Path Integrals in the Sky

Classical and Quantum Problems with Minimal Assumptions

by

Job Leon Feldbrugge

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Physics

Waterloo, Ontario, Canada, 2019

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner	Malcolm Perry, University of Cambridge
Supervisor	Neil Turok, Perimeter Institute, University of Waterloo
Internal/External Member	Jon Yard, University of Waterloo
Internal Members	Niayesh Afshordi, Perimeter Institute, University of Waterloo
	Kendrick Smith, Perimeter Institute
	Eric Poisson, University of Guelph

Author's declaration

This thesis consists of material all of which I authored or co-authored: see Statement of Contributions included in the thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Statement of Contributions

Chapter 3 of this thesis consists of material from the paper [139], co-authored with Neil Turok and Jean-Luc Lehners.

Chapter 4 of this thesis consists of material from the paper [140], co-authored with Neil Turok and Jean-Luc Lehners.

Chapter 5 of this thesis consists of material from the paper [142], co-authored with Neil Turok and Jean-Luc Lehners.

Chapter 6 of this thesis consists of material from the paper [141], co-authored with Neil Turok and Jean-Luc Lehners.

Chapter 7 of this thesis consists of material developed with Neil Turok, Laura Sberna and Angelika Fertig.

Chapter 9 of this thesis consists of material from the paper [143], co-authored with Neil Turok and Ue-Li Pen.

Chapter 10 of this thesis consists of material from the paper [144], co-authored with Rien van de Weygaert, Johan Hidding and Joost Feldbrugge.

Abstract

Cosmology has, after the formulation of general relativity, been transformed from a branch of philosophy into an active field in physics. Notwithstanding the significant improvements in our understanding of our Universe, there are still many open questions on both its early and late time evolution. In this thesis, we investigate a range of problems in classical and quantum cosmology, using advanced mathematical tools, and making only minimal assumptions. In particular, we apply Picard-Lefschetz theory, catastrophe theory, infinite dimensional measure theory, and weak-value theory.

To study the beginning of the Universe in quantum cosmology, we apply Picard-Lefschetz theory to the Lorentzian path integral for gravity. We analyze both the Hartle-Hawking no-boundary proposal and Vilenkin's tunneling proposal, and demonstrate that the Lorentzian path integral corresponding to the mini-superspace formulation of the two proposals is well-defined. However, when including fluctuations, we show that the path integral predicts the existence of large fluctuations. This indicates that the Universe cannot have had a smooth beginning in Euclidean de Sitter space. In response to these conclusions, the scientific community has made a series of adapted formulations of the no-boundary and tunneling proposals. We show that these new proposals suffer from similar issues.

Second, we generalize the weak-value interpretation of quantum mechanics to relativistic systems. We apply this formalism to a relativistic quantum particle in a constant electric field. We analyze the evolution of the relativistic particle in both the classical and the quantum regime and evaluate the back-reaction of the Schwinger effect on the electric field in $1+1$-dimensional spacetime, using analytical methods. In addition, we develop a numerical method to evaluate both the wavefunction and the corresponding weak-values in more general electric and magnetic fields.

We conclude the quantum part of this thesis with a chapter on Lorentzian path integrals. We propose a new definition of the real-time path integral in terms of Brownian motion on the Lefschetz thimble of the theory. We prove the existence of a σ-measure for the path integral of the non-relativistic free particle, the (inverted) harmonic oscillator, and the relativistic particle in a range of potentials. We also describe how this proposal extends to more general path integrals.

In the classical part of this thesis, we analyze two problems in late-time cosmology. Multi-dimensional oscillatory integrals are prevalent in physics, but notoriously difficult to evaluate. We develop a new numerical method, based on multi-dimensional Picard-Lefschetz theory, for the evaluation of these integrals. The virtue of this method is that its efficiency increases when integrals become more oscillatory. The method is applied to

interference patterns of lensed images near caustics described by catastrophe theory. This analysis can help us understand the lensing of astrophysical sources by plasma lenses, which is especially relevant in light of the proposed lensing mechanism for fast radio bursts.

