$, $C_{p_tp^k}^{(n)}$ and $C_{p_t}^{(n)}$ have second-order moments as lowest contribution. The highest order moments occurring in $C^{(n)}$ are of order $2n$, of order $2n+1$ for $C_q^{(n)}$, $C_t^{(n)}$, $C_p^{(n)}$ and $C_{p_t}^{(n)}$ and of order $2n+1+k$ in $C_{qp^k}^{(n)}$, $C_{tp^k}^{(n)}$, $C_{pp^k}^{(n)}$ and $C_{p_tp^k}^{(n)}$.

The structure of (E.3) – (E.6) implies that the lowest contributing order in the j- and ℓ-sums (on the constraint hypersurface) is $j+\ell+k\pm1$ and rises with k. Consequently, there exists a maximal k up to which constraints have to be studied if only moments up to a certain order are taken into account.

This can be seen by studying the constraint

$$C^{(n)} = \sum_{m=0}^{n} \sum_{j=0}^{m} \sum_{\ell=0}^{2(n-m)} \binom{n}{m} \binom{m}{j} \binom{2(n-m)}{\ell} p_t^{m-j} \frac{p^{2(n-m)-\ell}}{(2M)^{n-m}} G^{\ell,0}_{j,0} \ .$$

This specific constraint shall only serve as a representative. The other constraints, (E.1) – (E.6), are of the same structure. Assume that we are interested in moments up to order N. For such a fixed order N of moments, there is a factor of lowest and one of highest power of C. In $C^{(n)}$, e.g., the highest power is given for $j = 0$, $\ell = 0$ (with $m = n$) and is simply C^n, whereas the lowest power is given for $\ell = 0$, $j = N$ and is given by $n(n-1)\cdots(n-(N-1))C^{n-N}$.[13]

In the j-ℓ-summation, the relevant moments then occur for $j + \ell \pm 1 \leq N$. From this condition, a number of pairs (j, ℓ) result for which the sums occurring in (9.5.2), and thus also in (E.1) – (E.6), can be evaluated.

There remain sums over m containing p_t, which should be eliminated if we choose t as internal time to make contact with the quantum theory of the deparametrized system. (Our consistent approximation procedure, however, is more general and does not require the choice of an internal time.) We can achieve this by rewriting these as terms of the form $n(n-1)\cdots(n-g)C^{n-g-1}$ multiplied by powers of p and $2M$, where g is an integer depending on the values of j and ℓ, see (E.7) – (E.21) as examples. This is achieved by eliminating p_t via $C^{(1)} = C_Q \approx 0$ and illustrates the central role played by the principal quantum constraint C_Q.[14]

Since $C \approx -G^{2,0}_{0,0}/2M$, powers of second-order moments ensue — or higher q-moments if there is a potential. Together with powers of $\hbar$ in some of the terms, this must be compared with the orders of *higher* moments in order to approximate consistently.

One can now rewrite the sum over m for all those terms which produce factors with powers of C down to the lowest power occurring in front of the relevant moments. In $C^{(n)}$ this would correspond to C^{n-N}. One can therefore rewrite the constraints in the form

$$C^n Y_1 + nC^{n-1}Y_2 + n(n-1)C^{n-2}Y_3 + \cdots + R \approx 0 \ , \qquad (9.17)$$

where Y_i are functions linear in moments including those of order *smaller* than N, and R contains only moments which are of *higher* order. This allows one to successively solve the constraints for $n = 1$, $n = 2$, etc. and discard all constraints arising for $n \geq N+1$, $n > 0$. In each case, one has to find the terms of lowest order in the moment expansion, in combination with powers C^n, to see at which order a constraint becomes relevant.

[13] This term arises of course as well for $(\ell = N, j = 0)$, $(\ell = 1, j = N-1)$, etc.

[14] In our example of the free particle, we have $C_Q = p_t + p^2/2M + G^{2,0}_{0,0}/2M$. If there is a potential, there will be further classical terms as well as quantum variables $G^{0,n}_{0,0}$.

It is crucial for this procedure to work that C^n, which arises in all constraints, can be eliminated at least for all $n > n'$ through terms of higher-order moments using the principal constraint C_Q. This key property is indeed realized for *any* non-relativistic particle even in a potential, as long as p_t appears linearly. For relativistic particles, additional subtleties arise that have not been studied yet. While (E.1) – (E.6) change their form for a different classical constraint, the procedure sketched here still applies. Thus, it does not only refer to quadratic constraints but is sufficiently general for non-relativistic quantum mechanics.

We will explicitly demonstrate the procedure for the free particle in what follows. For that purpose, we rewrote the set of constraints in the required form (9.17) for moments up to third order as seen in Appendix E.

9.5.3 Consistent approximations: the parametrized, free non-relativistic particle as an example

Consistency of constraints for expectation values

At zeroth order, we keep only expectation values. All moments are of order $\mathcal{O}(\lambda^2)$ or higher. As only relevant constraints we therefore find $C^{(n)} \approx 0$, cf. Appendix E. Keeping only zeroth-order terms, this reduces to $C^{(n)} = C^n \approx 0$. This in turn corresponds to the single constraint $C \approx 0$ which can be used to eliminate p_t in terms of p. The system of constraints is obviously consistent at zeroth order and no constraints on variables associated with the pair (q, p) result.

The only constraint $C^{(1)} = C \approx 0$ generates a gauge flow on expectation values given by

$$\dot{p} = 0 \ , \qquad \dot{p}_t = 0 \ , \qquad \dot{q} = \frac{p}{M} \ , \qquad \dot{t} = 1 \ .$$

where a dot denotes derivation with respect to parameter time τ. The two observables of the system are therefore

$$\mathcal{P}^{(0)} = p \quad \text{and} \quad \mathcal{Q}^{(0)} = q - t\frac{p}{M} \quad \text{with} \quad \{\mathcal{Q}^{(0)}, \mathcal{P}^{(0)}\} = 1 \ , \tag{9.18}$$

These correspond to the two physical degrees of freedom corresponding to expectation values of canonical variables. Among the four original degrees of freedom of the system, p_t is eliminated via the constraint and t is a pure gauge degree of freedom. There are no further constraints to this order, which is thus consistently approximated.

Consistency of constraints up to second-order moments

At second order, we include second-order moments and orders of $\hbar$ — recall that $\hbar$ is of order λ^2 in the moment expansion — in addition to expectation

values. Third order contributions are set to zero. We find that in addition to $C^{(1)}$, the new constraints $C_q^{(1)}$, $C_t^{(1)}$, $C_{p_t}^{(1)}$ and $C_p^{(1)}$ arise. All other constraints are of higher order, since second-order moments enter in these equations only through quadratic terms or with a factor of $\hbar$, both of which are considered as higher-order terms, cf. Appendix E. The only non-trivial constraints are therefore

$$\begin{aligned}
C^{(1)} &= C + \frac{1}{2M} G^{2,0}_{0,0} \approx 0 \,, \\
C_q^{(1)} &= G^{0,1}_{1,0} + \frac{p}{M}\frac{\mathrm{i}\hbar}{2} + \frac{p}{M} G^{1,1}_{0,0} \approx 0 \\
C_t^{(1)} &= \frac{p}{M} G^{1,0}_{0,1} + G^{0,0}_{1,1} + \frac{\mathrm{i}\hbar}{2} \approx 0 \\
C_{p_t}^{(1)} &= G^{0,0}_{2,0} + \frac{p}{M} G^{1,0}_{1,0} \approx 0 \\
C_p^{(1)} &= G^{1,0}_{1,0} + \frac{p}{M} G^{2,0}_{0,0} \approx 0 \,,
\end{aligned}$$

where third-order contributions have been set to zero. In accordance with the order of expectation values, we use the first constraint to eliminate $p_t = -p^2/2M - G^{2,0}_{0,0}/2M$ and solve for second-order moments

$$\begin{aligned}
G^{0,1}_{1,0} &= -\frac{p}{M}\frac{\mathrm{i}\hbar}{2} - \frac{p}{M} G^{1,1}_{0,0} \quad , \qquad \frac{p}{M} G^{1,0}_{0,1} = -G^{0,0}_{1,1} - \frac{\mathrm{i}\hbar}{2} \\
G^{0,0}_{2,0} &= -\frac{p}{M} G^{1,0}_{1,0} \quad , \qquad\qquad G^{1,0}_{1,0} = -\frac{p}{M} G^{2,0}_{0,0} \,.
\end{aligned} \tag{9.19}$$

As constraints for $k > 1$ contain second-order moments only through C^n, they are trivial as well. This follows from the first constraint which sets $C^n \sim \left(G^{2,0}_{0,0}\right)^n \sim \mathcal{O}(\lambda^{2n})$.

Thus, as far as second-order moments are concerned, the system of constraints is consistent. The moments $G^{0,0}_{2,0}$, $G^{1,0}_{1,0}$, $G^{1,0}_{0,1}$ and $G^{0,1}_{1,0}$ are fully determined while all second-order moments associated with the pair (q,p) can be specified freely. All remaining constraints then determine *higher* moments. This is the same situation as experienced in the linear case, as far as solving the constraints for second-order moments is concerned. The inconsistency of Section 9.5.1 is avoided because $C^{(3)}$, which made C and thus $G^{2,0}_{0,0}$ vanish in the sharp truncation, is now realized as a *higher-order constraint* in the moment expansion.

Gauge transformations are generated by $C^{(1)}$, $C_q^{(1)}$, $C_t^{(1)}$, $C_{p_t}^{(1)}$ and $C_p^{(1)}$, where third-order contributions are set to zero as in (9.19). In comparison to Section 9.5.3 we have four additional gauge transformations. Whereas $\mathcal{P}^{(2)} := \mathcal{P}^{(0)}$ remains gauge invariant under these transformations as well, this is not the case for $\mathcal{Q}^{(0)}$. The latter has to be altered by adding second-order moments such that an observable

$$\mathcal{Q}^{(2)} = \mathcal{Q}^{(0)} - \frac{1}{M} G^{1,0}_{0,1} \tag{9.20}$$

results which also satisfies $\{\mathcal{Q}^{(2)}, \mathcal{P}^{(2)}\} = 1$.

Calculating the transformations generated by constraints on *second-order* moments shows that $\mathcal{G}^{pp(2)} = G^{2,0}_{0,0}$ is an observable, i.e. commutes with all five constraints on the hypersurface defined by these constraints. The form of gauge orbits suggests to make the ansatz

$$\begin{aligned}
\mathcal{G}^{qp(2)} &= G^{1,1}_{0,0} + G^{0,0}_{1,1} - \frac{t}{M} G^{2,0}_{0,0} + \frac{\mathrm{i}\hbar}{2} \,, \\
\mathcal{G}^{qq(2)} &= G^{0,2}_{0,0} - 2\frac{p}{M} G^{0,1}_{0,1} + \frac{p^2}{M^2} G^{0,0}_{0,2} - \frac{2t}{M}\left(G^{1,1}_{0,0} + G^{0,0}_{1,1} + \frac{\mathrm{i}\hbar}{2}\right) \\
&\quad + \frac{t^2}{M^2} G^{2,0}_{0,0} \,,
\end{aligned} \tag{9.21}$$

for the remaining two observables. They are invariant under gauge transformations. The term $\frac{\mathrm{i}\hbar}{2}$ is included such that Poisson brackets between $\mathcal{G}^{qq(2)}$ and the remaining two quantum observables are of the required form. They satisfy

$$\begin{aligned}
&\{\mathcal{G}^{pp(2)}, \mathcal{G}^{qp(2)}\} = -2\mathcal{G}^{pp(2)} \quad , \quad \{\mathcal{G}^{pp(2)}, \mathcal{G}^{qq(2)}\} = -4\mathcal{G}^{qp(2)} \,, \\
&\{\mathcal{G}^{qp(2)}, \mathcal{G}^{qq(2)}\} = -2\mathcal{G}^{qq(2)} \,.
\end{aligned}$$

Commutators between $\mathcal{Q}^{(2)}$, $\mathcal{P}^{(2)}$ and the physical quantum variables $\mathcal{G}^{qq(2)}$, $\mathcal{G}^{pp(2)}$ and $\mathcal{G}^{qp(2)}$ vanish.

Thus we showed that four of the ten second-order moments are eliminated directly by the constraints. Three of the remaining second-order moments, $G^{0,0}_{1,1}$, $G^{0,0}_{0,2}$ and $G^{0,1}_{0,1}$, are pure gauge degrees of freedom. Consequently three physical quantum degrees of freedom remain at second order. The observables can be used to determine the general motion of the system in coordinate time: From (9.18) and (9.20) together with (9.19) and (9.21) we obtain

$$\begin{aligned}
q(t) &= \mathcal{Q}^{(2)} + \frac{t}{M}\mathcal{P}^{(2)} + \frac{1}{M} G^p_t \approx \mathcal{Q}^{(2)} + \frac{t}{M}\mathcal{P}^{(2)} - \frac{1}{p}\left(G_{tp_t} + \frac{\mathrm{i}\hbar}{2}\right) \\
&= \mathcal{Q}^{(2)} + \frac{t}{M}\mathcal{P}^{(2)} - \frac{1}{\mathcal{P}^{(2)}}\left(\mathcal{G}^{qp(2)} + \frac{t}{M}\mathcal{G}^{pp(2)} - G^{qp}\right)
\end{aligned} \tag{9.22}$$

for the relational dependence between q, t and G^{qp}. Thus, the moments appear in the solutions for expectation values in coordinate time which illustrates the relation between expectation values and moments.

At this stage, we still have to choose a gauge if we want to relate the non-observables q, t and G^{qp} in this equation to properties in a kinematical Hilbert space. A convenient choice is to treat (t, p_t) like a fully constrained pair.

The choice should ensure that physical quantum variables take real values and the uncertainty relations are satisfied for *all*, physical as well as unphysical, quantum variables. This suggests to fix the gauge by requiring that $G_{tp_t} = -\frac{1}{2}\mathrm{i}\hbar$ has no real part. Then, *physical* quantum variables are real. Moreover, as in the linear case, we can gauge fix $G_{tt} = 0$, such that the uncertainty relation $G_{tt}G_{p_tp_t} - (G_{tp_t})^2 \geq \hbar^2/4$ is saturated independently of the behaviour of the (q, p)-variables. For $G_{tt} \neq 0$, it would depend on those variables via $G_{p_tp_t} \approx p^2 G^{pp}/M^2$ from (9.19). Finally, this is the only gauge condition for G_{tp_t} which works for all values of $\mathcal{P}^{(2)}$, including $\mathcal{P}^{(2)} = 0$ in (9.22).

In this gauge, we obtain

$$q(t) = \mathcal{Q}^{(2)} + \frac{\mathcal{P}^{(2)}}{M}t \quad , \quad G^{qp}(t) = \mathcal{G}^{qp(2)} + \frac{\mathcal{G}^{pp(2)}}{M}t$$

in agreement with the solutions one would obtain for the deparametrized free particle. In this case, there is no quantum back-reaction of quantum variables affecting the motion of expectation values because the particle is free. In the presence of a potential, equations analogous to those derived here would exhibit those effects.

Consistency of constraints up to third-order moments

Including third-order terms in the analysis, solutions to the constraints $C_q^{(1)}$, $C_t^{(1)}$, $C_{p_t}^{(1)}$ and $C_p^{(1)}$ become

$$\begin{aligned}
G^{0,1}_{1,0} &= -\frac{p}{M}\frac{\mathrm{i}\hbar}{2} - \frac{p}{M}G^{1,1}_{0,0} - \frac{1}{2M}G^{2,1}_{0,0}\,,\\
\frac{p}{M}G^{1,0}_{0,1} &= -G^{0,0}_{1,1} - \frac{\mathrm{i}\hbar}{2} - \frac{1}{2M}G^{2,0}_{0,1}\,,\\
G^{0,0}_{2,0} &= -\frac{p}{M}G^{1,0}_{1,0} - \frac{1}{2M}G^{2,0}_{1,0}\,,\\
G^{1,0}_{1,0} &= -\frac{p}{M}G^{2,0}_{0,0} - \frac{1}{2M}G^{3,0}_{0,0}\,.
\end{aligned}$$

As in the previous paragraph, they will be used to determine *second-order* moments. The constraint $C^{(1)}$ contains no third-order terms and thus remains unaltered. *Third-order* moments are determined by higher constraints $C^{(1)}_{qp}$, $C^{(1)}_{tp}$, $C^{(1)}_{p_tp}$, $C^{(1)}_{p^2}$ and $C^{(2)}_q$, $C^{(2)}_t$, $C^{(2)}_{p_t}$.

All other constraints contain third-order moments with a factor of $\hbar$ or of second or higher moments, both of which provides terms of higher order.

For instance, we may consider the constraints $C^{(1)}_{qp^2}$, $C^{(1)}_{tp^2}$, cf. (E.17), (E.18). They both contain third-order moments with a factor of C, which, after solving $C^{(1)}$, becomes a term of fifth order. The remaining second- and third-order terms occur with a factor of $\hbar$, and are thus of fourth and fifth order. From this consideration of orders in the moment expansion, we conclude that $C^{(1)}_{qp^2}$ and $C^{(1)}_{tp^2}$ do not constrain third-order moments but become relevant only at higher than third orders of the approximation scheme.