Finally, we analyze large-scale structure formation in terms of catastrophe theory. We show that the geometric structure of the three-dimensional cosmic-web is determined by both the eigenvalue and the eigenvector fields of the deformation tensor. We formulate caustic conditions, classifying caustics using properties of these fields. When applied to the Zel'dovich approximation of structure formation, the caustic conditions enable us to construct a caustic skeleton of the three-dimensional cosmic-web in terms of the initial conditions.

Acknowledgements

Several people have contributed directly or indirectly to this dissertation. Foremost, I am grateful to my advisor Neil Turok, who continues to inspire me with many engaging discussions. Neil has introduced me to the world of theoretical physics, while always insisting on minimal assumptions and aiming for a fundamental understanding. I would like to thank Rien van de Weygaert for believing in my scientific abilities, supporting me during my education in The Netherlands, and for our continuing collaborations on late time cosmology. I thank Ue-Li Pen and Jean-Luc Lehners for their mentoring, and collaborations in astrophysics and quantum cosmology. I treasure the discussions with Bernard Jones and am grateful for his valuable advice.

During my graduate years, I have had the pleasure to closely collaborate with Alice Di Tucci, Angelika Fertig, Beatrice Bonga, Derek Inman, Erik Schnetter, Johan Hidding, and Laura Sberna. We have learned a great deal about physics and I have greatly enjoyed their companionship. The PI community has brightened my life at the Perimeter Institute.

I am very grateful to my parents, Joost and Joke, and my sisters, Lelie and Eline for their support and care. I finally want to thank my Nynke, 'mijn verkering', for her endless love and support during these four years.

Table of Contents

List of Tables

List of Figures

Chapter 1

Introduction

> I would rather discover a single fact, even a small one, than debate the great issues at length without discovering anything at all.
>
> Galileo Galilei

Since the dawn of human civilization, people have been wondering about the beginning of life and the origin of the universe. Cosmology[1], in the elementary sense of an interest in the natural world and the heavenly phenomena, predates science and can be traced back for several thousands of years, before recorded history. Every culture has its own creation story to explain how it all began. The earliest surviving documents in 'western civilization' date from ancient Egypt, Mesopotamia, Greece, and Rome. Also in China, Japan, India, Persia, and pre-Columbian America, we find a rich collection of early creation myths. These cosmological stories often take religious forms, each with distinct explanations for what we see around us. For a summary of ancient cosmology see [263, 226].

For a long time, cosmology remained in the realm of religion and philosophy. It is only recently, with the development of the modern telescopes and advances in our understanding of matter and gravity, that cosmology started to enter the realm of science and in particular physics and astronomy. This branch of physics, known as *modern cosmology* or *precision cosmology*, deals with the physics of our universe as a whole and attempts to describe the origin and evolution of the cosmos. Even though the methods employed in the study of cosmology have changed dramatically, the existential questions have largely remained the same:

[1]Derived from the Greek *cosmos* meaning 'order', 'regular behaviour' and 'beauty'. The idea that the cosmos can be understood using the ratio was wildly ambitious, but now two millennia later, there is reason for optimism as a large part of the universe seem to follow a beautiful order.

- How did the universe originate, or has it always existed?
- What is the universe made of, *i.e.*, what is space and time?
- What does the universe contain?
- How does the universe evolve?
- What is the fate of our universe?
- And ultimately, how did life begin?

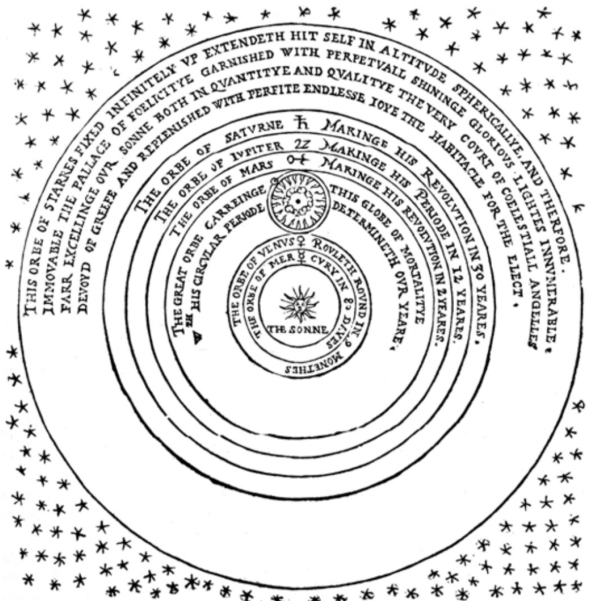
(the curious reader may skip to section 1.3 for the current understanding). Cosmology has during the last century – after the paradigm shift to general relativity – undergone large developments leading to the current ΛCDM[2] hot big bang model. Even though some of the existential questions have been largely answered, others remain open. At the same time, the mathematical techniques used in cosmology have remained limited to the analysis of differential equations and the quantization of quantum fields on curved spacetimes. In this thesis, we will introduce and extend the use of advanced mathematical techniques such as Picard-Lefschetz theory, catastrophe theory and infinite dimensional measure theory to the study of cosmology, in an attempt to shed new light on the open existential questions.

1.1 The history of cosmology

It is often claimed that modern science began with Nicolaus Copernicus' famous *De revolutionibus orbium coelestium* (on the revolutions of the heavenly spheres) published in 1543 [82]. In this work, he rejected the then prevalent Ptolemaic *geocentric* system, in favor of a *heliocentric* system where the planets orbit the sun instead of the earth (see Figure 1.1). In the following century, Johannes Kepler (1609) introduced elliptical orbits [220] and Galileo Galilei (1610) presented supporting observations made with one of the first telescopes [155]. Kepler did not only replace the to that point assumed circular orbit by an ellipse, but also described the dynamics of the orbit in terms of three laws of planetary motion. The planets accelerate when approaching, and decelerate when moving away from the sun. René Descartes in his *Traité du monde et de la lumière* (treatise on the world and the light), written between 1629 and 1633, used Kepler's laws to construct a heliocentric cosmology [95]. For this model, he postulated three Cartesian laws of motion

[2] A ΛCDM universe is a homogeneous and isotropic Friedmann Robertson Walker Lemaître universe consisting of baryonic and cold dark matter, together with a cosmological constant Λ or dark energy.

(a) The Ptolemaic universe prevalent in Medieval times, with the earth surrounded with ten heavenly spheres. Illustration from Petrus Apianus' *Cosmographicum liber* (1533).



(b) The Copernican universe, with the planets orbiting the sun. Illustration from Thomas Digges' *A Prognostication everlasting* (1576).

Figure 1.1: Ptolemaic and Copernican solar system.

which all particles must obey. He adds to this the assumption that the universe does not have empty space or voids. In Descartes' cosmology, the universe started from chaos in which the particles were at rest. When a single particle started to move, another particle also needed to move to take the former particle's place. This led to a universe in which the motion of the particles formed eddies, which grew to form the planetary orbits, where every star is a 'sun' at the center of its own whirlpool (see Figure 1.2). Inspired by Descartes' work, Benedict de Spinoza attempted to construct a mathematical framework for the evolution of the universe in which everything is completely deterministic. In this analysis, he proposed to follow the causal chain backward to the initial cause, the *causa sui* which was the starting point for our universe [93]. However, René Descartes' cosmology was soon overturned by Isaac Newton's derivation of the elliptical orbit. Isaac Newton in his *Philosophiae naturalis principia mathematica* (1687) transformed Kepler's three laws of planetary dynamics into three *universal laws of motion*, and introduced the