For $n = 1$ the constraints that actually do determine third-order moments are $C^{(1)}_{qp}$, $C^{(1)}_{tp}$, $C^{(1)}_{p_tp}$ and $C^{(1)}_{p^2}$. On the constraint hypersurface, they imply

$$\begin{aligned}
G^{1,1}_{1,0} &\approx -\frac{p}{M}G^{2,1}_{0,0} + \frac{1}{2M}G^{2,0}_{0,0}\left(G^{1,1}_{0,0} - \mathrm{i}\hbar\right) \ , \\
G^{1,0}_{1,1} &\approx \frac{1}{2M}G^{2,0}_{0,0}G^{1,0}_{0,1} - \frac{p}{M}G^{2,0}_{0,1} \ , \\
G^{1,0}_{2,0} &\approx \frac{1}{2M}G^{2,0}_{0,0}\left(\frac{1}{2M}G^{3,0}_{0,0} + \frac{p}{M}G^{2,0}_{0,0}\right) - \frac{p}{M}G^{2,0}_{1,0},, \\
G^{2,0}_{1,0} &\approx \frac{1}{2M}G^{2,0}_{0,0}G^{2,0}_{0,0} - \frac{p}{M}G^{3,0}_{0,0} \ .
\end{aligned}$$

Note that this holds on the constraint hypersurface defined by the constraints $C^{(1)}$, $C^{(1)}_q$, $C^{(1)}_t$, $C^{(1)}_{p_t}$ and $C^{(1)}_p$. Dropping fourth- and fifth-order terms, we find the simple relations

$$\begin{aligned}
G^{1,1}_{1,0} &\approx -\frac{p}{M}G^{2,1}_{0,0} \ , \quad G^{1,0}_{1,1} \approx -\frac{p}{M}G^{2,0}_{0,1} \ , \\
G^{1,0}_{2,0} &\approx -\frac{p}{M}G^{2,0}_{1,0} \ , \quad G^{2,0}_{1,0} \approx -\frac{p}{M}G^{3,0}_{0,0} \ .
\end{aligned}$$

This happens in a consistent manner because unconstrained third-order moments appear on the right-hand sides. No condition on the (q,p)-moments present in these equations arises in this way, but the third-order moments $G^{1,0}_{1,1}$ and $G^{0,0}_{2,1}$ associated with (t,p_t) remain unspecified at this stage. The constraints $C^{(2)}_q$, $C^{(2)}_t$, $C^{(2)}_{p_t}$ arising for $n = 2$ yield

$$\begin{aligned}
G^{0,1}_{2,0} &\approx \frac{p}{2M^2}G^{2,0}_{0,0}G^{1,1}_{0,0} \ , \\
G^{0,0}_{2,1} &\approx \frac{1}{M}\left(G^{2,0}_{0,0}\left(G^{0,0}_{1,1} + \frac{1}{2M}G^{2,0}_{0,1}\right) + \frac{p^2}{M}G^{2,0}_{0,1}\right) \ , \\
G^{0,0}_{3,0} &\approx 2\frac{p}{M}\left(-\frac{p^2}{2M^2}G^{3,0}_{0,0} + \frac{1}{2M}G^{2,0}_{0,0}\left(\frac{1}{2M}G^{3,0}_{0,0} + \frac{p}{2M}G^{2,0}_{0,0}\right)\right) \ ,
\end{aligned}$$

which, after setting higher-order terms to zero, sets

$$G^{0,1}_{2,0} \approx 0 \quad , \quad G^{0,0}_{2,1} \approx \frac{p^2}{M^2} G^{2,0}_{0,1} \quad , \quad G^{0,0}_{3,0} \approx -2\frac{p^3}{2M^3} G^{3,0}_{0,0} \, .$$

Now we turn to the elimination of degrees of freedom through gauge transformations. The inclusion of third-order terms and new constraints does not affect $\mathcal{P}^{(2)}$ and $\mathcal{Q}^{(2)}$. They remain constant under gauge transformations. We therefore write

$$\mathcal{P}^{(3)} := \mathcal{P}^{(0)} \quad \text{and} \quad \mathcal{Q}^{(3)} := \mathcal{Q}^{(2)} \, .$$

Accordingly, their Poisson bracket is unaltered. The situation is, however, altered for the second-order quantum variables. Only $\mathcal{G}^{pp(2)}$ remains invariant under the flow generated by third-order constraints. Now that third-order terms are included, $\mathcal{G}^{qp(2)}$ and $\mathcal{G}^{qq(2)}$ are no longer observables. The former changes under gauge transformations as follows

$$\begin{aligned} \{\mathcal{G}^{qp(2)}, C^{(1)}_q\} &= \frac{1}{2M} G^{2,1}_{0,0} \quad , \quad \{\mathcal{G}^{qp(2)}, C^{(1)}_t\} = \frac{1}{2M} G^{2,0}_{0,1} \, , \\ \{\mathcal{G}^{qp(2)}, C^{(1)}_{p_t}\} &= \frac{1}{2M} G^{2,0}_{1,0} \quad , \quad \{\mathcal{G}^{qp(2)}, C^{(1)}_p\} = \frac{1}{2M} G^{3,0}_{0,0} \, , \end{aligned}$$

whereas Poisson brackets with $C^{(1)}_{qp}$, $C^{(1)}_{tp}$, $C^{(1)}_{p_t p}$ and $C^{(1)}_{p^2}$ are of fourth order in the moment expansion. The terms on the right-hand side can be eliminated through the addition of a third-order moment by

$$\mathcal{G}^{qp(3)} := \mathcal{G}^{qp(2)} - \frac{1}{2M} G^{2,0}_{0,1} \, .$$

This has vanishing Poisson brackets with all constraints up to fourth-order terms. Moreover, it has vanishing Poisson bracket with $\mathcal{P}^{(3)}$ as well as $\mathcal{Q}^{(3)}$. The Poisson brackets with $\mathcal{G}^{pp(3)} := \mathcal{G}^{pp(2)}$ remains unaltered, $\{\mathcal{G}^{qp(3)}, \mathcal{G}^{pp(3)}\} = 2\mathcal{G}^{pp(3)}$.

Transformations generated by the constraints on $\mathcal{G}^{qq(2)}$ are of a more complicated form and I have not found a simple way of writing $\mathcal{G}^{qq(3)}$ in explicit form.

Nonetheless we can end our calculations because the applicability of effective constraints has been demonstrated.

9.6 Conclusions

I thoroughly motivated and briefly introduced the formalism of generalized effective equations. This is set up on quantum phase space coordinatized by expectation values and higher moments of the wave function which are also referred to as quantum variables. The main part of this chapter was concerned with the application of this scheme to constrained systems.

Two major tasks had to be solved in order to provide such an application. First of all, a handy set of constraints on quantum phase space had to be found that replaces Dirac's quantization condition. The major problem was to ensure that this set would provide neither stronger nor weaker restrictions than the original constraint quantization condition. Secondly, the resulting infinite set had to be dealt with in some way. That means we had to find a consistent approximation method. This method had to be chosen such that quantum variables can be treated order by order. Nonetheless, the approximated system should remove degrees of freedom up to a certain order correctly.

We developed an iteration procedure that allows one to determine a minimal set of quantum constraints. It was shown how this set would look like for linear constraints and moreover demonstrated that the thus chosen set meets all requirements. The degrees of freedom are removed correctly.

Furthermore, we presented a consistent approximation scheme that works for any artificially reparametrization-invariant theory. This procedure also applies to interacting systems. We can solve the constraints in the same manner and using the same orders of constraints. The main consequence in the presence of a potential $V(q)$ is that additional q-moments appear as extra terms in solutions at certain orders, whose precise form depend on the potential. For a small potential, this can be dealt with by perturbation theory around the free solutions.

The relativistic, or truly reparametrization-invariant case, awaits further study.

References:

The PhD thesis of Troy Schilling provides a very understandable review of the geometrical formulation of quantum theory, [83]. The foundations of the generalized effective equation scheme were laid in [81]. In [84] the relation of the framework to the effective action is discussed. Work presented in this chapter was published in [80]. The correct removal of degrees of freedom with linear constraint operator was shown by Artur Tsobanjan. I carried out the development of a consistent approximation scheme under supervision of Martin Bojowald.

Chapter 10

Conclusions

In this thesis, I gave a careful derivation of quantum general relativity. I paid particular attention to the problems of this approach. Quantum general relativity is usually blamed for not allowing a positive definite inner product and thus no Hilbert space structure. A positive definite inner product is however not what one would expect from a constrained quantum theory. To the contrary, one has to require that the theory does not allow a positive definite inner product because it still contains unphysical states.

The second reproach quantum general relativity has to face are divergences arising from operator products at the same space point. These, however, are an inherent feature of the functional Schrödinger picture. Every quantum field theory when formulated in this picture suffers from these divergences. Thus this type of divergences is a problem of the functional Schrödinger picture and not a specific deficit of quantum general relativity. The fact that these divergences cannot be simply removed by some renormalization procedure, is however owed to general relativity.

The third problem is related to the divergence of the Wheeler–DeWitt equation whenever the three-metric becomes degenerate.

This bridges the gap to the major topic of this thesis, namely the fate of singularities in quantum cosmology. First of all, I argued that we should not expect that quantum general relativity is a theory free of singularities. Simply because it is not the final and fundamental theory. Furthermore, I used the example of singular potentials in classical and quantum mechanics to demonstrate that we cannot expect that the singularities of classical general relativity do not occur on the quantum level as well. Namely, in the case of singular potentials, classical singularities persist on the quantum level. A familiar example is the inverse-square potential. From this point of view, *the motivation of quantum gravity through the existence of classical singularities is invalid*.

And indeed, the divergence of the Wheeler–DeWitt equation for degenerate three-metrics signals that *some* singularities persist on the quantum

level.

For the cosmological model this divergence arises for vanishing scale-factor, i.e. at the usual location of the big-bang singularity. I showed, however, that for a Friedmann–Robertson–Walker universe with cosmological constant for a whole class of factor-orderings a unique solution to the Wheeler–DeWitt equation exists which can be continuously extended to zero scale factor. So in the simple quantum-cosmological models considered here, the degeneracy of the three-metric does not prevent the theory from providing unique predictions.

A way to cope with this divergence of the Wheeler–DeWitt equation in the full theory, is to exclude degenerate three-metrics from the very beginning. This just amounts to an implementation of the positivity condition of the three-metric.

All classical singularities that do not correspond to a degenerate three-metric are absent on the quantum level — as long as they do not correspond to singular points of the matter Hamiltonian.

From this picture, two criteria seem to be sensible benchmarks to judge singularity resolution. The first one is the vanishing of the wave function at the location of the classical singularity. This, of course, can only apply if the classical singularity can be tracked down in configuration space. The vanishing of the wave function then simultaneously turns the Wheeler–DeWitt equation into a well-defined equation. Note that a point on which the wave function vanishes cannot contribute to the quantum theory.

The second criterion is the break-down of the semi-classical approximation. It applies especially to singularities which cannot be located in configuration space. Here, the dissolution of spacetime suffices to account for singularity resolution. The end points of classically incomplete geodesics are hidden in a quantum region. That means observers cannot 'fall' into the singularity because spacetime itself dissolves before the singularity is reached.

With these two criteria at hand, I approached two different cosmological scenarios. Both scenarios are dominated by singularities which end the evolution of the universe at *large* scale factor. The first scenario is motivated by phantom dark energy. This type of dark energy generically drives the universe into a big-rip singularity at which the scale factor diverges after a finite amount of time. So in this model, the singularity is found at the boundary of configuration space opposite to the big bang.

In the second scenario, the universe is governed by an anti-Chaplygin gas. This type of matter produces a big-brake singularity which ends the universe's evolution at finite, large scale factor.

Both singularities are resolved on the quantum level. The big-rip singularity does not correspond to a singularity of the Wheeler–DeWitt equation. As expected, it is therefore resolved through a mere spreading of wave packets, i.e. a break-down of the semi-classical approximation. As a side, let me

remark that an exact solution to the Wheeler–DeWitt equation was given for one of the phantom models. For each phantom model with big-rip singularity, a corresponding scalar-field model was studied. These exhibit a big-bang singularity. The big bang can be tracked down in configuration space and does correspond to a singular point of the Wheeler–DeWitt equation. Consequently it is, as expected, resolved through a vanishing of the wave function. The big-brake singularity, on the other hand, corresponds to a singularity of the matter potential. The wave function is consequently found to vanish at the big brake, but also spreads upon approach of the classical singularity.

So, we can conclude that these more exotic types of singularity are resolved on the quantum level. This resolution of singularities occurring at large scale factor stands as proof for the occurrence of quantum gravitational effects in the macroscopic universe.

The last chapter does not relate to the previous ones. It was devoted to the development of a formalism that allows the application of generalized effective equations to constrained systems. The two major hurdles to clear were the consistent formulation of a set of constraints on quantum phase space and the development of a consistent approximation scheme for these. Both difficulties were overcome. The consistency of the set of constraints was shown for a single, linear constraint and thus for all singly constrained systems at least locally. The applicability of the approximation scheme was demonstrated for artificially reparametrization-invariant theories. The extension to relativistic systems awaits further study.

Appendix A

Generalized Hamiltonian dynamics

In classical mechanics, one can distinguish between regular and singular systems. Singular systems have the property that the matrix of second derivatives of the Lagrangian with respect to the velocities is degenerate, i.e.

$$\det\left(\frac{\partial^2 L(q,\dot{q})}{\partial\dot{q}^i\dot{q}^j}\right) = 0$$

for a system with Lagrangian $L(q,\dot{q})$ depending on positions q^i and velocities $\dot{q}^i = \frac{\mathrm{d}q^i}{\mathrm{d}t}$ for each degree of freedom $i = 1,\dots N$. There are certain peculiarities arising from this degeneracy, in the Lagrangian as well as in the Hamiltonian framework.

As the Hamiltonian framework is the starting point for quantization, I will focus on the setup of the Hamiltonian formulation for such a singular system. This formulation is given by the *Bergmann–Dirac algorithm*.

A.1 The Bergmann–Dirac algorithm for classical mechanics

As described above, we start from a system with action

$$\mathrm{S} = \int \mathrm{d}t\, L(q,\dot{q}) \ ,$$

where q^i and $\dot{q}^i$ for $i = 1,\dots N$ are the positions and velocities of a mechanical system. We assume that this system is consistent but singular. The aim is to arrive at a Hamiltonian formulation for it. In the following we will use q, $\dot{q}$ or p to refer to *all* q^i, $\dot{q}^i$ and p_i.

A.1.1 Canonical momenta and primary constraints

The Hamiltonian framework is set up on phase space Γ consisting of all positions q and canonically conjugate momenta p defined via

$$p_i = \frac{\partial L(q,\dot{q})}{\partial \dot{q}^i} \ .$$

Because we are dealing with a singular system, the matrix $W = \left(\frac{\partial p_i}{\partial \dot{q}^j}\right)$ is degenerate, i.e. it does not have full rank. Let the rank be $R < N$. This means that we can eliminate just R of the velocities in terms of the momenta and the remaining velocities,

$$\dot{q}^a = f^a(q, p_\alpha, \dot{q}^\rho) \ ,$$

where $a = 1, \dots R$, $\alpha = 1, \dots R$ and $\rho = R+1, \dots N$. The velocities $\dot{q}^\rho$ remain in the formalism. We thus have

$$p_i = g_i(q, p_\alpha, \dot{q}^\rho)$$

for the canonical momenta. The set of momenta dissociates into two groups: those which still *contain* velocities

$$p_\alpha = g_\alpha$$

and those which do *not*

$$p_r = g_r(q, p_\alpha, \dot{q}^\rho) \ , \qquad \text{where } r = R+1, \dots N \ . \tag{A.1}$$

The latter are constraints. As they arise in the definition of the momenta, they are called primary constraints.

A singular Lagrangian system thus corresponds to a Hamiltonian system with $2N$ variables $(q, p_\alpha, \dot{q}^\rho)$ set up on a hypersurface Γ_p of phase space defined by the primary constraints, (A.1).

A.1.2 Canonical Hamiltonian and equations of motion

We define the quantity

$$\mathrm{H_c}(q, p_\alpha, \dot{q}^\rho) = p_i \dot{q}^i - L(q, \dot{q}) \ .$$

Inserting the defining equations for the momenta, this becomes

$$\mathrm{H_c}(q, p_\alpha, \dot{q}^\rho) = p_\alpha \dot{f}^\alpha + g_r \dot{q}^r - L(q, \dot{q}) \ .$$

Differentiating this quantity with respect to the variables $(q, p_\alpha, \dot{q}^\rho)$, we find the equations

$$\dot{q}^a = \frac{\partial \mathrm{H}_c}{\partial p_a} - \frac{\partial g_r}{\partial p_a}\dot{q}^r \ , \tag{A.2}$$

$$\dot{p}_i = -\frac{\partial \mathrm{H}_c}{\partial q^i} + \frac{\partial g_r}{\partial q^i}\dot{q}^r \ , \tag{A.3}$$

whereas $\frac{\partial \mathrm{H}_c}{\partial \dot{q}^\rho} = 0$. These equations hold by means of the Euler–Lagrange equations, i.e. when the equations of motion are satisfied (because we replaced $\frac{\partial \mathrm{L}}{\partial \dot{q}^i}$ by $\dot{p}_i$).

The Hamiltonian thus yields $N + R$ equations of motion. The $N - R$ velocities $\dot{q}^r$ remain undetermined. Thus we have equations of motion on the constraint hypersurface Γ_p – but only for the q^a and all p_i. The equations look very similar to the Hamiltonian equations of motion for a regular system but are not symmetric in p and q (more equations for p than for q) and contain undetermined functions $\dot{q}^r$.

The decisive step in the Bergmann–Dirac algorithm consists now in *extending* this set of equations as well as the Hamiltonian *to the entire phase space* Γ. To this end it is convenient to introduce the notions of weak and strong equalities.

A.1.3 Weak and strong equalities

Weak and strong equalities can be defined for any function $F(p, q)$ on a neighbourhood of Γ_p. The function $F(p, q)$ is *weakly* equal to zero if $F(p, q)|_{\Gamma_p} = 0$. We write

$$F(p, q) \approx 0 \qquad \text{if} \qquad F(p, q)|_{\Gamma_p} = 0.$$

The function $F(p, q)$ vanishes *strongly* if itself and its gradient vanish weakly. We use the notation

$$\begin{aligned} F(p, q) \simeq 0 \qquad \text{if} \qquad & F(p, q)|_{\Gamma_p} = 0 \\ \text{and} \qquad & \left(\tfrac{\partial F}{\partial q^i}, \tfrac{\partial F}{\partial p_i}\right)|_{\Gamma_p} = 0 \ . \end{aligned}$$

Fixing a function on the primary constraint hypersurface up to first-order derivatives is sufficient for the reformulation of the Hamilton equations of motion, because they only contain first-order derivatives with respect to phase-space variables.