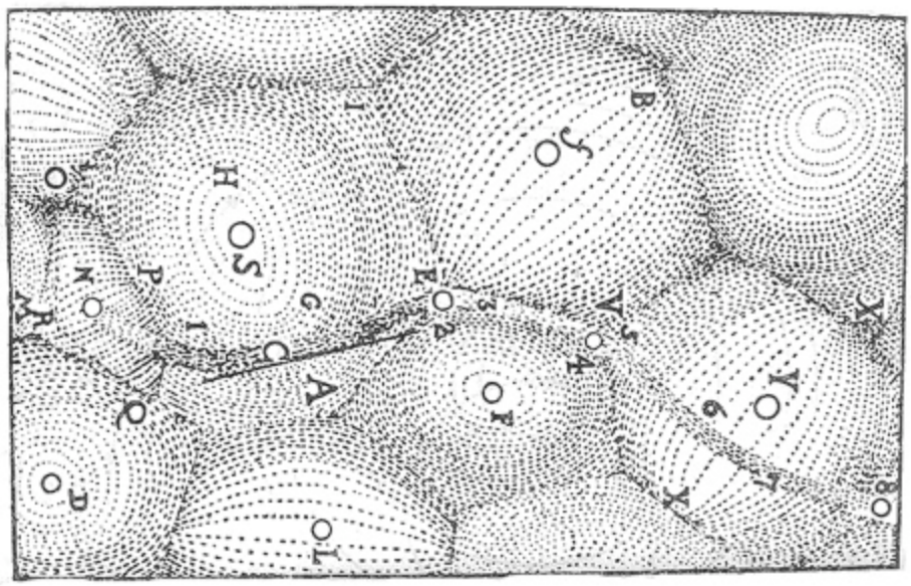


Figure 1.2: René Descartes' cosmological model based on whirlpools.

concept of the *universal gravitational force*[3] [261]. The revolutionary step by Newton was to assume that all massive bodies exert a gravitational force on each other and that that same gravitational force governs both falling apples, the tidal forces of the oceans and planetary orbits. This allowed Newton to develop a cosmological model which contrasted the one by Descartes. In Newton's universe, matter on large scale is uniformly distributed canceling out the attractive gravitational forces. Newton's cosmos is static, has always existed but is unstable with respect to matter fluctuations. This model is an example of a *steady-state universe*. After Newton, cosmology was discussed by many philosophers among whom Immanuel Kant (1724 - 1804) [219]. Kant proposed a hierarchical universe in which matter is clustered on ever-larger scales of hierarchy. The Kantian universe started with the primeval chaos of particles at rest in an infinite void. This initial state was unstable, as denser particles attract more tenuous particles. In order to avoid the possibility that all particles would form one massive object, Kant introduced an unspecified repulsive force. The universe would form structure at ever larger scales. He speculated about the formation of the solar system. Although some of his philosphical ideas about the order of the universe have turned out to be incorrect, his speculations about the formation of the Milky Way were surprisingly accurate, as he claimed that the galaxy is a disc-like structure

[3]See *Feynman's lost lecture: the motion of planets around the sun* [168] for a contemporary geometric derivation of Newtons result.

consisting of stars orbiting the galactic center. He, even more remarkably, suggested that observed nebulae are other *island universes* consisting of many stars like our Milky Way. Kant also discussed the dimensionality of space and the nature of space and time.

Modern cosmology can be considered to have commenced in 1917 with Albert Einstein's paper *Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie* (cosmological considerations of the general theory of relativity) [126]. Two years earlier, Einstein replaced Newton's universal law of gravity with his theory of *general relativity* making gravity consistent with his theory of *special relativity* (1915) [125, 124]. In general relativity, gravity is no longer a force but rather the consequence of the curvature of spacetime. Spacetime instantaneously became a dynamical entity which can be pulled and stretched by the matter and radiation it contains. When applied to the universe as a whole, a revolutionary new cosmological theory emerged. Einstein, as Newton had done before, proposed a steady-state model for the universe in which the gravitational forces are balanced. This premise required him to introduce a *cosmological constant*. However, the Einstein model did not last as it is unstable with respect to small energy fluctuations. Willem de Sitter considered an empty universe with a cosmological constant, obtaining an exponentially contracting and expanding solution now known as de Sitter spacetime (1917) [91]. This solution is currently seen as a good candidate for the fate of our universe. When Alexander Friedmann applied general relativity to the universe as a whole while including an energy content – assuming a homogeneous and isotropic perfect fluid – he obtained a finite spherical model which expands and subsequently collapses (1922) [152]. Applied to hyperbolically curved space, he obtained a forever expanding universe (1924) [153]. Friedmann proposed a cyclic universe, avoiding the problem of having to describe the beginning. These results were independently derived by Georges Lemaître (1927) [237]. However, Lemaître differed in opinion with respect to the latter point: he considered a universe with a beginning. Albert Einstein initially criticized the dynamical cosmological models [127], for he preferred a steady-state solution, but revised his opinion after Edwin Hubble's observation of the recessional velocity of extra-galactic nebulae (1929) [208][4]. This observation is now known as the *Hubble law*. As a response, Einstein together with de Sitter proposed a 'simple' flat, matter-dominated universe without a cosmological constant (1932) [128]. This model is currently known as the Einstein-de Sitter universe and is commonly used to study large-scale structure formation. Arthur Eddington combined Einstein's original static model with the de Sitter universe to unite the steady-state model with the observed expansion (1930) [122]. A few years later, Howard Robertson and Arthur Walker extended the work of Friedmann and Lemaître, and constructed a class of isotropic and homogeneous mod-

[4]It should be noted that the same observation had already been made by Georges Lemaître [238]. The corresponding paper was however delayed.

els, now known as the Friedmann Lemaître Robertson Walker (FLRW) universe, which includes all the above mentioned models (1935)[5] [285, 337]. Depending on its matter content, the universe can contract and expand while following the Friedmann equations. It is remarkable how Einstein's discovery of general relativity led in a span of twenty years, to a completely new scientific world view. Cosmological models became predictive and entered the realm of science.

While the consequences of Einstein's theory of general relativity were being developed and digested, another group of physicists started to study the quantum nature of matter. In 1931, in the paper *The beginning of the world from the point of view of quantum theory* [239], Lemaître used his insights in both general relativity and quantum mechanics to propose a model of the *primeval atom*; a precursor of the current *big bang model*. In his short paper – consisting of 457 words – Lemaître responds to Arthur Eddington, his previous collaborator who at that point favored his steady-state cosmology:

> Sir Arthur Eddington states that, philosophically, the notion of a beginning of the present order of Nature is repugnant to him. I would rather be inclined to think that the present state of quantum theory suggests a beginning of the world very different from the present order of Nature. Thermodynamical principles from the point of view of quantum theory may be stated as follows: (1) Energy of constant total amount is distributed in discrete quanta. (2) The number of distinct quanta is every increasing. If we go back in the course of time we must find fewer and fewer quanta, until we find all the energy of the universe packed in a few or even in a unique quantum.

He describes these initial quanta by discussing the emergence of space and time:

> Now, in atomic processes, the notions of space and time are no more than statistical notions; they fade out when applied to individual phenomena involving but a small number of quanta. If the world has begun with a single quantum, the notions of space and time would altogether fail to have any meaning at the beginning; they would only begin to have a sensible meaning when the original quantum had been divided into a sufficient number of quanta. If this suggestion is correct, the beginning of the world happened a little before the beginning of

[5]It is remarkable that the evolution of a matter-dominated universe, can be fully described in terms of Newtonian forces. However, it is only after the discovery of Einstein's cosmology that this was realized. Friedmann Robertson Walker Lemaître universes containing radiation or a cosmological constant cannot be described using Newton's law of gravitation.

> space and time. I think that such a beginning of the world is far enough from the present order of Nature to be not at all repugnant.

According to Lemaître, this unique primeval atom would decay into the universe following "a kind of super-radioactive process". This initial atom is like the *causa sui* of Spinoza in that it is the initial cause of the universe. It, however, differs in form as it, by its quantum nature, does not determine the complete evolution of the universe:

> Clearly the initial quantum could not conceal in itself the whole course of evolution; but, according to the principle of indeterminacy, that is not necessary. Our world is now understood to be a world where something really happens; the whole story of the world need not have been written down in the first quantum like a song on the disk of a phonograph. The whole matter of the world must have been present at the beginning, but the story it has to tell may be written step by step.

In this, Lemaître sees, in a sense, room for *free will* in quantum physics. Georges Lemaître, in the first accepted draft of the paper, united his cosmological with his religious views (being a Catholic priest) by adding the sentence "I think that everyone who believes in a supreme being supporting every being and every acting, believes also that God is essentially hidden and may be glad to see how present physics provides a veil hiding the creation." He, however, decided to remove this final paragraph from his paper.

The proposal of a universe with a beginning left many physicists worried. In particular, since the estimated age of the universe seemed to have been shorter than the estimated age of certain stars. As Willem de Sitter during a meeting in 1931 [92] stated

> I am afraid all we can do is to accept the paradox and try to accommodate ourselves to it, as we have done to so many paradoxes lately in modern physical theories. We shall have to get accustomed to the idea that the change of the quantity R, commonly called the 'radius of the universe', and the evolutionary changes of stars and stellar systems are two different processes, going on side by side without any apparent connection between them. After all, the 'universe' is a hypothesis, like the atom, and must be allowed the freedom to have properties and to do things which would be contradictory and impossible for a finite material structure.