One can then define the *primary constraint hypersurface* by a set of weakly vanishing functions

$$G_r(p, q) = p_r - g_r(q, p_\alpha) \approx 0 \ . \tag{A.4}$$

It can be shown (for the proof see Sundermeyer's book [23], p. 48) that every weakly vanishing function is strongly equal to a linear combination of the primary constraints G_r,

$$F(p,q) \approx 0 \qquad \Leftrightarrow \qquad F(p,q) \simeq G_r(p,q)\frac{\partial F(p,q)}{\partial p_r} \ . \tag{A.5}$$

A.1.4 Primary Hamiltonian and extended equations of motion

The generalized Hamiltonian H

With this result and these definitions, one can extend the Hamiltonian H_c over the entire phase space. This generalized Hamiltonian H' has to be equivalent to the canonical Hamiltonian on Γ_p, $\mathrm{H}_c \approx \mathrm{H}'$. According to the previous result (A.5) for $F(p,q) = \mathrm{H}_c - \mathrm{H}' \approx 0$ it follows that

$$\mathrm{H}_c - \mathrm{H}' + \frac{\partial \mathrm{H}'}{\partial p_r} G_r(p,q) \simeq 0 \ ,$$

where we used the fact that H_c does not depend on p_r. By definition, the derivatives of this expression with respect to all positions and momenta vanish weakly. Employing the equations of motion (A.2), we find that

$$\begin{aligned} \dot{q}^a &\approx \frac{\partial}{\partial p_a}\left[\mathrm{H}' - \frac{\partial \mathrm{H}'}{\partial p_r} G_r(p,q) - g_r \dot{q}^r\right] \ , \\ -\frac{\partial L}{\partial q^i} &\approx \frac{\partial}{\partial q^i}\left[\mathrm{H}' - \frac{\partial \mathrm{H}'}{\partial p_r} G_r(p,q) - g_r \dot{q}^r\right] \ . \end{aligned}$$

We can now replace g_r by G_r in these equations as the derivatives of p_r with respect to p_a and q^i vanish. Note further that

$$\frac{\partial G_r}{\partial p_i}\dot{q}^r = \frac{\partial G_r}{\partial p_a}\dot{q}^r \ ,$$

so that we finally obtain a symmetric form of the equations of motion

$$\begin{aligned} \dot{q}^i &\approx \{q^i, \mathrm{H} + \dot{q}^r G_r\} \ , \\ \frac{\partial L}{\partial q^i} &\approx \{p_i, \mathrm{H} + \dot{q}^r G_r\} \ , \end{aligned}$$

where $\mathrm{H} = \mathrm{H}' - G_r \frac{\partial \mathrm{H}'}{\partial p_r}$. The price we pay for the additional equations of motion for the $\dot{q}^r$ is that the equations of motion are now only weak equations. Also the equations contain the not completely fixed function H. The latter problem can easily be circumvented as $\mathrm{H}_c \simeq \mathrm{H}$. Therefore the equations of motion can be written as

$$\begin{aligned} \dot{q}^i &\approx \{q^i, \mathrm{H}_c + \dot{q}^r G_r\} \ , \\ \frac{\partial L}{\partial q^i} &\approx \{p_i, \mathrm{H}_c + \dot{q}^r G_r\} \ . \end{aligned}$$

The primary Hamiltonian H_p

In a last step, one takes care of those cases in which the constraints cannot be solved for the momenta p_r. In general this is not possible and we thus cannot assume that the primary constraint hypersurface be described by equations of the form G_r, (A.4). We therefore characterize Γ_p by the weakly vanishing

$$\phi_r(p,q) \approx 0 \ , \qquad \text{where} \qquad r = R+1, \ldots N \ .$$

We relate this to the former expression G_r by taking derivatives with respect to q and p_a,

$$\begin{aligned} \frac{\mathrm{d}\phi_r}{\mathrm{d}q^i} &= \frac{\partial \phi_r}{\partial q^i} + \frac{\partial \phi_r}{\mathrm{d}p_s}\frac{\partial g_s}{\mathrm{d}q^i} \approx 0 \ , \\ \frac{\mathrm{d}\phi_r}{\mathrm{d}p_a} &= \frac{\partial \phi_r}{\partial p_a} + \frac{\partial \phi_r}{\mathrm{d}p_s}\frac{\partial g_s}{\mathrm{d}p_a} \approx 0 \ . \end{aligned}$$

The matrix $V = (\frac{\partial \phi_r}{\mathrm{d}p_s})$ is invertible and therefore we can solve these equations for $\frac{\partial g_s}{\mathrm{d}q^i}$ and $\frac{\partial g_s}{\mathrm{d}p_a}$. We can again substitute g_r by G_r weakly. Thus we can rewrite the equations of motion

$$\begin{aligned} \dot{q}^i &\approx \{q^i, \mathrm{H}_c\} + \dot{q}^s V^{-1}_{sr} \frac{\partial \phi_r}{\mathrm{d}p_i} \ , \\ \dot{p}_i &\approx \{p_i, \mathrm{H}_c\} + \dot{q}^s V^{-1}_{sr} \frac{\partial \phi_r}{\mathrm{d}q^i} \ . \end{aligned}$$

Introducing $\mu_r = \dot{q}^s V^{-1}_{sr}$, we can define the *primary Hamiltonian*

$$\mathrm{H}_p = \mathrm{H}_c + \mu^r \phi_r$$

with the help of which the time-development for an arbitrary phase space function $A(p,q)$ can be written as

$$\dot{A} \approx \{A, \mathrm{H}_p\} + \mu^r \{A, \phi_r\} \ .$$

These μ_r are then seen to contain the arbitrariness carried in the velocities $\dot{q}^r$ which could not be expressed in terms of momenta.

A.1.5 Consistency and secondary constraints

The system can only be described consistently if the constraints themselves are conserved in time. This can yield new constraints or restrict the multplier functions μ^r, depending on the form of the consistency conditions. These are

$$\dot{\phi}_r \approx \{\phi_r, \mathrm{H}_c\} + \mu^s \{\phi_r, \phi_s\} \ .$$

Using the short-hands $h_r = \{\phi_r, \mathrm{H}_c\}$ and $P_{rs} = \{\phi_r, \phi_s\}$, the following four cases can arise.

Case I: $h \not\approx 0$

Case IA: $\det P \not\approx 0$

In this case, P can be inverted and multiplier functions μ be determined through

$$\mu^s \approx -P^{-1}_{rs} h_r \ .$$

The equations of motion thus become

$$\dot{A} \approx \{A, \mathrm{H}_p\} - \{A, \phi_s\} P^{-1}_{rs} \{\phi_r, \mathrm{H}_c\} \ ,$$

and do no longer contain any ambiguities. The motion is completely determined.

Case IB: $\det P \approx 0$

The matrix P does not have full rank. Let its rank be denoted by $M < N - R$. In this case, P has $N - R - M$ null eigenvectors $e^{(\alpha)}$ yielding new constraints

$$e^{(\alpha)}_r P_{rs} \approx 0 \qquad \Rightarrow \qquad e^{(\alpha)}_r h_r \approx 0 \ .$$

These constraints restrict the primary constraint hypersurface further to a hypersurface $\Gamma' \subset \Gamma_p$. The new constraints are called *secondary constraints*.

Case II: $h \approx 0$

Case IIA: $\det P \not\approx 0$

Now the fact that P can be inverted fixes the μ^r — they have to vanish weakly $\mu^s P_{rs} \approx 0 \Leftrightarrow \quad \mu^s \approx 0$. Thus the canonical Hamiltonian is equal to the primary one.

If $\mathrm{H}_c = 0$, one therefore has to demand that $\det P \approx 0$ — otherwise the theory is trivial. This is an additional, secondary constraint.

Case IIB: $\det P \approx 0$

In this case, the condition $\mu^s P_{rs} \approx 0$ yields constraints on the multipliers μ. Let the rank of P be M, then $N-R-M$ multipliers get weakly fixed through these equations.

Requiring the constraints to be conserved in time yields a number of secondary constraints that further shrink the constraint hypersurface. Moreover, some of the multiplier functions get fixed. We denote the new constraint hypersurface by Γ'. It is defined via

$$\begin{aligned} \phi_r(p,q) &\approx 0\ , \qquad r = R+1, \dots N \qquad \text{and} \\ \chi_{\rho'}(p,q) &\approx 0\ , \qquad \rho' = 1, \dots L'\ . \end{aligned}$$

Weak equalities from now on refer to Γ'.

From the arisal of new constraints, two iteration procedures result.

First of all, the primary constraints have to be conserved on the *new* constraint hypersurface Γ'. From this requirement one might get more (tertiary) constraints that restrict the surface to Γ'' etc. In the end we will arrive at a constraint hypersurface Γ'' described by the weak vanishing of

$$\phi_r(p,q) \approx 0\ , \qquad r = R+1, \dots N \qquad \text{and} \tag{A.6}$$

$$\chi_{\rho''}(p,q) \approx 0\ , \qquad \rho'' = 1, \dots L''\ , \tag{A.7}$$

where $L'' \geq L'$. We will call all new constraints arising in this process secondary constraints.

The second iteration procedure takes care of the time-conservation of these new, secondary constraints. Whether or not new constraints arise at this stage depends now on the matrix

$$\begin{pmatrix} \{\phi_r, \phi_s\} \\ \{\chi''_\rho, \phi_s\} \end{pmatrix}\ .$$

For every null eigenvector of this matrix we get a new constraint from multiplication of $\{\phi_r, \mathrm{H}_c\} + \{\chi_\rho, \mathrm{H}_c\}$ by the null eigenvector.

Iteration of these two procedures yields the final *constraint hypersurface* Γ_c defined by a set of primary and secondary constraints

$$\begin{aligned} \phi_r(p,q) &\approx 0\ , \qquad r = R+1, \dots N \qquad \text{and} \\ \chi_\rho(p,q) &\approx 0\ , \qquad \rho = 1, \dots L \geq L''\ . \end{aligned}$$

For every null eigenvector $e^{(i)}$ of

$$D = \begin{pmatrix} \{\phi_r, \phi_s\} \\ \{\chi_\rho, \phi_s\} \end{pmatrix}$$

the equations

$$e_r^{(i)}\{\phi_r, \mathrm{H}_c\} + e_\rho^{(i)}\{\chi_\rho, \mathrm{H}_c\} \approx 0$$

are satisfied on Γ_c. Furthermore the functions μ^s are subject to the conditions

$$\{\phi_r, \mathrm{H}_c\} + \mu^s\{\phi_r, \phi_s\} \approx 0 \tag{A.8}$$
$$\{\chi_\rho, \mathrm{H}_c\} + \mu^s\{\chi_\rho, \phi_s\} \approx 0 \tag{A.9}$$

on Γ_c. In practice, this procedure terminates after a few steps and in each step it is very obvious what to do without recurring to matrices P or D and their rank.

A.1.6 Determination of multiplier functions μ

In the last paragraph we just noted that the multiplier functions get fixed through consistency relations for certain primary and secondary constraints. We now want to look somewhat closer on how this happens. More specifically, we want to know *how many* of the multiplier functions get fixed.

To answer this question, it is useful to introduce the notion of first- and second-class constraints. A *first-class constraint* has weakly vanishing Poisson brackets with all other constraints *and* all linear combinations of them. All constraints that do not have this property are called *second class*. Note that here and in the following, weak vanishing refers to the final constraint hypersurface Γ_c.

With these notions at hand, we can discuss the fixing of multiplier functions. Conditions on μ arise whenever the matrix D in (A.8) has full rank $N - R$. This is just Case IA discussed above. If the rank of D is $K < N - R$, then K multipliers get fixed through the above equations. From the $N - R - K$ null eigenvectors $e^{(J)}$ of D, one can construct the new constraints

$$\phi^J = e_r^{(J)}\phi_r, \qquad J = 1, \ldots N - R - K$$

through linear combination. Because the $e^{(J)}$ are null eigenvectors of D, we find that

$$\{\phi^J, \phi_r\} \approx 0\ , \qquad \{\phi^J, \chi_\rho\} \approx 0\ .$$

The new constraints are first class. We call the remaining primary constraints ϕ^i where $i = N - R - K + 1, \ldots N - R$. These are then necessarily second class. We carry out a similar split for the secondary constraints. The first-class ones will be denoted by χ^A, the second-class ones by χ^a. In this step, *primary and secondary constraints get mixed*. Thus we will only speak

of primary first- and second-class constraints ϕ but the remaining constraints χ will no longer be referred to as secondary as they are linear combinations of secondary *and* primary constraints. We write therefore $\mu_s\phi^s = \phi^J v_J + \phi^i u_i$. Inserting this in (A.8), we obtain the following equations

$$\{\phi^J, \mathrm{H}_c\} \approx 0 \tag{A.10}$$
$$\{\chi^A, \mathrm{H}_c\} \approx 0 \tag{A.11}$$
$$\{\phi^i, \mathrm{H}_c\} + u_j\{\phi^i, \phi_j\} \approx 0 \tag{A.12}$$
$$\{\chi^a, \mathrm{H}_c\} + u_j\{\chi^a, \phi_j\} \approx 0 \ . \tag{A.13}$$

There are several observations one can make at this point. First of all, we see that any first-class constraint necessarily commutes with the Hamiltonian. Secondly, the multipliers associated with the primary first-class constraints all drop out. Thus there are as many undetermined multiplier functions as there are *primary* first-class constraints. Thirdly, the multiplier functions associated with the second-class constraints are determined by (A.10).

They are fixed to

$$u_j \approx -\Delta^{-1}_{j\mu}\{\xi^\mu, \mathrm{H}_c\} \ ,$$

where $\xi^\mu = (\phi^i, \chi^a)$ stands for any second-class constraint and the matrix Δ is given by

$$\Delta = \begin{pmatrix} \{\phi^i, \phi^j\} \\ \{\chi^a, \phi^j\} \end{pmatrix} \ .$$

The derivation of these results can be found in Sundermeyer's book, [23], p. 58 — it is essentially the same calculation as in Case IA above.

The equation of motion for a general phase space function $A(p, q)$ is then given by

$$\dot{A} \approx \{A, \mathrm{H}_c\} + v_J\{A, \phi^J\} - \{A, \xi^\nu\}\Delta^{-1}_{\mu\nu}\{\xi^\mu, \mathrm{H}_c\} \ . \tag{A.14}$$

So whereas all second-class constraints appear in a symmetric manner, the primary first-class constraints play a distinguished role.

A.1.7 First-class constraints and gauge transformations

Gauge transformations

The equation of motion (A.14) shows that there is some arbitrariness in the time evolution of phase space functions due to the undetermined v_J. More specifically, given an initial value A_0 of a phase space function A at time $t = t_0$, the evolution of $A(t)$ is not uniquely determined. The transformation between two solutions $A(t)$ obtained through the evolution with two different

parameters v is called a *gauge transformation. A priori this has nothing to do with gauge theories nor the invariance of the Lagrangian under certain (rigid or local) gauge transformations.* However, there is a link between gauge transformations defined here and local gauge invariance of the Lagrangian. This connection will be discussed now.

We assume for the time being that the system at hand has only first-class constraints so that the primary Hamiltonian is $\mathrm{H}_p = \mathrm{H}_c + \lambda_J \phi^J$. The gauge transformation maps two solutions $A_\lambda(t)$, $A_{\lambda'}(t)$ with $A_\lambda(t_0) = A_{\lambda'}(t_0) = A_0$ onto each other. If we expand

$$A_\lambda(t) = A_0 + \{A_0, \mathrm{H}_c\}t + \lambda_J(0)\{A_0, \phi^J\}t + \text{ higher derivatives },$$

corrBthe following transformation is a gauge transformation

$$A_\lambda(t) - A_{\lambda'}(t) = t\left(\lambda_J(0) - \lambda'_J(0)\right)\{A_0, \phi^J\} = \delta_\epsilon A = \epsilon_J\{A_0, \phi^J\}\ . \quad \text{(A.15)}$$

First of all, one can see from calculating the commutator of two such gauge transformations δ_ϵ, δ_η that in general these transformations do *not* form a group. The group property holds only weakly. Moreover, the primary first-class constraints are not the only constraints generating gauge transformations. Evolution over a larger amount of time makes it necessary to include also second derivatives $\ddot{A}(t)$. Thus also $\dot{\phi}^J$ generate gauge transformations. In general, one can construct the set of all gauge transformations iteratively, including higher-order time derivatives in (A.15). Let the set G_0 be the set of all first-class primary constraints. Then we find from the inclusion of second derivatives that $G_1 = G \cup \{G_0, \mathrm{H}_p\}$ is the set of gauge transformations. Here, $\{G_0, \mathrm{H}_p\}$ stands for the set of constraints that arises from the derivation of the constraints in G with respect to time. This procedure has to be iterated, $G_{i+1} = G_i \cup \{G_i, \mathrm{H}_p\}$, until $G_{i+1} = G_i$. We call the resulting set of gauge transformation-generating constraints G. It follows from the construction that, firstly, all primary first-class constraints generate gauge transformations. Secondly, G does not necessarily exhaust the set of all first-class constraints. *There may be first-class constraints that do not generate gauge transformations.* Before we turn to the content of the Dirac conjecture, we want to briefly shed light on the relationship between gauge transformations as they are defined here and local gauge transformations.

Gauge transformations and Lagrangian invariance

In the Lagrangian approach, the name of gauge transformations is used for transformations that leave the Lagrangian invariant. Depending on the character of these transformations, one speaks of *rigid* or *local gauge transformations.* They are the subject of the first and second Noether theorem, respectively.

Assume a system with action

$$S[\phi^A] = \int \mathrm{d}^\mu x \, \mathcal{L}[\phi^A, \partial\phi^A, x^\mu]$$

describing a theory with fields ϕ^A depending on coordinates x^μ. Let $\mathcal{L}[\phi^A, \partial\phi^A, x^\mu]$ be the system's Lagrangian density and $\partial\phi^A$ a short-hand denoting the dependence of the Lagrangian on $\partial_\mu\phi^A = \frac{\partial\phi^A}{\partial x^\mu}$.

If under an infinitesimal transformation

$$x^\mu \longrightarrow x'^\mu = x^\mu + \delta x^\mu \ , \tag{A.16}$$

$$\phi^A(x) \longrightarrow \phi'^A(x') = \phi^A(x) + \delta\phi^A \ , \tag{A.17}$$

the Lagrangian remains invariant, we say that the transformations (A.16) generate gauge transformations of the system. Inserting (A.16) into the Lagrangian and requiring $\delta\mathcal{L} = \mathcal{L}[\phi'^A, \partial'\phi'^A, x'^\mu] - \mathcal{L}[\phi^A, \partial\phi^A, x^\mu] = 0$, one finds the condition

$$\partial_\mu \left[\mathcal{L}\delta x^\mu + \frac{\delta\mathcal{L}}{\delta(\partial_\mu\phi^A)}\delta^*\phi^A \right] + \mathrm{L}_A\delta^*\phi^A = 0 \ , \tag{A.18}$$

where $\delta^*\phi^A = \delta\phi^A - \partial_\rho\phi^A\delta x^\rho$ is the substantial variation of ϕ^A and $\mathrm{L}_A = \frac{\delta\mathcal{L}}{\delta\phi^a} - \partial_\mu\left(\frac{\delta\mathcal{L}}{\delta(\partial_\mu\phi^a)}\right)$ are the Euler derivatives. This statement holds irrespective of the form of δx^μ and $\delta\phi^A$. We specify these further. Let the transformations be parametrized by r parameters ϵ_k,

$$\delta x^\mu = \epsilon_k\chi_k^\mu \qquad \delta\phi^A = \epsilon_k\xi_k^A \ .$$

corrBIf these are constant, we speak of *rigid* gauge transformations. In this case, Noether's theorem tells us that there exist r conserved currents — if the equations of motion are satisfied (just use (A.18)). We are however interested in *local* gauge transformations. These are transformations with x^μ-dependent $\epsilon_k(x)$. Writing

$$\delta\phi^A = \epsilon_k\xi_k^A + \epsilon_{k,\mu}\,\psi_k^{A\mu} \ ,$$

we get from (A.18) the so-called generalized Bianchi identities

$$\mathrm{L}_A\left(\xi_k^A - (\partial_\mu\phi^A)\chi_k^\mu\right) - \partial_\mu\left(\mathrm{L}_A\psi_k^{A\mu}\right) = 0 \ . \tag{A.19}$$

These are trivially satisfied if the equations of motion are. But they hold also if the equations of motion are not satisfied, showing that the set of Euler–Lagrange equations is not independent. Also, one can show that (A.19) implies that the Lagrangian is singular. Therefore, *any locally gauge-invariant theory yields constrained Hamiltonian dynamics.* (The reverse,

however, is not true: Singular Lagrangians with second-class constraints on the Hamiltonian level are not locally gauge-invariant.)

On the Hamiltonian level, such a local gauge transformation is a canonical transformation. One finds that the generators of canonical transformations corresponding to local gauge transformations are in fact linear combinations of *all* first-class constraints. On the other hand, out of the set of transformations $\delta_{\underline{\epsilon}} A = \epsilon_\alpha \{A, \varphi^\alpha\}$, where $\varphi^\alpha = (\phi^J, \chi^A)$ is the set of *all* fist-class constraints, one can construct those which leave the Lagrangian invariant. The condition $\delta_{\underline{\epsilon}} \mathcal{L} = 0$ restrict the ϵ_α. The local gauge transformation that leaves the Lagrangian invariant is usually a linear combination of first-class constraints.

Thus it is generally not possible to construct such a transformation out of the primary first-class constraints alone.

A.1.8 The Dirac conjecture

Because, firstly, local gauge transformations are generally a linear combination of *all* first-class constraints and, secondly, the set of gauge transformations generally exceeds the set of primary first-class constraints, one may be tempted to generalize the Hamiltonian even further and introduce the extended Hamiltonian

$$\mathrm{H}_E = \mathrm{H}_p + \mu_A \chi^A \ ,$$

containing *all* first-class constraints. In fact, Dirac conjectured that all first-class constraints generate gauge transformations and consequently the replacement of H_p by H_E was admissible in the equations of motion. This is just the content of the *Dirac conjecture*. There are certain arguments that speak in favour of such a generalization of the Hamiltonian.

First of all, depending on whether one uses a first- or second-order action formalism, the role of primary and secondary constraints may be interchanged. Moreover, the secondary first-class constraints are in a sense already contained in the primary Hamiltonian. Namely, one can rewrite $\mathrm{H}_c = \mathrm{H}' + \lambda_A \chi^A$, where $\lambda_A = \lambda_A(q, p_a)$, see Matschull [5] for illustrative examples. Thirdly, due to the ambiguous functions $\dot{q}^r$, the positions q^r themselves remain undetermined. But as $\lambda_A = \lambda_A(q, p_a)$, these are also undetermined and therefore generate gauge transformations. Lastly, one can state that H_E at least captures the full gauge aspect of the theory and therefore is suitable for the analysis of gauge issues.

On the other hand, there exist examples, see [23], where G is not the set of all first-class constraints — thus violating the Dirac conjecture. Moreover, there is no indication in the Bergmann–Dirac algorithm itself for such an extension of the primary Hamiltonian. Note that here, all statements were only local as they refer to local coordinate charts. Sundermeyer adds that a geometrical and global formulation as it was given by Gotay, Nester and

Hinds (see the reference in Sundermeyer's book) does neither indicate the generalization to the extended Hamiltonian.

Conclusively, whether or not Dirac's conjecture holds has to be specified on a case-by-case basis. So in the end, one has to check for the specific model at hand which transformations are indeed gauge transformations and which are not.

A.2 The Bergmann–Dirac algorithm for field theories

If one tries to generalize the Bergmann–Dirac algorithm to field theories, complications arise. These are essentially based on the fact that not all conclusions drawn in the above section carry over from the finite-dimensional to the infinite-dimensional case.

I will give a brief list of the complications that arise and refer again to Sundermeyer's book for an illustrative example and further explanations.

Let the field theory be described by fields Q and momenta Π. The primary constraints then generally contain 'spatial' derivatives of these fields, $\phi_r = \phi_r[Q, \Pi, \partial_i Q, \partial_j \Pi]$.[1]

So, first of all, the constraints are no longer algebraic relations but differential equations. Moreover, for each $\phi_r \approx 0$ (now referring to the primary constraint hypersurface) its spatial derivatives and spatial integrals also vanish weakly. Thus we can no longer conclude that a weakly vanishing functional is a linear combination of the primary constraints. Recall that this was a decisive step in the derivation of the generalized Hamiltonian.

Due to the spacetime dependence of the fields and momenta, each constraint $\phi_r \approx 0$ stands for an infinite number of constraints: $\phi_r \approx 0$ at each space point. All summations are thus augmented by an integration over all space points. For example, the primary Hamiltonian becomes

$$\mathrm{H}_p = \mathrm{H}_c + \sum_r \int \mathrm{d}^3x \, u_r(x)\phi_r(x) \ ,$$

ensuing the following consistency relations

$$\{\phi_s, \mathrm{H}_c\} + \int \mathrm{d}^3y \, u_r(y)\{\phi_s(x), \phi_r(y)\} \approx 0 \ . \tag{A.20}$$

The matrix P is consequently infinite-dimensional,

$$P_{rs}(x^a, y^b) = \{\phi_s(x), \phi_r(y)\}_{x^0=y^0} \ ,$$

[1] 'Spatial' here depends on the choice of time variable with respect to which velocities are defined.

where (x^a, y^b) denote dependence just on the spatial coordinates of x and y. If its determinant is non-zero on the constraint hypersurface, an inverse exists. But this inverse is no longer unique.

Equivalently, one can look at (A.20) and expand

$$\begin{aligned} \{\phi_s(x), \phi_r(y)\}_{x^0=y^0} &= a_{rs}\delta(x^a - y^a) + b^i_{rs}\partial_i\delta(x^a - y^a) \\ &+ \text{higher-order derivatives .} \end{aligned}$$

If $b^i_{rs} \not\approx 0$, (A.20) is a differential equation for u_r. Thus, boundary conditions are needed to get a unique solution u_r from (A.20). The same holds if the determinant of P is weakly vanishing. Then the null eigenvectors are not uniquely defined and again we need boundary conditions to fix the arbitrariness in the formalism.

So to make the Bergmann–Dirac algorithm work for field theories, additional conditions in the form of boundary conditions are needed. Otherwise the secondary constraints cannot be found nor the multiplier functions uniquely be fixed.

References:

In this appendix, I followed the presentation given in Sundermeyer's book on 'Constrained Dynamics', [23], very closely. I focussed on the relevant properties of systems with first-class constraints only. This is the relevant case for general relativity. A thorough exposition of Hamiltonian systems with second-class constraints and their relation to the Poisson bracket can be found in Sundermeyer's book. He also exhibits very clearly the relation between Lagrangian singular systems and their Hamiltonian formulation. I only used those bits relevant to understand the definition of gauge transformations in the canonical context.

Appendix B

The canonical quantization scheme

B.1 Canonical Quantization in Classical Mechanics

Quantization here shall be understood as a description of how to pass from a given classical system to a quantum theoretical one. The oldest attempt at a formalisation of such a procedure is given *by canonical quantization*. This method was mainly devised by Weyl, von Neumann and Dirac.

In the original form, canonical quantization makes use of a phase space $\mathbf{\Gamma} = \mathbb{R}^n \times \mathbb{R}^n$ with coordinates (p, q) where p, q are short-hands for all p_i, q_i and $i = 1, \ldots N$ labels the number of degrees of freedom. The task of quantization is then to find a map $\widehat{\ }$ from a suitable subset C of the space of functions $f(p, q)$ on phase space, $C \subseteq C^\infty(\mathbf{\Gamma})$, into the space of self-adjoint operators, $SELFADJ(\mathfrak{H})$, acting on the Hilbert space $\mathfrak{H} = L^2(\mathbb{R}^n, d^n q)$,

$$\widehat{\ }: \quad f \longmapsto Q_f \ ,$$

such that the following conditions hold:

(q1) The map

$$\begin{array}{rccl} \widehat{\ }: & C & \longrightarrow & SELFADJ(\mathfrak{H}), \\ & f & \longmapsto & Q_f = \hat{f} \end{array}$$

is linear.

(q2) The function with constant value one is mapped onto the identity element on the Hilbert space: $Q_1 = \mathbf{1}$.

(q3) Any function on phase space is realized as an operator on Hilbert space via the *von-Neumann rule*.

For any function $\phi : \mathbb{R} \to \mathbb{R}$ for which $Q_{\phi\circ f}$ and $\phi(Q_f)$ are well-defined, $Q_{\phi\circ f} = \phi(Q_f)$.

Note that this rule gives a prescription of how the multiplication of phase-space functions is realized on the space of operators. It thus requires the preservation of the associative structure on C given by multiplication of functions on phase space.

(q4) The quantization map is consistent with the Schrödinger representation, i.e.

$$Q(q^i)\psi(q) = q^i\psi(q) \ , \qquad Q(p_i)\psi(q) = -\mathrm{i}\hbar\frac{\partial\psi(q)}{\partial q_i} \ ,$$

for $\psi \in L^2(\mathbb{R}^n, d^n q)$.

(q5) A correspondence between the classical Poisson bracket and the quantum commutator exists in the following way

$$Q_{\{f,g\}} = \frac{1}{\mathrm{i}\hbar}\left[Q_f, Q_g\right] \ ,$$

for all quantizable observables $f, g \in C$.

The Stone–von Neumann theorem ensures that (q4) is the only irreducible representation — up to finite multiplicity — of the canonical variables p and q such that the canonical commutation relations hold as they follow from the Poisson bracket: $\left[\hat{q}_i, \hat{p}^j\right] = \mathrm{i}\hbar\,\delta_i^j$ and all others vanish.

The Lie algebra spanned by $(p, q, \mathbf{1})$ under the Poisson bracket is called *Heisenberg algebra*. One can therefore replace (q4) by:

(q4') The Heisenberg algebra has to be represented irreducibly (up to finite multiplication) on $\mathfrak{H}$.

The problem with this simple list is that any three out of the four conditions (q1), (q3), (q4) and (q5) are inconsistent, [17].[1] A famous example of such an inconsistency proof is given by Groenewold, later elaborated on by van Hove [21, 22]. Their statement is simply that (q5) cannot be satisfied whenever (q1) and (q4) are *and* the space of quantizable functions is given by *all* polynomials in (p, q) of degree smaller than four. This theorem makes no use of the von-Neumann rule. More importantly, it is part of the proof of the theorem to show that the multiplicative structure is partially conserved by the quantisation map — by virtue of the conditions (q1), (q2), (q4) and

[1] The second axiom (q2) is redundant, as it follows e.g. from (q5) with $f = q$ and $g = p$.

(q5) alone. More specifically, the von-Neumann rule is trivially satisfied for q^2, p^2 and pq, see also [18].

If any three out of four conditions are inconsistent, one condition obviously has to be dropped. Essentially, two ways have been chosen out of this predicament: One can drop the von-Neuman rule (q3). But, as was pointed out above, this does not suffice to circumvent the Groenewold–van Hove theorem. So further, one has to restrict the space of quantizable observables *suitably*. E.g. if C is restricted to the space of all polynomials of at most second order $C = C_{\mathrm{poly}(2)}$, the axiom (q5) can be satisfied. This holds also if C contains all polynomial functions at most linear *in* p but of arbitrary order in q, $C = C_{\mathrm{poly}(\infty,1)}$. Thus there is more than one suitable choice of quantizable variables.

Another possibility is to lessen (q5) and generalize the Poisson bracket–commutator relation.

Modern formulations of the canonical quantization programme try to implement these ideas and furthermore extend quantization to situations where phase space is given by an arbitrary symplectic manifold.

We will restrict the general outline of the canonical quantization programme to the case that the underlying phase space is a symplectic manifold of dimension $2n$ which at the same time is a cotangent bundle over some configuration space. We will denote the manifold by $\mathbf{\Gamma}$ and the symplectic form by ω, i.e. ω is a closed, non-degenerate two-form. Configuration space is given by $\mathcal{Q}$, the cotangent bundle consequently by $T^*\mathcal{Q}$. The dimensions are $dim\mathcal{Q} = n$, $dimT^*\mathcal{Q} = 2n$. Due to Darboux' theorem, the manifold looks locally like $\mathbb{R}^{2n}$. More precisely, we can equip $\mathbf{\Gamma}$ locally with coordinates (q^i, p_i), $i = 1, \ldots n$. In these coordinates, $\omega = \mathrm{d}q^i \wedge \mathrm{d}p_i$ and $\wedge$ denotes the wedge product. For two functions f and g on $\mathbf{\Gamma}$, the Poisson bracket is defined via

$$\{f, g\} = -\omega(X_f, X_g) \ ,$$

where X_f, X_g are the Hamiltonian vector fields of f and g, respectively. In the local coordinate chart, the Hamiltonian vector field of a function f is given by

$$X_f = \frac{\partial f}{\partial q^i}\frac{\partial}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial}{\partial q^i} \ .$$

And thus the above definition coincides with the usual Poisson bracket,

$$\{f, g\} = \frac{\partial f}{\partial q^i}\frac{\partial g}{\partial p_i} - \frac{\partial f}{\partial p_i}\frac{\partial g}{\partial q^i} \ .$$

We then require the quantization map $\widehat{\ }: \quad C \longrightarrow SELFADJ(\mathfrak{H})$, $f \longmapsto Q_f$ to satisfy the conditions (q1), (q2) and (q5). We require that (q4) holds in a local coordinate chart and in the case that $\mathbf{\Gamma} = \mathbb{R}^{2n}$. We drop the

von-Neumann rule and keep in mind that we have no general rule how to select the suitable subspace $C \subseteq C^\infty(\mathbf{\Gamma})$ which satisfies (q5).

There are several fixed structures in this procedure: first of all, operators are defined on a Hilbert space. Secondly, operators are self-adjoint. Thirdly, the representation of the fundamental variables is fixed. Each of these structures has been challenged in quantum gravity.

B.2 Canonical Quantization of Field Theories

So far, we gave a more or less axiomatic system to be realized for the quantization of a finite dimensional system. The whole procedure becomes more elaborate but not altered in principle if one generalizes these ideas further to field theory. This is known as the *functional Schrödinger representation* of quantum field theory.

The functional Schrödinger representation constitutes a generalization of the usual Fock space representation. The latter is restricted to situations where the particle interpretation applies. But the Schrödinger representation applies also when the particle interpretation does not exist — it can be generalized to arbitrarily curved spacetimes. But in fact, for a free scalar field theory in Minkowski space, it can be shown that both descriptions are equivalent.

In either case, *we have to specify a constant-time hypersurface.*

The relation between the Schrödinger and Fock representations is one way to carry over the quantization scheme of the previous section to a field theory, [10]. Another way is to discretize the hypersurface into cubes of volume δV_i where i numbers the cube, [35]. Then the field in one such volume is approximated by its value at the centre x_i of this cube:

$$q_{ri}(t) := \phi_r(x_i, t) = \phi_r(i, t) \ ,$$

where $r = 1, \ldots N$ labels the different degrees of freedom, different fields or components of a vector field, for instance. Replacing similarly derivatives by differences at neighbouring lattice points, one can rewrite the Lagrangian of the theory $\mathcal{L} = \mathcal{L}(\phi_r, \dot{\phi}_r, \phi_{r,a})$ in a discretized form as

$$\mathrm{L}(t) = \delta V_j \mathcal{L}_j(\phi_r(i,t), \dot{\phi}_r(i,t)\phi_r(i',t)) \ ,$$

where i' is the difference of fields at neighbouring lattice points and summation over j is assumed. The conjugate momentum is thus

$$p_{ri}(t) = \frac{\partial \mathrm{L}}{\partial \dot{q}_{ri}} = \pi_r(i,t)\delta V_i \ ,$$

where

$$\pi_r(i,t) = \frac{\partial \mathcal{L}_i}{\partial \dot{\phi}_r(i,t)} \ .$$

Thus actually the conjugate momentum is a density of weight one. We choose, to be consistent with the literature on canonical quantum gravity, the conjugate fields $\pi_r(i,t)$. In the limit of vanishing volume, $\delta V_i \to 0$, we obtain

$$\mathrm{H} = \delta V_j \left[\pi_r(j,t)\dot{\phi}_r(j,t) - \mathcal{L}_j\right] \quad \longrightarrow \quad \mathrm{H} = \int \mathrm{d}V\, \mathcal{H}(x) \ ,$$

where

$$\mathcal{H}(x) = \pi_r(x)\dot{\phi}_r(x) - \mathcal{L}(\phi_r(x,t), \dot{\phi}_r(x,t), \phi_{r,a}\,(x,t)) \ .$$

References:

I used the review article by Ali and Englis, [17], as well as Giulini's contribution to the book 'Quantum Gravity — From Theory to Experimental Search', [18] in the first section to give an axiomatic approach to canonical quantization. For the generalization of this quantization scheme to field theories, I used Mandl and Shaw's book on quantum field theory, [35].

Appendix C

'Singularity resolution': the Coulomb potential

The standard example for singularity resolution cited in the quantum general relativity literature is an electron moving in the attractive Coulomb potential of a nucleus. The claim is that here, the classical singularity consisting in the fall of the particle to the centre is resolved on the quantum level.

C.1 Singular potentials in quantum and classical mechanics

The Coulomb potential is more generally discussed in the study of singular potentials in classical and quantum physics. Let the potential be described by a single coordinate r which stands for the radial distance.

Then *singular potentials* in classical, *non-relativistic* mechanics are all attractive potentials which fall off faster than the inverse-square potential r^{-2} as the origin is approached. The inverse-square potential itself is a *transitional potential*, i.e. it exhibits singular features only for a certain range of coupling. In relativistic mechanics already the Coulomb potential is transitional, thus singular for certain parameter ranges. All potentials that fall off faster at the origin are *truly* singular.

Physically, such classically singular potentials are characterized by the existence of trajectories along which the particle spirals into the origin. The velocity diverges in this approach. And so does the scattering angle. Infinitely many trajectories go through the origin — all with the same tangent. Thus, in- and outgoing trajectories can only be matched if additional conditions are imposed, as, for example, energy and angular momentum conservation. It is the occurrence of an infinite curvature of the particle's trajectory which prevents a unique continuation of the trajectory through the origin.

Upon quantization, *singular potentials remain singular*. That means a problem such as the motion of a particle in an attractive inverse-square potential which exhibits singular features on the classical level, will do so as well on the quantum level. For singular potentials, *no singularity resolution takes place upon quantization*. In non-relativistic quantum mechanics, for the radial Schrödinger equation all potentials which fall off as fast as the inverse-square potential or faster, are problematic. In the relativistic case, the Klein–Gordon and Dirac equation exhibit singular features already for the Coulomb potential. All potentials that fall off faster than the Coulomb potential are truly singular potentials of these equations.

Physically, singular potentials on the quantum level are characterized by the fact that no unique solution to the quantum equation exists. For the Schrödinger equation with inverse-square potential, this implies for example that *both* solutions vanish at the origin — but differ by a phase factor (again, this holds for a certain parameter range). This phase factor cannot be determined and *mirrors* the non-uniqueness of outgoing trajectory for a given ingoing one that is found on the classical level. More generally, a boundary condition does not pick a unique solution for a quantum equation with singular potential. From this results an arbitrary phase in scattering amplitudes. Generally, the bound-state spectrum is not unique — there is an infinity of bound states. Lastly, the energy is generally unbounded from below.

Mathematically, these singular potentials correspond to strong singularities at $r = 0$ of the corresponding differential equation. Equivalently, one can say they stem from a Hamiltonian operator which has no unique self-adjoint extension to $r = 0$.