It might not be surprising that Lemaître's views were not directly accepted in academic circles – it probably did not help that he was a Catholic priest. For a few decades, the evolution of the universe remained a subject of academic debate with, in particular, Hermann

Bondi and Thomas Gold's (1948) [50] and Fred Hoyle's (1948) [207] new proposals for steady-state models. It was Fred Hoyle who, in a BBC radio interview in 1949, coined the term *big bang* to contrast it with the steady-state models. The controversy was resolved in 1964 when Arno Penzias and Robert Wilson accidentally discovered the Cosmic Microwave Background (CMB) radiation field [270]. Penzias and Wilson observed the afterglow of a hot photo-electron plasma which permeates the universe, emitted during an earlier phase of our universe. This phenomenon was at the same time predicted by Robert Dicke, Jim Peebles, Peter Roll and David Wilkinson [105]. This afterglow, now known as the cosmic microwave background radiation field, invalidated all steady-state models. In 1978, Penzias and Willson received the Nobel prize for their discovery. After the observation of the CMB, the academic community was forced to take the big bang models more seriously. The question arose whether these singular models were generic to general relativity or rather an artifact of the homogeneity and isotropy assumptions of the FLRW universe. This was two years later resolved by Roger Penrose and Stephen Hawking's famous singularity theorems (1966) [191, 192, 193], which prove that the big bang singularity is indeed generic, assuming the universe contains only 'conventional' forms of matter and radiation. For a more detailed history of modern cosmology, we refer the interested reader to [263, 226, 216].

1.2 Observations

Also more recently, astrophysical observations have been immensely informative for theoretical cosmology. Since 1990, several dramatic advances in observational cosmology have reshaped the subject from a largely speculative science to precision cosmology. We here list a few of the key discoveries. For a detailed overview see [216].

- Among the many advances, a remarkable shift came from the observations of distant supernovae. In 1998, the Supernova Cosmology Project and the High-Z Supernova Search Team both used type Ia supernovae to measure the recessional speed of distant stars [284, 292]. They observed that the universe is expanding in an accelerated rather than the expected decelerated fashion. After Hubble's discovery of the recessional velocity of extra-galactic nebulae (1929) [208], the scientific community directed attention to cosmological models without a cosmological constant[6]. However, the observation of the accelerated expansion forced the cosmological constant

[6] John Wheeler: "I heard Einstein say to Gamow about the cosmological constant, 'That was the biggest blunder of my life.' "