C.2 The classical motion of a particle in the attractive Coulomb potential

On the non-relativistic level, the motion of the electron is perfectly well behaved. As long as the particle has non-vanishing angular momentum, it does not fall to the centre. In the case of a vanishing angular momentum, the particle falls to the centre, but does so on a straight line. Because the total energy remains constant and finite, the kinetic energy and thus the velocity diverges as the origin is approached, thus compensating the infinitely negative potential energy. The equation of motion poses a well-defined initial value problem. The predictions inferred are unique (for a given set of initial conditions). The classical motion is *not* singular. The fact that the fall of the electron to the centre is *not in agreement with observation* is a different question and might at this level be justified by the argument that the state of exactly vanishing angular momentum might never be exactly

realized in Nature.[1]

In the relativistic case, however, the particle spirals into the centre whenever angular momentum becomes smaller than the coupling, i.e. the total charge. In this case, the scattering angle diverges and we encounter the problems listed above.[2]

C.3 The quantized Coulomb problem

Quantizing the motion in the Coulomb potential, one ends up with the radial Schrödinger equation and the additional condition, following from the self-adjointness of the Hamiltonian operator, that its solutions vanish at the origin and infinity sufficiently quickly.[3] This fixes the wave function uniquely. The radial Schrödinger equation thus poses a well-defined problem. The resulting wave function corresponding to zero angular momentum remains constant at the origin. The energy is unbounded from below. The quantum motion is *not* singular — as one would have expected from the quantization of a non-singular classical system. But it is now in (better) accord with observation: instead of a classical trajectory which leads straight into the nucleus, the electron is described by a wave function with zero probability density at the origin.

The case is, again, different in the relativistic setting. The Klein–Gordon equation with Coulomb potential is singular for certain angular-momentum eigenvalues. The boundary condition at the origin here does not pick a unique solution.

C.4 Conclusion

So I do not know why and how one can speak of singularity resolution in the case of the Coulomb potential. To the contrary: the Coulomb potential is an example where either *both*, quantum and classical motion, are well-defined or *both* pose an ill-defined problem. In fact, this is what is expected from the study of singular potentials.

What happens upon quantization of the Coulomb problem in the non-relativistic case is that the wrong prediction of the classical framework is substituted by a correct prediction through the quantum framework. The singularity occurring in the special relativistic description is due to the fact

[1] This resembles the point of view taken before the formulation of the singularity theorems in general relativity: The singularities occur only because we make idealized assumptions about Nature (like exact spherical symmetry, for example) which are not realized.

[2] The divergence in the relativistic case is due to the fact that the effective potential contains the square of the real potential — in contrast to the non-relativistic case were the potential enters the effective potential linearly.

[3] The condition is that $r\psi(r) = 0$ as $r \to 0$ and as $r \to \infty$, where r is the radial coordinate.

that in special relativity instantaneous interactions are not admissible. The Coulomb potential has to be substituted by a photon-mediated interaction. This, in a sense, is a consequence of the fact that the Klein–Gordon equation does not allow a consistent interpretation and has to be subjected to *second quantization*. In this way, one arrives at quantum field theory. The gauge principle than yields the correct description of the interaction. Thus the correct statement would be that the singularity of the Coulomb potential occurring in special relativity is cured by the replacement of an instantaneous interaction through the gauge principle *and* the field quantization of the resulting problem.

Another example where the classical *and* the quantum motion are singular is the radial Schrödiger equation with attractive inverse-square potential.

References: Singular potentials in quantum mechanics are discussed in the review by Frank and Land, [42]. Moreover, the singular inverse-square case is discussed in Landau and Lifshitz's book, [44]. The specific case of the Coulomb potential can be found in Newton's textbook, [45] and the article by Case, [46].

Appendix D

Approximating the Wheeler–DeWitt equation for the big-brake model

D.1 The Born–Oppenheimer approximation

The Born–Oppenheimer approximation is used in atomic physics to simplify the full Schrödinger equation comprising the full motion of all electrons and nuclei. It relies on the fact that nuclei have much larger mass than electrons. It can be carried over to the quantum general relativity case where the role of the nuclei is played by gravitation and the matter fields take the place of the electrons. I will give here a brief outline of the Born–Oppenheimer approximation in atomic physics as well as the conditions under which it applies.

I will use R to denote the coordinate of the nucleus and r to denote the electron location. By simply substituting the metric for R and the matter fields for r, one arrives at the quantum gravitational case. The parameter M used in the following is originally given by the mass of the nucleus and in quantum general relativity turns into (4.7). The time-independent Schrödinger equation is then of the form

$$\left[-\frac{\hbar^2}{2M}\Delta_R^2 + \hat{h}(r,R)\right]\Psi(r,R) = 0 \ , \tag{D.1}$$

where $\hat{h}(r,R)$ is the electron-part of the Hamiltonian and Δ_R is the Laplacian with respect to the nucleus' coordinate. One makes the ansatz

$$\Psi(r,R) = \sum_n C_n(R)\varphi_n(R,r) \ ,$$

where

$$\hat{h}(r,R)\varphi_n(R,r) = E_n(R)\varphi_n(R,r)$$
$$\text{and} \quad \int \mathrm{d}r\, \varphi_m^*(R,r)\varphi_n(R,r) = \delta_{mn} \ .$$

Upon inserting this equation in (D.1) and multiplying the resulting equation by $\varphi_n^*(R,r)$, integrating over r and using the orthogonality of the electron-eigenfunctions, one obtains

$$\left[-\frac{\hbar^2}{2M}\Delta_R^2 + E_n(R)\right] C_n(R)$$
$$- \quad \frac{\hbar^2}{2M}\sum_k \int \mathrm{d}r\, \varphi_n^*(R,r)\left[2\Delta_R C_m \Delta_R \varphi_m + C_m \Delta_R^2 \varphi_m\right] = 0 \ .$$

The Born–Oppenheimer approximation now consists in *neglecting the off-diagonal terms*. The Schrödinger equation then simplifies to

$$\left[-\frac{\hbar^2}{2M}\Delta_R^2 + E_n(R)\right] C_n(R)$$
$$- \quad \frac{\hbar^2}{2M}\int \mathrm{d}r\, \varphi_n^*(R,r)\left[2\Delta_R C_n \Delta_R \varphi_n + C_n \Delta_R^2 \varphi_n\right] = 0 \ .$$

When does this approximation apply? This can be seen from a rewriting of the neglected terms as

$$\int \mathrm{d}r\, \varphi_m^* \Delta_R \varphi_n \sim \langle\varphi_m|\hat{P}_R|\varphi_n\rangle = \frac{\langle\varphi_m|\left[\hat{P}_R, \hat{h}\right]|\varphi_n\rangle}{E_n(R) - E_m(R)} \ , \tag{D.2}$$

were P_R is the canonically conjugate momentum to R. This is thus negligible if the energy levels are well separated. This justifies the neglection of the *first* off-diagonal term. The second one is basically given by $\langle\varphi_m|\hat{P}_R^2|\varphi_n\rangle$ and the assumption is that $\langle\varphi_m|\hat{P}_R^2|\varphi_n\rangle \approx (\langle\varphi_m|\hat{P}_R|\varphi_n\rangle)^2$. Thus the spread in the momentum of the nucleus has to be small.

For a real-valued wave function, the diagonal terms vanish: the first one, $\langle\varphi_n|\hat{P}_R|\varphi_n\rangle$, due to the self-adjointness of $\hat{P}_R$, the second one again in the approximation $\langle\varphi_n|\hat{P}_R^2|\varphi_n\rangle \approx (\langle\varphi_n|\hat{P}_R|\varphi_n\rangle)^2$.

Note that the last two assumptions make some difficulty when applied to quantum general relativity: First of all, the wave functional does not have to be real and secondly, the operator corresponding to the kinetic energy of the gravitational field is generally not self-adjoint.

Moreover, for exotic matter potentials as will be used later in this work, see Chapters 7, 8, the matter Hamiltonian also depends on the gravitational coupling and thus on the Planck mass. The general assumption that gravity is 'heavier' than matter then cannot be made, as will be seen in the following.

D.2 Approximating the Wheeler–DeWitt equation

We start from the full Wheeler–DeWitt equation (8.10),

$$\frac{\hbar^2}{2}\left(\frac{\kappa^2}{6}\frac{\partial^2}{\partial\alpha^2}-\frac{\partial^2}{\partial\phi^2}\right)\psi(\alpha,\phi)-\frac{\tilde{V}_0}{|\phi|}e^{6\alpha}\psi(\alpha,\phi)=0\ ,$$

where $\tilde{V}_0 = V_0 a_{\mathrm{ref}}/3\kappa^2$. Into this equation, we insert the ansatz $\psi(\alpha,\phi) = \sum_n C_n(\alpha)\varphi_n(\alpha,\phi)$, where $\varphi_n(\alpha,\phi)$ is the solution of (8.11),

$$\hat{\mathrm{H}}_\phi\varphi_n(\alpha,\phi)=E_n(\alpha)\varphi_n(\alpha,\phi)\ ,$$

where $\hat{\mathrm{H}}_\phi = \frac{1}{2}\hat{p}_\phi^2 + \hat{a}^6 V(\hat{\phi})$. The eigenfunctions of the matter Hamiltonian $\hat{\mathrm{H}}_\phi$ are just the eigenfunctions of the radial Schrödinger equation for the Coulomb potential with vanishing orbital angular momentum. They are given by (8.12) and obey the orthogonality relation (8.13). Inserting the ansatz, the Wheeler–DeWitt equation turns into

$$\sum_n\left\{\frac{\hbar^2}{2}\frac{\kappa^2}{6}\left[\ddot{C}_n\varphi_n+2\dot{C}_n\dot{\varphi}_n+C_n\ddot{\varphi}_n\right]+E_nC_n\varphi_n\right\}=0\ ,$$

where dots are used to denote derivatives with respect to α. This equation can be simplified by multiplication with φ_m and integration over ϕ. This may not be very obvious at this point, but recall the procedure used in the Born–Oppenheimer approximation, cf. Section D.1. After using the orthogonality relation, the Wheeler–DeWitt equation becomes

$$\begin{aligned}&\frac{\hbar^2}{2}\frac{\kappa^2}{6}\left[\ddot{C}_m+E_mC_m\right]Z(\alpha)\\ +\ &\sum_n\frac{\hbar^2}{2}\frac{\kappa^2}{6}\left[2\dot{C}_n\langle\varphi_m|\dot{\varphi}_n\rangle+C_n\langle\varphi_m|\ddot{\varphi}_n\rangle\right]=0\ .\end{aligned}$$

In the following, Dirac brackets denote an inner product given by integration over ϕ which is suggested by the orthogonality relation (8.13).

I will now carry out an approximation in two steps. First, the cross terms $\langle\varphi_m|\dot{\varphi}_n\rangle$, $\langle\varphi_m|\ddot{\varphi}_n\rangle$ will be neglected in a Born–Oppenheimer approximation. In a second step, I will show that the remaining diagonal terms containing derivatives of φ_n with respect to α can also be neglected. In atomic physics, this follows trivially from the Born–Oppenheimer approximation and the fact that all operators are self-adjoint. Here, counting powers of the gravitational coupling κ^2 will be employed.

D.2.1 Born–Oppenheimer approximation of the Wheeler–DeWitt equation

We have to show that two terms are negligible: $\langle\varphi_m|\dot{\varphi}_n\rangle$ and $\langle\varphi_m|\ddot{\varphi}_n\rangle$.

I will start with the first term. Rewriting this term as

$$\begin{aligned}\langle\varphi_m|\dot{\varphi}_n\rangle &= \langle\varphi_m|\frac{\partial}{\partial\alpha}|\varphi_n\rangle = \langle\varphi_m|a\frac{\partial}{\partial a}|\varphi_n\rangle \\ &= -\frac{1}{\mathrm{i}\hbar}\langle\varphi_m|a\hat{p}_a|\varphi_n\rangle \ .\end{aligned}$$

The last equation can then be rewritten in the well-known way, cf. (D.2), as

$$\langle\varphi_m|a\hat{p}_a|\varphi_n\rangle = a\frac{\langle\varphi_m|\left[\hat{p}_a,\hat{H}_\phi\right]|\varphi_n\rangle}{E_n - E_m} \ ,$$

where square brackets denote commutators. Thus, we obtain

$$\langle\varphi_m|\dot{\varphi}_n\rangle = -\frac{6\mathrm{i}\hbar a^6}{E_n - E_m}\langle\varphi_m|V(\hat{\phi})|\varphi_n\rangle \ .$$

If the expectation value of the potential remains finite, the term on the left hand side is negligible, if the energy eigenvalues are well separated. Using the energy eigenvalues, (8.14), this reads

$$\langle\varphi_m|\dot{\varphi}_n\rangle = -\frac{6\mathrm{i}\hbar^3}{V_\alpha}\frac{1}{\left(\frac{1}{n^2}-\frac{1}{m^2}\right)}\langle\varphi_m|V(\hat{\phi})|\varphi_n\rangle \ .$$

So if $\langle\varphi_m|V(\hat{\phi})|\varphi_n\rangle$ remains finite, the approximation applies. The expectation value of the potentials yields an integral of the form

$$\langle\varphi_m|V(\hat{\phi})|\varphi_n\rangle = \tilde{V}_0\mu\nu\int \mathrm{d}\phi\phi e^{-\left(\frac{\mu+\nu}{2}\right)\phi}\mathrm{L}^1_{n-1}(\mu\phi)\mathrm{L}^1_{m-1}(\nu\phi) \ ,$$

where $\mu = \frac{2V_\alpha}{\hbar^2 m}$ and similarly $\nu = \frac{2V_\alpha}{\hbar^2 n}$. This can be integrated with the help of

$$\begin{aligned}&\int_0^\infty \mathrm{d}x e^{-bx}x^\beta L_l^\beta(\nu x)L_j^\beta(\mu x) = \frac{\Gamma\left(j+l+\beta+1\right)}{j!l!}\frac{\left(b-\nu\right)^l\left(b-\mu\right)^j}{b^{j+l+\beta+1}} \\ &\times F\left(-j,-l;-j-l-\beta;\frac{b\left(b-\nu-\mu\right)}{\left(b-\nu\right)\left(b-\mu\right)}\right) \ ,\end{aligned}$$

where $F(a,b;c;x)$ denotes the hypergeometric function, [78]. This yields

$$\begin{aligned}\langle\varphi_m|V(\hat{\phi})|\varphi_n\rangle &= \frac{\tilde{V}_0}{K_{m,n}}\frac{\Gamma(m+n)}{(m-1)!(n-1)!} \\ &\times F\left(-m+1,-n+1;-m-n+1;\left(\frac{n+m}{n-m}\right)^2\right) \ ,\end{aligned}$$

Recall that $n \neq m$. This is obviously finite.

We therefore conclude that the term $\langle\varphi_m|\dot{\varphi}_n\rangle$ can be safely neglected due to the fact that the spectrum of the matter Hamiltonian is discrete and the energy levels are well separated.

The second term, $\langle\varphi_m|\ddot{\varphi}_n\rangle$ can basically be rewritten as $\langle\varphi_m|\hat{p}_a^2|\varphi_n\rangle$.[1] The assumption is that

$$\langle\varphi_m|\hat{p}_a^2|\varphi_n\rangle \approx \langle\varphi_m|\hat{p}_a|\varphi_n\rangle^2$$

and, as the right-hand side vanishes, so does the left-hand side. In this way, we get rid of the off-diagonal terms. This is analogous to the Born–Oppenheimer approximation as it was presented in Section D.1.

With these approximations, the Wheeler–DeWitt equation simplifies to

$$\begin{aligned} & \frac{\hbar^2}{2}\frac{\kappa^2}{6}\left[\ddot{C}_m + E_m C_m\right] Z(\alpha) \\ + \quad & \frac{\hbar^2}{2}\frac{\kappa^2}{6}\left[2\dot{C}_m\langle\varphi_m|\dot{\varphi}_m\rangle + C_m\langle\varphi_m|\ddot{\varphi}_m\rangle\right] = 0 \ . \end{aligned}$$

D.2.2 Neglection of matter motion

Whereas in atomic physics, the diagonal terms, $\langle\varphi_m|\dot{\varphi}_m\rangle$, $\langle\varphi_m|\ddot{\varphi}_m\rangle$, vanish trivially, here some work is needed to show that they can actually be neglected. We thus have to show that the term $\ddot{C}_m Z(\alpha) = \ddot{C}_m\langle\varphi_m|\varphi_m\rangle$ dominates over $\dot{C}_m\langle\varphi_m|\dot{\varphi}_m\rangle$ and $C_m\langle\varphi_m|\ddot{\varphi}_m\rangle$. This is rather easily seen. First of all note that the derivatives of $C_m(\alpha) = c_2 \mathrm{K}_0\left(\frac{1}{\sqrt{6}}\frac{V_\alpha}{\hbar^2 m\kappa}\right)$ with respect to α, cf. (8.16) satisfy

$$C_m \sim C_m \ , \qquad \dot{C}_m \sim V_\alpha \ , \qquad \ddot{C}_m \sim V_\alpha^2 \ .$$

On the other hand, as a consequence of the orthogonality relation, (8.13),

$$\langle\varphi_m|\varphi_m\rangle \sim \langle\varphi_m|\dot{\varphi}_m\rangle \sim \langle\varphi_m|\ddot{\varphi}_m\rangle \sim \frac{1}{V_\alpha} \ .$$

Therefore, the different terms behave as

$$\begin{aligned} \ddot{C}_m\langle\varphi_m|\varphi_m\rangle &\sim V_\alpha \sim \frac{e^{6\alpha}}{\kappa^2} \\ \dot{C}_m\langle\varphi_m|\dot{\varphi}_m\rangle &\sim 1 \\ C_m\langle\varphi_m|\ddot{\varphi}_m\rangle &\sim \frac{1}{V_\alpha} \sim \frac{\kappa^2}{e^{6\alpha}} \ . \end{aligned}$$

[1] It also gives one contribution proportional to $\langle\hat{p}_a\rangle$ which can be negelcted due to the arguments given above.

Recall that V_α depends on κ,

$$V_\alpha = \tilde{V}_0 e^{6\alpha} = \frac{a_{\mathrm{ref}} V_0}{3\kappa^2} \ .$$

There are thus two arguments that justify the neglection of the terms containing derivatives of φ_m. First of all, one can use an expansion in terms of the Planck mass. This is analogous to the mass parameter $M = \frac{1}{4\kappa^2}$ introduced in the derivation of the approximate functional Schrödinger equation, cf. Section 4.4.1, (4.7). Then the $\ddot{C}_m$ term gives the highest order contribution.[2] Moreover, in the vicinity of the big-brake singularity the scale factor is large. This is another reason why we can safely neglect the derivative terms.[3]

Thus, we finally end up with the approximated Wheeler–DeWitt equation (8.15),

$$\ddot{C}_n(\alpha) - \frac{6{V_\alpha}^2}{\hbar^4 n^2 \kappa^2} C_n(\alpha) = 0 \ .$$

[2]Recall that $E_m \sim V_\alpha^2$.

[3]Recall that, anyway, the potential used here is a good approximation to the full potential only in that regime.