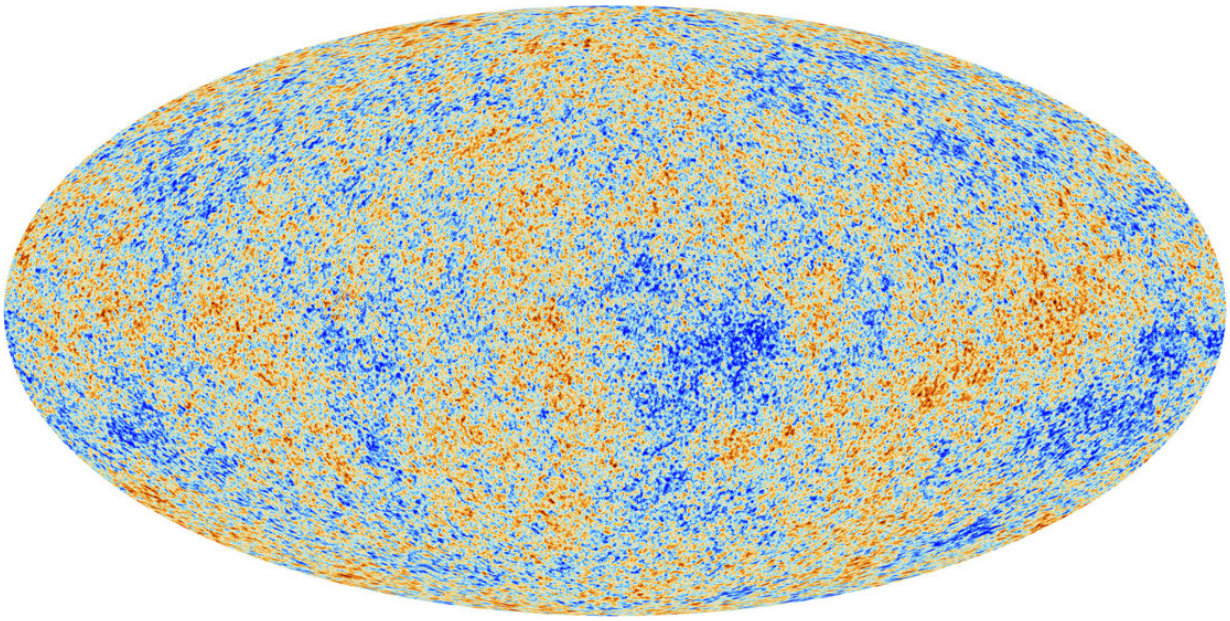


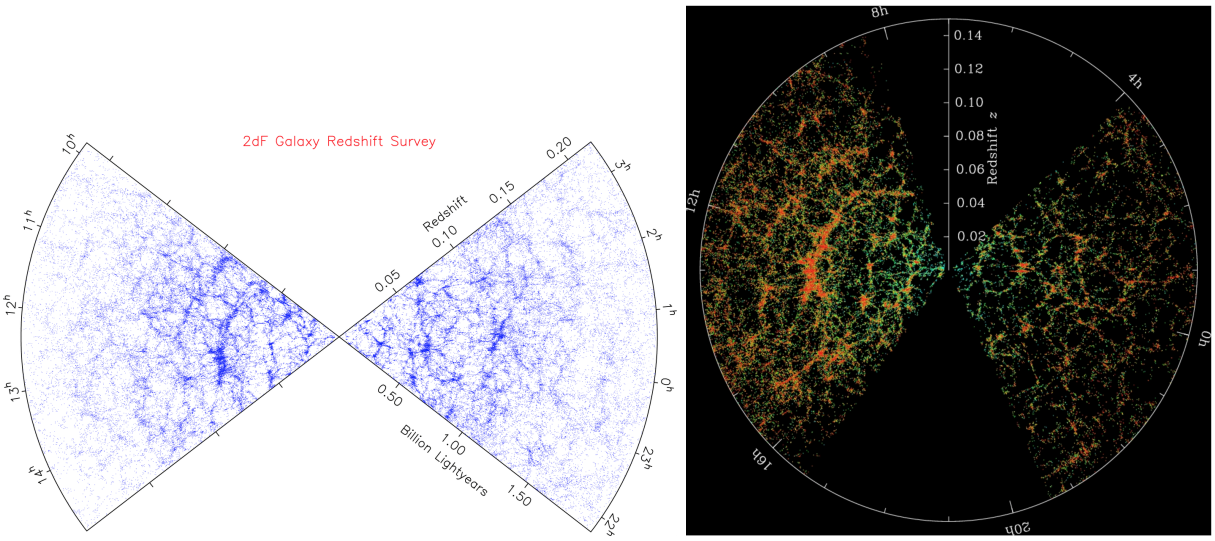
Figure 1.3: The cosmic microwave radiation field observed by the Planck satellite (2013).

Λ, also known as dark energy, back into mainstream cosmology. In 2011, the leaders of the two supernova projects were awarded the Nobel prize.

- Cosmology has over the last decades benefited from ever more precise observations of the CMB. In particular noteworthy are the three satellite missions COBE, WMAP, and Planck (see Figure 1.3 for the observation by Planck) [273]. The cosmic microwave background radiation field turned out to be extremely close to a thermal black body, with a temperature of 2.725 Kelvin, and with tiny Gaussian fluctuations. By fitting the spectrum of these fluctuations to a ΛCDM model, the Planck collaboration estimated the universe to be approximately 13.8 billion years old, and approximately flat and to consist of 5% baryonic matter, 27% dark matter and 68% dark energy.

- Finally it should be noted that as of February 2016, the LIGO collaboration announced the first gravitational wave detection, originating from the merger of two black holes [1]. It is expected that further measurement of gravitational waves will not only probe general relativity but might also shed light on cosmology.