Appendix E

System of constraints for the parametrized free particle

General expression for the constraints are

$$C^{(n)} = \sum_{m=0}^{n}\sum_{j=0}^{m}\sum_{\ell=0}^{2(n-m)}\binom{n}{m}\binom{m}{j}\binom{2(n-m)}{\ell}p_t^{m-j}\frac{p^{2(n-m)-\ell}}{(2M)^{n-m}}G_{j,0}^{\ell,0}\ , \tag{E.1}$$

$$\begin{aligned} C_q^{(n)} &= \sum_{m=0}^{n}\sum_{j=0}^{m}\sum_{\ell=0}^{2(n-m)}\binom{n}{m}\binom{m}{j}\binom{2(n-m)}{\ell}p_t^{m-j}\frac{p^{2(n-m)-\ell}}{(2M)^{n-m}} \\ &\quad\times\left(qG_{j,0}^{\ell,0}+G_{j,0}^{\ell,1}+\frac{\mathrm{i}\hbar}{2}\ell G_{j,0}^{\ell-1,0}\right)\ , \end{aligned}$$

$$\begin{aligned} C_t^{(n)} &= \sum_{m=0}^{n}\sum_{j=0}^{m}\sum_{\ell=0}^{2(n-m)}\binom{n}{m}\binom{m}{j}\binom{2(n-m)}{\ell}p_t^{m-j}\frac{p^{2(n-m)-\ell}}{(2M)^{n-m}} \\ &\quad\times\left(tG_{j,0}^{\ell,0}+G_{j,1}^{\ell,0}+\frac{\mathrm{i}\hbar}{2}jG_{j-1,0}^{\ell,0}\right)\ , \end{aligned}$$

$$\begin{aligned} C_{p_t}^{(n)} &= \sum_{m=0}^{n}\sum_{j=0}^{m}\sum_{\ell=0}^{2(n-m)}\binom{n}{m}\binom{m}{j}\binom{2(n-m)}{\ell}p_t^{m-j}\frac{p^{2(n-m)-\ell}}{(2M)^{n-m}} \\ &\quad\times\left(p_tG_{j,0}^{\ell,0}+G_{j+1,0}^{\ell,0}\right)\ , \end{aligned} \tag{E.2}$$

$$\begin{aligned} C_{p^k}^{(n)} &= \sum_{m=0}^{n}\sum_{j=0}^{m}\sum_{\ell=0}^{2(n-m)}\sum_{r=0}^{k}\binom{n}{m}\binom{m}{j}\binom{2(n-m)}{\ell}\binom{k}{r}p_t^{m-j}\frac{p^{2(n-m)+k-\ell-r}}{(2M)^{n-m}} \\ &\quad\times G_{j,0}^{\ell+r,0}\ , \end{aligned} \tag{E.3}$$

$$C^{(n)}_{tp^k} = \sum_{m=0}^{n}\sum_{j=0}^{m}\sum_{\ell=0}^{2(n-m)}\sum_{r=0}^{k}\binom{n}{m}\binom{m}{j}\binom{2(n-m)}{\ell}\binom{k}{r}p_t^{m-j}\frac{p^{2(n-m)+k-\ell-r}}{(2M)^{n-m}} \times\left(tG^{\ell+r,0}_{j,0}+G^{\ell+r,0}_{j,1}+\frac{\mathrm{i}\hbar}{2}jG^{\ell+r,0}_{j-1,0}\right)\ , \tag{E.4}$$

$$C^{(n)}_{qp^k} = \sum_{m=0}^{n}\sum_{j=0}^{m}\sum_{\ell=0}^{2(n-m)}\sum_{r=0}^{k}\binom{n}{m}\binom{m}{j}\binom{2(n-m)}{\ell}\binom{k}{r}p_t^{m-j}\frac{p^{2(n-m)+k-\ell-r}}{(2M)^{n-m}} \times\left(qG^{\ell+r,0}_{j,0}+G^{\ell+r,1}_{j,0}+\frac{\mathrm{i}\hbar}{2}(\ell+r)G^{\ell+r-1,0}_{j,0}\right)\ , \tag{E.5}$$

$$C^{(n)}_{p_tp^k} = \sum_{m=0}^{n}\sum_{j=0}^{m}\sum_{\ell=0}^{2(n-m)}\sum_{r=0}^{k}\binom{n}{m}\binom{m}{j}\binom{2(n-m)}{\ell}\binom{k}{r}p_t^{m-j}\frac{p^{2(n-m)+k-\ell-r}}{(2M)^{n-m}} \times\left(p_tG^{\ell+r,0}_{j,0}+G^{\ell+r,0}_{j+1,0}\right)\ . \tag{E.6}$$

In addition to those written explicitly here, there are those involving higher polynomials also in q, t and p_t. The first two types of those constraints are more lengthy due to reorderings in the quantum variables. The constraints listed suffice for considerations in Chapter 9.

In a moment expansion, the leading terms of these constraints are

$$
\begin{aligned}
C^{(n)} &= C^n_{\rm class} + nC^{n-1}_{\rm class}\frac{1}{2M}G^{2,0}_{0,0} \\
&\quad +n(n-1)C^{n-2}_{\rm class}\left[\frac{p^2}{2M^2}G^{2,0}_{0,0}+\frac{p}{M}G^{1,0}_{1,0}+\frac{1}{2}G^{0,0}_{2,0}+\frac{1}{2M}G^{2,0}_{1,0}\right. \\
&\qquad\qquad \left. +\frac{p}{2M^2}G^{3,0}_{0,0}+\frac{1}{8M^2}G^{4,0}_{0,0}\right] \\
&\quad +n(n-1)(n-2)C^{n-3}_{\rm class}\left[\frac{p^2}{2M^2}G^{2,0}_{1,0}+\frac{p}{2M}G^{1,0}_{2,0}+\frac{1}{6}G^{0,0}_{3,0}+\frac{p^3}{6M^3}G^{3,0}_{0,0}+X_1\right] \\
&\quad +n(n-1)(n-2)(n-3)R_1 = 0\ , \qquad \text{(E.7)} \\
C^{(n)}_q &= qC^{(n)} + nC^{n-1}_{\rm class}\left[\frac{p}{M}\frac{\mathrm{i}\hbar}{2}+\frac{p}{M}G^{1,1}_{0,0}+G^{0,1}_{1,0}+\frac{1}{2M}G^{2,1}_{0,0}\right] \\
&\quad +n(n-1)C^{n-2}_{\rm class}\left[\frac{\mathrm{i}\hbar}{2}\frac{1}{M}\left(G^{1,0}_{1,0}+\frac{3p}{2M}G^{2,0}_{0,0}\right)+\frac{p^2}{2M^2}G^{2,1}_{0,0}+\frac{p}{M}G^{1,1}_{1,0}\right. \\
&\qquad\qquad \left. +\frac{1}{2}G^{0,1}_{2,0}+\frac{\mathrm{i}\hbar}{2}\frac{1}{2M^2}G^{3,0}_{0,0}+X_2\right] \\
&\quad +n(n-1)(n-2)C^{n-3}_{\rm class}\left[\frac{\mathrm{i}\hbar}{2}\left(\frac{p^2}{M^2}G^{1,0}_{1,0}+\frac{p^3}{2M^3}G^{2,0}_{0,0}\right.\right. \\
&\qquad\qquad \left.\left. +\frac{p}{2M}G^{0,0}_{2,0}+\frac{3p}{2M^2}G^{2,0}_{1,0}+\frac{p^2}{M^3}G^{3,0}_{0,0}+\frac{1}{2M}G^{1,0}_{2,0}\right)+X_3\right] \\
&\quad +n(n-1)(n-2)(n-3)C^{n-4}_{\rm class}\left[\frac{\mathrm{i}\hbar}{2}\left(\frac{p^4}{6M^4}G^{3,0}_{0,0}+\frac{p^3}{2M^3}G^{2,0}_{1,0}\right.\right. \\
&\qquad\qquad \left.\left. +\frac{p^2}{2M^2}G^{1,0}_{2,0}+\frac{p}{6M}G^{0,0}_{3,0}\right)+X_4\right] \\
&\quad +n(n-1)(n-2)(n-3)(n-4)R_2 = 0\ , \qquad \text{(E.8)} \\
C^{(n)}_t &= tC^{(n)} + nC^{n-1}_{\rm class}\left[\frac{\mathrm{i}\hbar}{2}+\frac{p}{M}G^{1,0}_{0,1}+G^{0,0}_{1,1}+\frac{1}{2M}G^{2,0}_{0,1}\right] \\
&\quad +n(n-1)C^{n-2}_{\rm class}\left[\frac{\mathrm{i}\hbar}{2}\frac{1}{2M}G^{2,0}_{0,0}+\frac{p^2}{2M^2}G^{2,0}_{0,1}+\frac{p}{M}G^{1,0}_{1,1}+\frac{1}{2}G^{0,0}_{2,1}+X_5\right] \\
&\quad +n(n-1)(n-2)C^{n-3}_{\rm class}\left[\frac{\mathrm{i}\hbar}{2}\left(\frac{p}{M}G^{1,0}_{1,0}+\frac{1}{2}G^{0,0}_{2,0}+\frac{p^2}{2M^2}G^{2,0}_{0,0}+\frac{p}{2M^2}G^{3,0}_{0,0}\right.\right. \\
&\qquad\qquad \left.\left. +\frac{1}{2M}G^{2,0}_{1,0}\right)+X_6\right] \\
&\quad +n(n-1)(n-2)(n-3)C^{n-4}_{\rm class}\left[\frac{\mathrm{i}\hbar}{2}\left(\frac{p^3}{6M^3}G^{3,0}_{0,0}+\frac{p^2}{2M^2}G^{2,0}_{1,0}+\frac{p}{2M}G^{1,0}_{2,0}\right.\right. \\
&\qquad\qquad \left.\left. +\frac{1}{6}G^{0,0}_{3,0}\right)+X_7\right] \\
&\quad +n(n-1)(n-2)(n-3)(n-4)R_3 = 0\ , \qquad \text{(E.9)}
\end{aligned}
$$

$$
\begin{aligned}
C^{(n)}_{p_t} &= p_t C^{(n)} + nC^{n-1}_{\text{class}}\left[\frac{p}{M}G^{1,0}_{1,0} + G^{0,0}_{2,0} + \frac{1}{2M}G^{2,0}_{1,0}\right] \\
&\quad + n(n-1)C^{n-2}_{\text{class}}\left[\frac{p^2}{2M^2}G^{2,0}_{1,0} + \frac{p}{M}G^{1,0}_{2,0} + \frac{1}{2}G^{0,0}_{3,0} + X_8\right] \\
&\quad + n(n-1)(n-2)R_4 = 0 \ , \qquad \text{(E.10)}
\end{aligned}
$$

$$
\begin{aligned}
C^{(n)}_{p} &= pC^{(n)} + nC^{n-1}_{\text{class}}\left[G^{1,0}_{1,0} + \frac{1}{2M}G^{3,0}_{0,0} + \frac{p}{M}G^{2,0}_{0,0}\right] \\
&\quad + n(n-1)C^{n-2}_{\text{class}}\left[\frac{p^2}{2M^2}G^{3,0}_{0,0} + \frac{p}{M}G^{2,0}_{1,0} + \frac{1}{2}G^{1,0}_{2,0} + X_9\right] \\
&\quad + n(n-1)(n-2)R_5 = 0 \ , \qquad \text{(E.11)} \\
C^{(n)}_{p^2} &= 2pC^{(n)}_{p} - pC^{(n)} + C^{n}_{\text{class}}G^{2,0}_{0,0} + nC^{n-1}_{\text{class}}\left[\frac{p}{M}G^{3,0}_{0,0} + G^{2,0}_{1,0} + \frac{1}{2M}G^{4,0}_{0,0}\right] \\
&\quad + n(n-1)R_6 = 0 \ , \qquad \text{(E.12)}
\end{aligned}
$$

$$
\begin{aligned}
C^{(n)}_{tp} &= tC^{(n)}_{p} + pC^{(n)}_{t} + C^{n}G^{1,0}_{0,1} + nC^{n-1}_{\text{class}}\left[\frac{p}{M}G^{2,0}_{0,1} + \frac{1}{2M}G^{3,0}_{0,1} + G^{1,0}_{1,1}\right] \\
&\quad + n(n-1)C^{n-2}_{\text{class}}\left[\frac{\mathrm{i}\hbar}{2}\left(\frac{p}{M}G^{2,0}_{0,0} + G^{1,0}_{1,0} + \frac{1}{2M}G^{3,0}_{0,0}\right) + X_{10}\right] \\
&\quad + n(n-1)(n-2)C^{n-3}_{\text{class}}\left[\frac{\mathrm{i}\hbar}{2}\left(\frac{p^2}{2M^2}G^{3,0}_{0,0} + \frac{1}{2}G^{1,0}_{2,0} + \frac{p}{M}G^{2,0}_{1,0}\right) + X_{11}\right] \\
&\quad + n(n-1)(n-2)(n-3)R_7 = 0 \ , \qquad \text{(E.13)}
\end{aligned}
$$

$$
\begin{aligned}
C^{(n)}_{qp} &= qC^{(n)}_{p} + pC^{(n)}_{q} + C^{n}_{\text{class}}\left[G^{1,1}_{0,0} + \frac{\mathrm{i}\hbar}{2}\right] \\
&\quad + nC^{n-1}_{\text{class}}\left[3\frac{\mathrm{i}\hbar}{2}\frac{1}{2M}G^{2,0}_{0,0} + \frac{p}{M}G^{2,1}_{0,0} + G^{1,1}_{1,0} + \frac{1}{2M}G^{3,1}_{0,0}\right] \\
&\quad + n(n-1)C^{n-2}_{\text{class}}\left[\frac{\mathrm{i}\hbar}{2}\left(\frac{3p^2}{2M^2}G^{2,0}_{0,0} + \frac{2p}{M}G^{1,0}_{1,0} + \frac{1}{2}G^{0,0}_{2,0} + \frac{2p}{M^2}G^{3,0}_{0,0}\right.\right. \\
&\qquad\qquad \left.\left. + \frac{3}{2M}G^{2,0}_{1,0}\right) + X_{12}\right] \\
&\quad + n(n-1)(n-2)C^{n-3}_{\text{class}}\left[\frac{\mathrm{i}\hbar}{2}\left(\frac{3p^2}{2M^2}G^{2,0}_{1,0} + \frac{2p^3}{3M^3}G^{3,0}_{0,0} + \frac{p}{M}G^{1,0}_{2,0}\right.\right. \\
&\qquad\qquad \left.\left. + \frac{1}{6}G^{0,0}_{3,0}\right) + X_{13}\right] \\
&\quad + n(n-1)(n-2)(n-3)R_8 = 0 \ , \qquad \text{(E.14)}
\end{aligned}
$$

$$
\begin{aligned}
C^{(n)}_{p_t p} &= p_t C^{(n)}_{p} + pC^{(n)}_{p_t} + C^{n}_{\text{class}}G^{1,0}_{1,0} + nC^{n-1}_{\text{class}}\left[\frac{p}{M}G^{2,0}_{1,0} + G^{1,0}_{2,0} + \frac{1}{2M}G^{3,0}_{1,0}\right] \\
&\quad + n(n-1)R_9 = 0 \ , \qquad \text{(E.15)}
\end{aligned}
$$

$$C^{(n)}_{p^3} = 3pC^{(n)}_{p^2} - 3p^2C^{(n)}_{p} + p^3C^{(n)} + C^n_{\rm class}G^{3,0}_{0,0} + nC^{n-1}_{\rm class}X_{14} + n(n-1)R_{10} = 0 \ , \quad \text{(E.16)}$$

$$C^{(n)}_{tp^2} = tC^{(n)}_{p^2} - p^2C^{(n)}_{t} + 2pC^{(n)}_{tp} - 2ptC^{(n)}_{p} + C^n_{\rm class}G^{2,0}_{0,1} + nC^{n-1}_{\rm class}\left[\frac{\mathrm{i}\hbar}{2}G^{2,0}_{0,0} + X_{15}\right] + n(n-1)C^{n-2}_{\rm class}\left[\frac{\mathrm{i}\hbar}{2}\left(\frac{p}{M}G^{3,0}_{0,0} + G^{2,0}_{1,0}\right) + X_{16}\right] + n(n-1)(n-2)R_{11} = 0, \quad \text{(E.17)}$$

$$C^{(n)}_{qp^2} = qC^{(n)}_{p^2} - p^2C^{(n)}_{q} + 2pC^{(n)}_{qp} - 2pqC^{(n)}_{p} + C^n_{\rm class}G^{2,1}_{0,0} + nC^{n-1}_{\rm class}\left[\frac{\mathrm{i}\hbar}{2}\left(3\frac{p}{M}G^{2,0}_{0,0} + 2G^{1,0}_{1,0} + 4\frac{1}{2M}G^{3,0}_{0,0}\right) + X_{17}\right] + n(n-1)C^{n-2}_{\rm class}\left[\frac{\mathrm{i}\hbar}{2}\left(\frac{2p^2}{M^2}G^{3,0}_{0,0} + 3\frac{p}{M}G^{2,0}_{1,0} + G^{1,0}_{2,0}\right) + X_{18}\right] + n(n-1)(n-2)R_{12} = 0 \ , \quad \text{(E.18)}$$

$$C^{(n)}_{p_tp^2} = p_tC^{(n)}_{p^2} - p^2C^{(n)}_{p_t} + 2pC^{(n)}_{p_tp} - 2pp_tC^{(n)}_{p} + C^n_{\rm class}G^{2,0}_{1,0} + nC^{n-1}_{\rm class}X_{19} + n(n-1)R_{13} = 0 \ , \quad \text{(E.19)}$$

$$C^{(n)}_{tp^3} = tC^{(n)}_{p^3} + p^3C^{(n)}_{t} - 3p^2C^{(n)}_{tp} + 3^2ptC^{(n)}_{p} + 3pC^{(n)}_{tp^2} - 3ptC^{(n)}_{p^2} + C^n_{\rm class}G^{3,0}_{0,1} + nC^{n-1}_{\rm class}\left[\frac{\mathrm{i}\hbar}{2}G^{3,0}_{0,0} + X_{20}\right] + n(n-1)R_{14} = 0 \ , \quad \text{(E.20)}$$

$$C^{(n)}_{qp^3} = qC^{(n)}_{p^3} + p^3C^{(n)}_{q} - 3p^2C^{(n)}_{qp} + 3^2pqC^{(n)}_{p} + 3pC^{(n)}_{qp^2} - 3pqC^{(n)}_{p^2} + C^n_{\rm class}\left[G^{3,1}_{0,0} + 3\frac{\mathrm{i}\hbar}{2}G^{2,0}_{0,0}\right] + nC^{n-1}_{\rm class}\left[\frac{\mathrm{i}\hbar}{2}\left(4\frac{p}{M}G^{3,0}_{0,0} + 3G^{2,0}_{1,0}\right) + X_{21}\right] + n(n-1)R_{15} = 0 \ , \quad \text{(E.21)}$$

where X_i and R_i are linear functions of higher, i.e. at least fourth, order moments.

Appendix F

Figures

F.1 Wave packet for the big-brake model

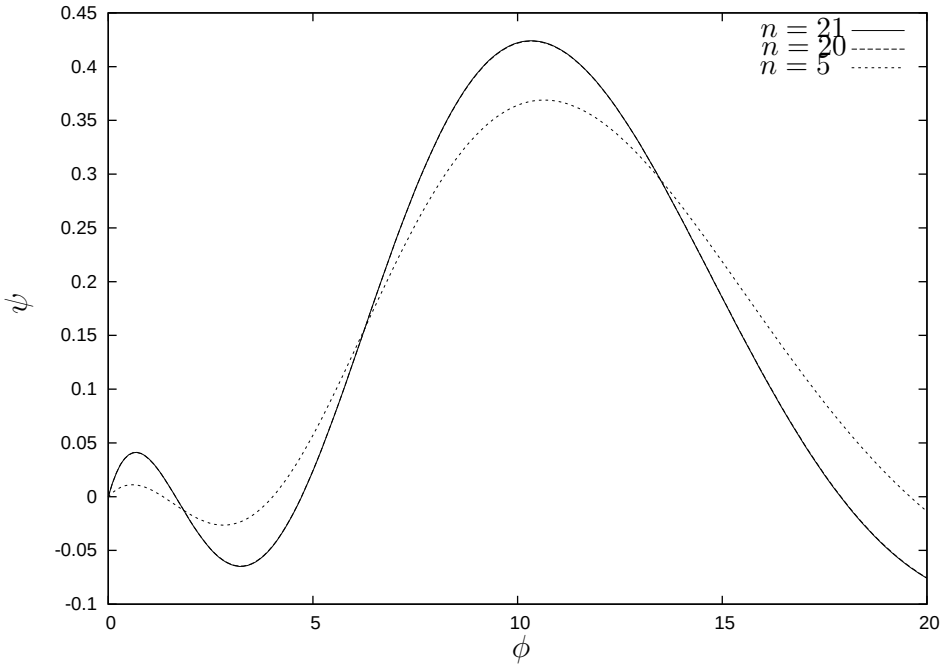

Figure F.1: This figure shows the wave packet for the big-brake model at $\tau = 1$. Here, the sum in equation (8.24) is carried out up to different n. It can be seen that it is sufficient to sum the series up to $n = 20$. Including more terms does not alter the form of the wave packet: The packet obtained for $n = 21$ is exactly of the same shape as the one obtained from summation up to $n = 20$.

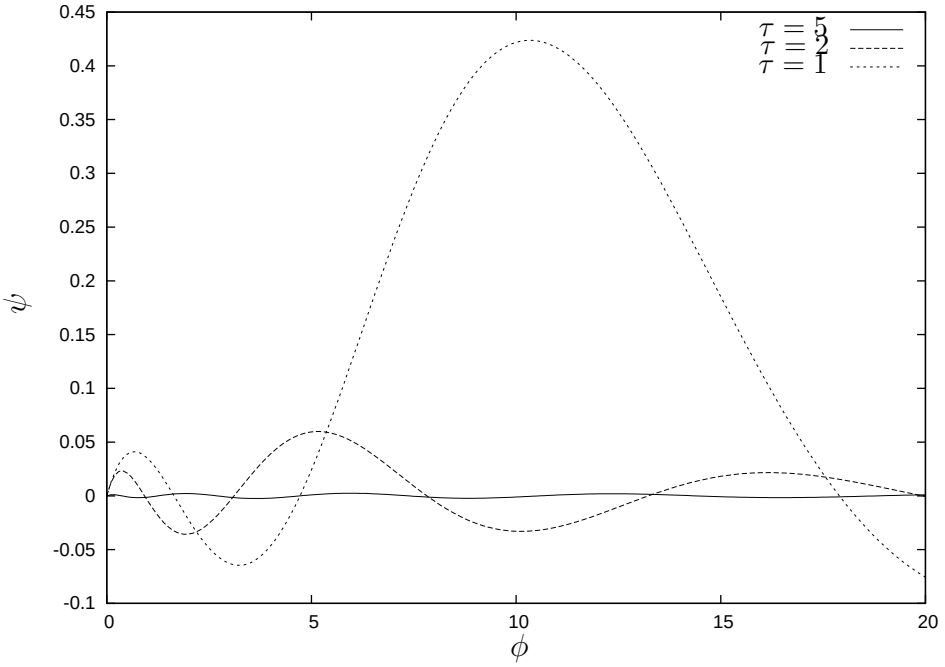

Figure F.2: This plot shows how the wave packet for the big-brake model looses its form. The different lines show the packet at different τ. Summation is carried out up to $n = 20$. The initially Gaussian packet at $\tau = 1$ disperses as $\tau = 2$ and at $\tau = 5$ is already a mere oscillation around zero. The classical singularity occurs at $\tau \approx 7$.

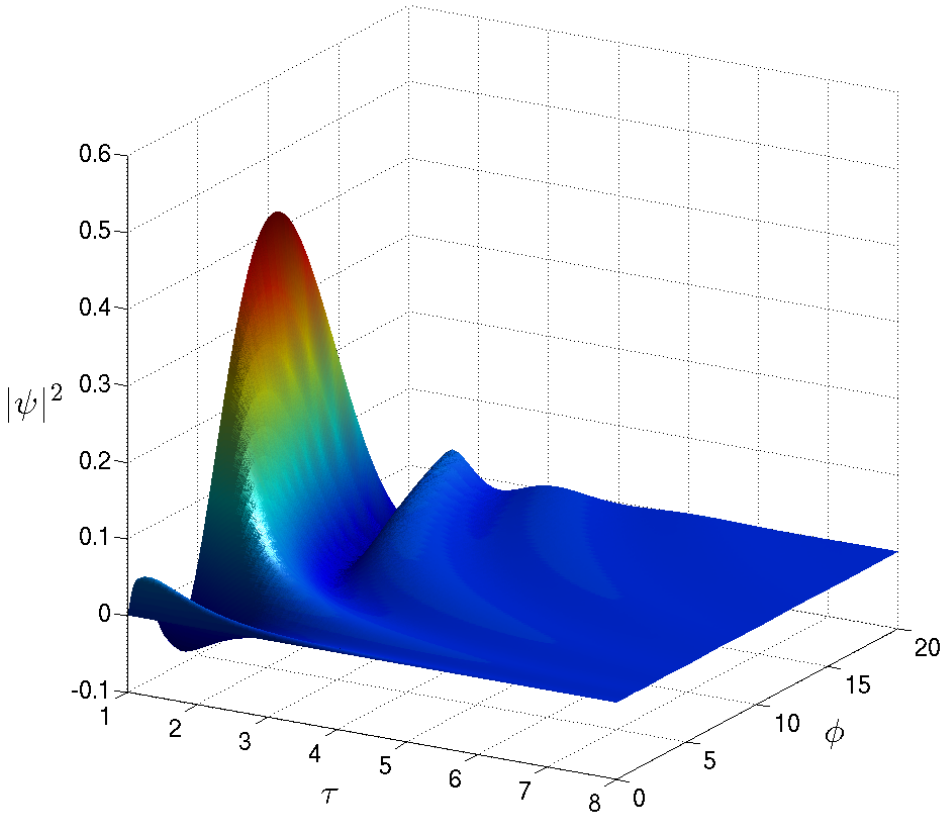

Figure F.3: The figure shows the wave packet for the big-brake model. One clearly sees that the initially peaked packet disperses quickly. The classical singularity occurs at $\tau \approx 7$, $\phi = 0$ — where the wave packet is zero anyway. The centre of the Gaussian at $\tau = 1$ is chosen to be $\phi_0 = 10$. Units are determined such that $\hbar = 1$ and $\kappa = 0.05$. As parameter $V_0 = 1$ was chosen.

F.2 Wave packet for the phantom- and scalar-field models

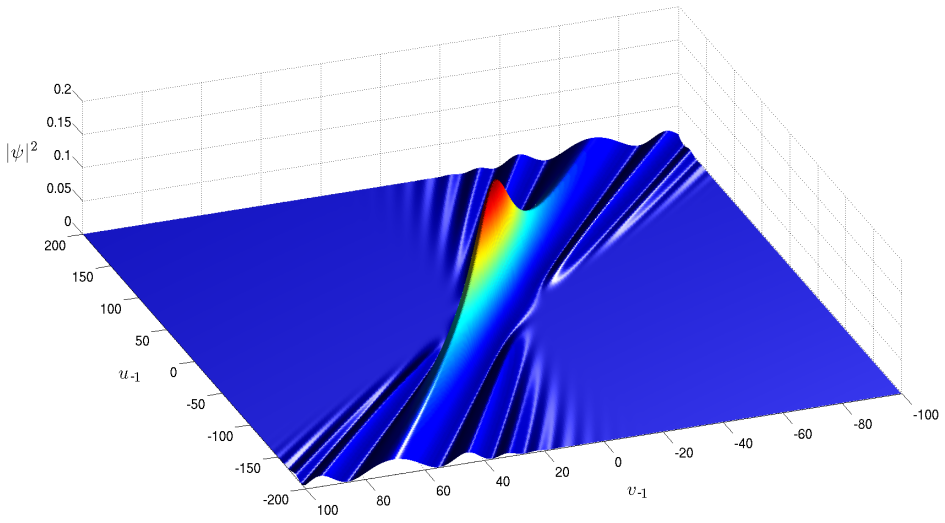

Figure F.4: This figure shows the wave packet for the phantom model with exponential potential. The packet is displayed in the (u_{-1}, v_{-1})-plane. It is clearly peaked around the classical trajectory (7.29) but spreads as $|u_{-1}|$, $|v_{-1}|$ become large, which is just the region of the classical big-rip singularity. In this plot, $\lambda = \sqrt{3}$ was chosen. The gravitational constant has units such that $\kappa^2 = 6$.

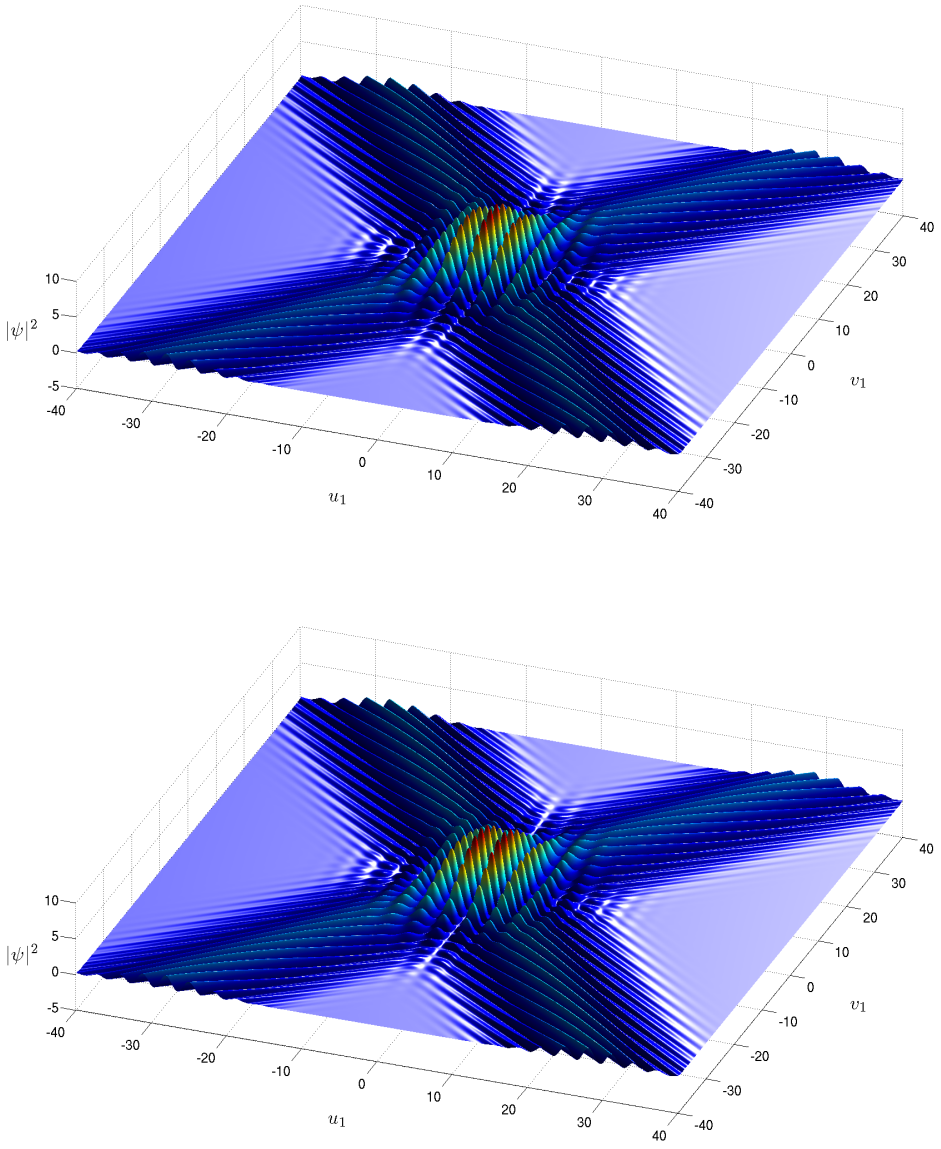

Figure F.5: These two figures exhibit the wave packet for the scalar field model with exponential potential, Chapter 7. In the upper plot, the initial conditions are chosen such that the two solutions add up at the origin of the (u_1, v_1)-plane. For wave packet shown below, the initial conditions have been chosen such that the packet vanishes on $u_1 = 0$. Recall that $u_1 > 0$ and one quarter of the (u_1, v_1)-plane corresponds to the entire (α, ϕ)-plane. Parameters are chosen as in the corresponding phantom model, $\lambda = \sqrt{3}$ and $\kappa^2 = 6$.

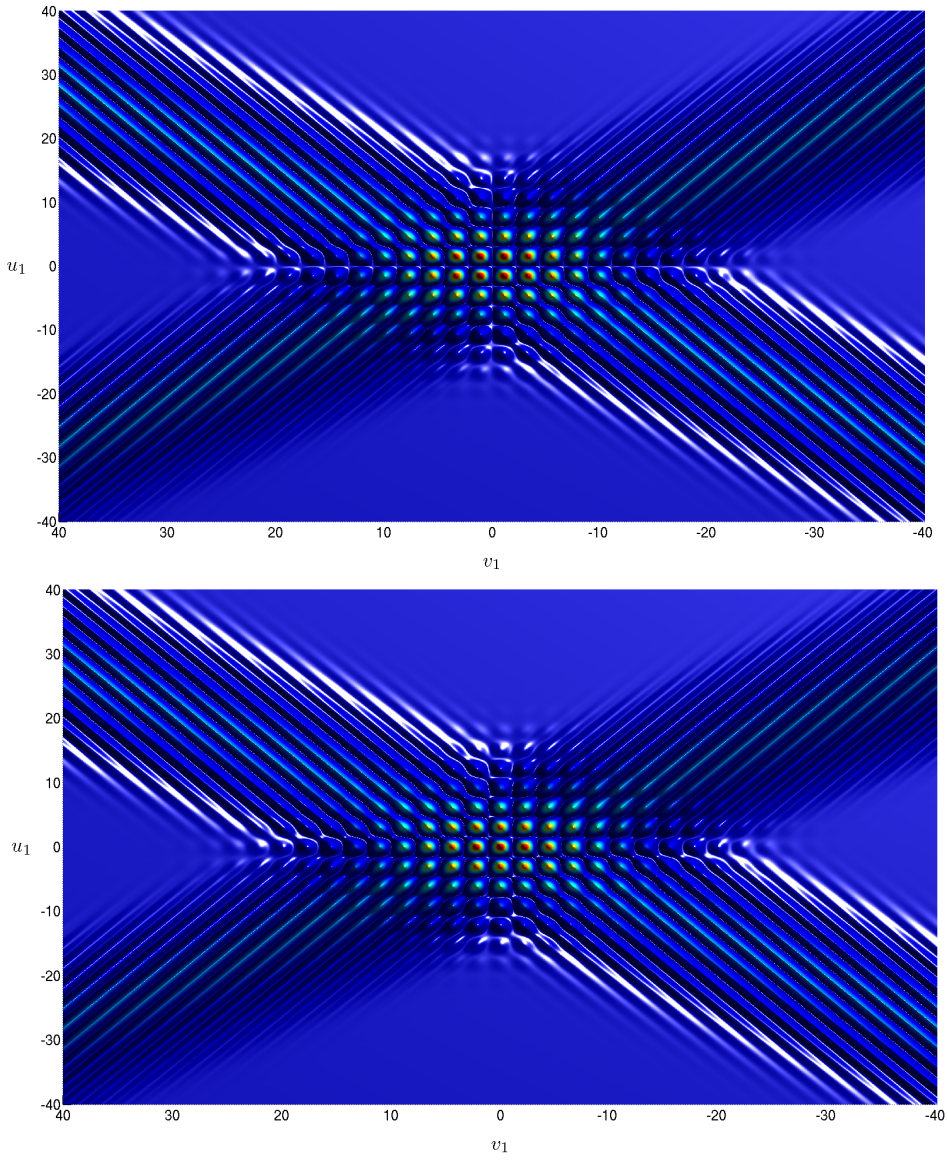

Figure F.6: This figure shows the same wave packets as F.5. From this perspective, one can clearly see that the boundary condition $\psi = 0$ at $u_1 = 0$ is satisfied for the upper wave packet.

Bibliography

[1] P. A. M. Dirac, 'Generalized Hamiltonian Dynamics', Proc. Roy. Soc. London **A 246**, 326 (1958).

[2] P. A. M. Dirac, 'The Theory of Gravitation in Hamiltonian Form ', Proc. Roy. Soc. London **A 246**, 333 (1958).

[3] P. A. M. Dirac, 'Lectures on Quantum Mechanics', in: Belfer Graduate School of Science Monograph Series, **2**, Yeshiva University, New York (1964).

[4] B. S. DeWitt, 'Quantum Theory of Gravity. I. The Canonical Theory', Phys. Rev. **160**, 1113 (1967).

[5] H. J. Matschull, 'Dirac's Canonical Quantization Programme', arXiv:quant-ph/9606031 (1996).

[6] R. Arnowitt, S. Deser, C. Misner, 'The Dynamics of General Relativity', in: 'Gravitation: An Introduction to Current Research', ed. by: L. Witten; Wiley, New York (1962).

[7] K. Kuchar, 'Canonical Quantization of Gravity', in: 'Relativity, Astrophysics and Cosmology', ed. by: W. Israel; D. Reidel Publishing Company, Dordrecht (1973).

[8] K. Kuchar, 'Canonical Methods of Quantization', in: 'Quantum Gravity 2: A Second Oxford Symposium', ed. by: C. J. Isham, R. Penrose, D. W. Sciama; Clarendon Press, Oxford (1981).

[9] K. Kuchar, 'On equivalence of parabolic and hyperbolic super-Hamiltonians', J. Math. Phys. **19**, 390 (1978).

[10] C. J. Isham, 'An Introduction to Quantum Gravity', in: 'Quantum Gravity 2: A Second Oxford Symposium', ed. by: C. J. Isham, R. Penrose, D. W. Sciama; Clarendon Press, Oxford (1981).

[11] C. J. Isham, 'Canonical Quantum Gravity and the Problem of Time', Lectures of the NATO Summer School, Salamanca (1992). arXiv:gr-qc/9210011.

[12] S. A. Hojman, K. V. Kuchar, C. Teitelboim, 'Geometrodynamics regained', Ann. Phys. **96**, 88 (1976).

[13] P. G. Bergmann, A. Komar, 'The Phase Space Formulation of General Relativity and Approaches Toward its Canonical Quantization', in: 'General Relativity and Gravitation', Vol. 1, ed. by: A. Held; Plenum Press, New York (1980).

[14] P. G. Bergmann, A. Komar, 'The Coordinate Group Symmetries of General Relativity', Int. J. Theor. Phys. **5**, 15 (1972).

[15] J. A. Wheeler, 'Superspace and the Nature of Quantum Geometrodynamics', in: 'Quantum Cosmology', ed. by: L. Z. Fang, R. Ruffini; World Scientific, Singapore (1987).

[16] A. E. Fischer, J. E. Marsden, 'The initial value problem and the dynamical formulation of general relativity', in: 'General Relativity: An Einstein Centenary Survey', ed. by: S. W. Hawking, W. Israel; Cambridge University Press, Cambridge (1979).

[17] S. T. Ali, M. Englis, 'Quantization Methods: A Guide for Physicists and Analysts', Rev. Math. Phys. **17**, 391 (2005).

[18] D. Giulini, 'That strange Procedure Called Quantization', in: 'Quantum Gravity — From Theory to Experimental Search', ed. by: D. Giulini, C. Kiefer, C. Lämmerzahl; Springer Verlag, Berlin (2003).

[19] K. Kuchar, 'Canonical Quantum Gravity', arXiv:gr-qc/9304012v1 (1993).

[20] P. A. M. Dirac, 'The Principles of Quantum Mechanics', The International Series of Monographs on Physics **27**, Oxford University Press, Oxford (1958).

[21] H. J. Groenewold, 'On the principles of elementary quantum mechanics', Physica **12**, 405 (1946).

[22] L. van Hove, 'Sur certaines représentations unitaires d'un groupe infini de transformations', Mem. Acad. Roy. Belgique Cl. Sci. (6) **26**, 1, (1951).
L. van Hove, 'Sur le problème des relations entre les transformations unitaires de la mécanique quantique et les transformations canoniques de la mécanique classique', Acad. Roy. Belgique Bull. Cl. Sci. (5) **37**, 610, (1951).

[23] K. Sundermeyer, 'Constrained Dynamics', Lecture Notes in Physics **169**, Springer Verlag, Berlin (1982).

[24] J. Klauder, 'Soluble models of quantum gravitation', in: 'Relativity', ed. by: M. Carmeli, S. Flicker, L. Witten; Plenum Press, New York (1970).

[25] A. Komar, 'Constraints, Hermiticity, and correspondence', Phys. Rev. D **19**, 2908 (1979).

[26] A. Komar, 'Consistent factor-ordering of general-relativistic constraints', Phys. Rev. D **20**, 830 (1979).

[27] J. L. Anderson, 'Factor Sequences in Quantized General Relativity', Phys. Rev. **114**, 1182 (1959).

[28] L. H. Ryder, 'Quantum Field Theory', Cambridge University Press, Cambridge (1996).

[29] C. G. Torre, 'Is general relativity an 'already parametrized' theory?', Phys. Rev. D **46**, R3231 (1992).

[30] J. L. Anderson, in: 'Proceedings of the First Eastern Theoretical Physics Conference 1962', ed. by: M. E. Rose; Gordon and Breach, New York (1963).

[31] J. Schwinger, 'Quantized Gravitational Field. II', Phys. Rev. **132**, 1317 (1963).

[32] N. C. Tsamis, R. P. Woodard, 'The factor-ordering problem must be regulated', Phys. Rev. D **36**, 3641 (1987).

[33] J. L. Friedman, I. Jack, 'Formal commutators of the gravitational constraints are not well defined: A translation of Ashtekar's ordering to the Schrödinger representation', Phys. Rev. D **37**, 3495 (1988).

[34] P. A. M. Dirac, 'The Quantization of the Gravitational Field', in: 'Contemporary Physics: Trieste Symposium', Vol. 1, IAEA, Vienna (1969).

[35] F. Mandl, G. Shaw, 'Quantenfeldtheorie', Aula-Verlag, Wiesbaden (1993).

[36] A. O. Barvinsky, 'Unitarity approach to quantum cosmology', Phys. Rep. **230**, 237 (1993).

[37] C. Kiefer, 'The Semi-classical Approximation to Quantum Gravity', in: 'Canonical Gravity: From Classical to Quantum', ed. by: J. Ehlers, H. Friedrich; Springer Verlag, Berlin (1993).

[38] C. Kiefer, 'Decoherence in quantum cosmology', in: 'Proceedings of the 10th Seminar on Relativistic Astrophysics and Gravitation', ed. by: S. Gottlöber, J.P. Mücket, V. Müller; World Scientific, Singapore (1992).

[39] J. B. Hartle, S. W. Hawking, 'Wave function of the Universe', Phys. Rev. D **28**, 2960 (1983).

[40] M. Blagojevic, 'Gravitation and Gauge Symmetry', IOP Publishing (2002).

[41] R. M. Wald, 'General Relativity', The University of Chicago Press, Chicago (1984).

[42] W. M. Frank, D. J. Land, R. M. Spector, 'Singular Potentials', Rev. Mod. Phys. **43**, 36 (1971).

[43] K. Jänich, 'Analysis für Physiker und Ingenieure', Springer-Lehrbuch, Springer Verlag, Berlin (1990).

[44] L. D. Landau, E. M. Lifshitz, 'Course of Theoretical Physics. Volume 3: Quantum mechanics. Non-relativistic theory', Pergamon Press, London (1959).

[45] R. G. Newton, 'Scattering Theory of Waves and Particles', McGraw-Hill Book Company, New York (1966).

[46] K. M. Case, 'Singular Potentials', Phys. Rev. **80**, 797 (1950).

[47] H. Fischer, H. Kaul, 'Mathematik für Physiker 2', Teubner Studienbücher, B. G. Teubner, Stuttgart (1998).

[48] A. Y. Kamenshchik, C. Kiefer, B. Sandhöfer, 'Quantum cosmology with a big-brake singularity', Phys. Rev. D **76**, 064032 (2007).

[49] E. Komatsu et al., 'Five-year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Cosmological Interpretation', arXiv:0803.0547v2 (2008).

[50] A. G. Riess et al., 'Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant', Astronom. J. **116**, 1009 (1998).

[51] S. Perlmutter et al., 'Measurements of Ω and Λ from 42 High-Redshift Supernovae', Astrophys. J. **517**, 565 (1999).

[52] X. Wang, M. Tegmark, M. Zaldarriaga, 'Is cosmology consistent?', Phys. Rev. D **65**, 123001 (2002).

[53] J. L. Tonry et al., 'Cosmological results from High-z Supernovae', Astrophys. J. **594**, 1 (2003).

[54] M. Tegmark et al., 'Cosmological parameters from SDSS and WMAP', Phys. Rev. D **69**, 103501 (2004).

[55] R. R. Caldwell, 'A Phantom Menace? Cosmological consequences of a dark energy component with super-negative equation of state', Phys. Lett. **545**, 23 (2002).

[56] A. G. Riess et al., 'The Farthest Known Supernova: Support for an Accelerating Universe and a Glimpse of the Epoch of Deceleration', Astrophys. J. **560**, 49 (2001).

[57] R. R. Caldwell, M. Kamionkowski, N. N. Weinberg, 'Phantom Energy: Dark Energy with $w < -1$ Causes a Cosmis Doomsday', Phys. Rev. Lett. **91**, 071301 (2003).

[58] M. Kowalski et al., 'Improved Cosmological Constraints from New, Old, and Combined Supernova Data Sets', Astrophys. J. **686**, 749 (2008).

[59] J. Hao, X. Li, 'Attractor solution of phantom field', Phys. Rev. D **67**, 107303 (2003).

[60] J. Hao, X. Li, 'Phantom cosmic dynamics: Tracking attractor and cosmic doomsday', Phys. Rev. D **70**, 043529 (2004).

[61] E. J. Copeland, A. R. Liddle, D. Wands, 'Exponential potentials and cosmological scaling solutions', Phys. Rev. D **57**, 4686 (1998).

[62] P. G. Ferreira, M. Joyce, 'Cosmology with a primordial scaling field', Phys. Rev. D **58**, 023503 (1998).

[63] C. Kiefer, 'Wave packets in minisuperspace', Phys. Rev. D **38**, 1761 (1988).

[64] M. Abramowitz, I. A. Stegun (eds.), 'Handbook of Mathematical Functions', U. S. National Bureau of Standards, Washington (1972), Sec. 9.3.6.

[65] M. Abramowitz, I. A. Stegun (eds.), *op. cit.*, Sec. 10.4.59.

[66] M. Dabrowski, C. Kiefer, B. Sandhöfer, 'Quantum phantom cosmology', Phys. Rev. D **74**, 044022 (2006).

[67] *This citation has been provided by A. Kamenshchik. I have not been able to access the journal myself.* J. D. Barrow, G. J. Galloway, F. J. Tipler, Mon. Not. R. astron. Soc. **223**, 835 (1986).

[68] V. Gorini, A. Kamenshchik, U. Moschella, V. Pasquier, 'Tachyons, scalar fields, and cosmology', Phys. Rev. D **69**, 123512 (2004).

[69] A. Sen, 'Rolling tachyon', JHEP **0204**, 048 (2002).

[70] B. Carter, 'Duality relation between charged elastic strings and superconducting cosmic strings', Phys. Lett. B **224**, 61 (1989); A. Vilenkin, 'Effect of small-scale structure on the dynamics of cosmic strings', Phys. Rev. D **41**, 3038 (1990).

[71] A. Kamenshchik, U. Moschella, V. Pasquier, 'An alternative to quintessence', Phys. Lett. B **511**, 265 (2001).

[72] I. S. Gradshteyn, I. M. Ryzhik, 'Table of Integrals, Series and Products', Academic Press, San Diego, (2000), Sec. 2.132.3.

[73] A. Ashtekar, M. Bojowald, J. Lewandowski, 'Mathematical structure of loop quantum cosmology ', Adv. Theor. Math. Phys. **7**, 233 (2003); A. Ashtekar, T. Pawlowski, P. Singh, 'Quantum nature of the big bang: An analytical and numerical investigation', Phys. Rev. D **73**, 124038 (2006); A. Ashtekar, T. Pawlowski, P. Singh, 'Quantum Nature of the Big Bang', Phys. Rev. Lett. **96**, 141301 (2006).

[74] M. Bojowald, D. Cartin, G. Khanna, 'Lattice refining loop quantum cosmology, anisotropic models and stability', Phys. Rev. D **76**, 064018 (2007).

[75] A. Ashtekar, T. Pawlowski, P. Singh, 'Quantum nature of the big bang: Improved dynamics', Phys. Rev. D **74**, 084003 (2006).

[76] A. Ashtekar, T. Pawlowski, P. Singh, K. Vandersloot, 'Loop quantum cosmology of $k = 1$ FRW models', Phys. Rev. D **75**, 024035 (2007).

[77] C. F. Dunkl, 'A Laguerre Polynomial Orthogonality and the Hydrogen Atom', arXiv:math-ph/0011021v2 (2002).

[78] I. S. Gradshteyn, I. M. Ryzhik, *op. cit*, Sec. 7.414.4.

[79] T. W. B. Kibble, 'Geometrization of quantum mechanics', Commun. Math. Phys **65**, 189 (1979).

[80] M. Bojowald, B. Sandhöfer, A. Skirzewski, A. Tsobanjan, 'Effective Constraints for Quantum Systems', accepted for publication in: Rev. Math. Phys., arXiv:0804.3365v1 [math-ph] (2008).

[81] M. Bojowald, A. Skirzewski, 'Effective Equations of Motion for Quantum Systems', Rev. Math. Phys. **18**, 713 (2006).

[82] R. F. Baierlein, D. H. Sharp, J. A. Wheeler, 'Three-Dimensional Geometry as Carrier of Information about Time', Phys. Rev. **126**, 1864 (1962).

[83] T. Schilling, 'Geometry Of Quantum Mechanics', PhD Thesis, Institute for Gravitation and the Cosmos, Pennsylvania State University (1996). *Can be downloaded from:* http://gravity.psu.edu/research/archives/thesis/1996/troy.pdf

[84] M. Bojowald, A. Skirzewski, 'Quantum Gravity and Higher Curvature Actions', Int. J. Geom. Methods Mod. Phys. **4**, 25 (2007).

Acknowledgements

Theorem 1 (R. Nagel, 2008, oral communication) *Suppose* $\mathfrak{A}$ *to be an achievement of person* $\mathfrak{P}$*. We then can describe* $\mathfrak{A}$ *as the direct sum of three subspaces,* $\mathfrak{A} = \mathfrak{A}_E \oplus \mathfrak{A}_H \oplus \mathfrak{A}_L$ *, where*

(i) the subspace $\mathfrak{A}_E$ *is determined by the effort of person* $\mathfrak{P}$*,*

(ii) the help of good friends of person $\mathfrak{P}$ *spans* $\mathfrak{A}_H$ *and*

(iii) the subspace $\mathfrak{A}_L$ *can be uniquely identified as a good portion of luck.*

Taking the assumptions of this theorem for granted, let me first of all thank my supervisor Prof. Claus Kiefer for introducing me to the exciting field of quantum gravity. Of course, I am indebted to him for the choice of topic for this thesis, but I am particularly thankful for the freedom admitted to pursue my own research interests and the richness of collaborations he introduced me to.

Moreover, I want to thank Prof. Martin Bojowald, for the kind reception at the Institute for Gravitation and the Cosmos, for an interesting collaboration and a very patient and open ear for my questions and doubts.

I am indebted to Prof. Alexander Kamenshchik for suggesting the big-brake model but also for many interesting discussions. I guess, I should thank him and Prof. Andrei Barvinsky for urging me to finally submit this thesis. I also want to thank Prof. Andrei Barvinsky for useful discussions on the relation between the canonical and the covariant approach to quantum general relativity.

Finally, I thank Prof. Hehl for discussions and for always very carefully selected and useful literature on whatever topic. I also profitted very much from his critical attitude and remarks during our seminars.

Having covered the official side, I now have to face the task to embrace the many who, in one or the other way, have contributed to this thesis.

First of all, I want to thank Philipp Kuhn, who has accompanied this thesis — and me — with help, advice, tea and a good portion of humour through all four years. I especially want to thank him for his pedantic

judgement of and unbreakable enthusiasm for plots, his patience with Matlab and with myself.

But this thesis had never been written if it was not for Diana Doert and Anna Sperlich who kept me company during the final phase of writing up. In this way, I did not only enjoy the best coffee-break entertainment, but also learned how to open moldy coffee mugs, wear jackets inside out and other practical things that might help me to get along in life. I just hope they profitted in turn at least a bit from the vast experience of a theoretical physicist.

A big thanks goes to the Institute for Physical and Theoretical Chemistry at the University of Bonn which released two of its highly qualified staff members, Christoph Reimann and Christoph Riplinger, in order that they may fight my orthography and my very capricious punctuation. I particularly want to thank Christoph Riplinger for being such a diligent advocate of consistent spelling and for supervising all the correction business.

Splitting work between two universities might be good, including a third one is even be better. In this sense, I want to thank Britta Dorn for a careful reading of a first version of Chapter 5 — and to be fair I should include the entire Arbeitsgruppe Funktionalanalysis of the University of Tübingen which collectively decided that a physicists writing is simply not understandable.

Everybody who knows me better, can surely appreciate what an endeavour this thesis was. I therefore want to seize the opportunity and thank all those who 'simply' provided support and encouragement. These are my parents, brothers and sister, Christoph Riplinger, Philipp Kuhn and Anne Franzen, who all, for some unknown reason, always had a very firm and unshakeable faith that everything will somehow work out. I furthermore want to thank Pantea Bashi for making me a fighter of sexual equality and the what-are-we-cooking-tonight Chaos Team in Bonn for providing me with at least one hot meal — and several cold beer — per week.

Lastly, I want to thank Thomas Kappeller for his elucidating instructions on applications, captions and the establishment and Christoph Sandhöfer (an acknowledgement cannot be good if not at least three Christophs are mentioned) for putting his master's words into such nice form, see the above Theorem 1 and an efficient proof reading.

I want to thank the Friedrich–Ebert–Stiftung for financial support, during my undergraduate as well as graduate studies.

Erklärung

Ich versichere, daß ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit — einschließlich Tabellen, Karten und Abbildungen —, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; daß diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; daß sie — abgesehen von unten angegebenen Teilpublikationen — noch nicht veröffentlicht worden ist sowie, daß ich eine solche Veröffentlichung vor Abschluß des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Claus Kiefer betreut worden.

Köln, den

Teilpublikationen:

'Effective Constraints for Quantum Systems',
M. Bojowald, A. Skirzewski, **B. Sandhöfer** and A. Tsobanjan,
angenommen zur Publikation in Rev. of Math. Phys.,
(arXiv:0804.3365).

'Quantum phantom cosmology',
B. Sandhöfer,
Proceedings of the Eleventh Marcel Grossmann Meeting on General Relativity, ed. von H. Kleinert, R.T. Jantzen, R. Ruffini; World Scientific, Singapore, 2008, S. 1887.

‘Quantum Cosmology’,
C. Kiefer und **B. Sandhöfer**,
Beitrag zu ‘Beyond the Big Bang’, ed. von R. Vaas; Springer Frontiers Collection, 2009,
(arXiv:0804.0672).

‘Quantum Cosmology with big-brake singularity’ ,
A. Kamenshchik, C. Kiefer und **B. Sandhöfer**,
Phys. Rev. **D76** 064032 (2007).

‘Quantum phantom cosmology’ ,
M. P. Dabrowski, C. Kiefer und **B. Sandhöfer**,
Phys. Rev. **D74** 044022 (2006).

Lebenslauf

Kontakt Barbara Sandhöfer
Institut für Theoretische Physik
Universität zu Köln
Zülpicher Straße 77
50937 Köln

Telefon: 0049-221-470 3642
Email: bs324@thp.uni-koeln.de

Persönliche Angaben

geboren 24.02.1981 in Saarbrücken

Staatsangehörigkeit deutsch

Ausbildung

Dissertation	Institut für Theoretische Physik, Universität zu Köln Titel: The Fate of Singularities in Quantum Cosmology and the Application of Generalized Effective Equations to Constrained Quantum Systems Betreuer: Prof. Claus Kiefer, Universität zu Köln ausgezeichnet mit dem Graduiertenstipendium der Friedrich-Ebert-Stiftung
C.A.S.M.	Department of Applied Mathematics and Theoretical Physics, University of Cambridge, (2004) Essay: The Functional Schrödinger Equation in Quantum Field Theory Betreuer: Prof. Manton Gesamtnote: Merit
Vordiplom	Fakultät für Mathematik und Physik, Eberhard-Karls-Universität Tübingen, (2003)

	Gesamtnote: sehr gut ausgezeichnet mit dem Stipendium der Friedrich-Ebert-Stiftung
Abitur	Landesschule Pforta, Schulpforte, (2000) Gesamtnote: 1.3 ausgezeichnet mit dem Preis der Deutschen Physikalischen Gesellschaft

9. Februar 2009