The above-mentioned observations provide a solid basis for the cosmological predictions made in the first decades of the 20th century and the corresponding big bang model. The



(a) The 2dF Galaxy Redshift Survey. (b) The Slown Digital Sky Survey.

Figure 1.4: Galaxy redshift surveys.

universe evolved through several stages; it started in the hot radiation-dominated phase, became neutral in the matter-dominated phase and currently undergoes an accelerated expansion driven by the cosmological constant.

Apart form having inspired fundamental progress in physical cosmology, observational astronomy has dramatically improved our understanding of the structure the universe contains. Several large galaxy redshift surveys – including the European 2dF Galaxy Redshift Survey (2dfGRS) running from 1997 till 2002 [78] and the American Sloan Digital Sky Survey (SDSS) starting data collection in 2000 and scheduled to run till 2020 – have mapped the position and spectra of millions of galaxies (see Figure 1.4a) [47]. These surveys unveiled that the galaxies are distributed in a gigantic foam-like structure, consisting of clusters, string-like filaments, walls, and gigantic voids.

The observation of this cosmic structure, known as the *cosmic-web* and its formation, was already anticipated by pioneers James Peebles in the United States and Yakov Zel'dovich in the Soviet Union in the 1970s. Two competing models for structure formation were developed. The Soviet school constructed the *adiabatic scenario*, in which the first structures to form are extremely massive. Galaxies then form by successive processes of fragmentation of these large objects. For this reason, the adiabatic scenario is also called a *top-down model*. The American school constructed the *isothermal scenario*, in which the

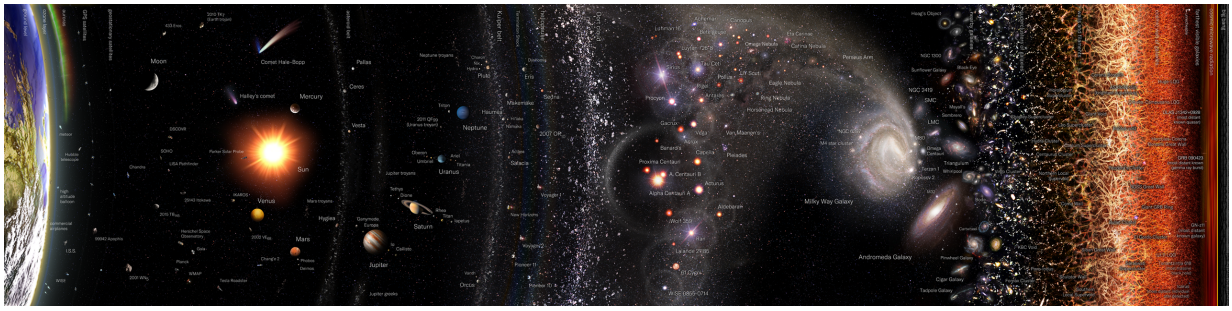


Figure 1.5: Logarithmic map of the evolution of the universe from today (left) to the big bang (right). Artist impression by Pablo Carlos Budassi.

first structures, proto-clouds or protogalaxies, are formed on a much smaller mass scale. Structures on larger scales form by successive mergers of these smaller objects in a process known as *hierarchical clustering*. For this reason, the isothermal scenario is described as *the bottom-up model*. Note that these two scenarios reflected the political views of the two superpowers. Both models were abandoned in the 1980s, as they did not match observations. In subsequent years, the emphasis was placed on the effect of dark matter. The Soviet school placed emphasis on Hot Dark Matter (HDM) whereas the American school placed emphasis on Cold Dark Matter (CDM), along the lines of the adiabatic and isothermal scenarios.

In the last decades, these models for structure formation have been further developed and, with the advance of supercomputers, studied with large N-body simulations (such as the Millennium [53], the EAGLE [289] and the Illustris simulation [259]). The simulations seem to agree with observations of the cosmic-web on large scales, where the evolution is mildly nonlinear and easy to predict. It is still controversial how well the simulations do on small scales, where gravitational collapse turns highly non-linear and the baryonic matter comes into its own. The complexity of the formation of the cosmic-web remains an active field of investigation.

The observational efforts in cosmology have helped us to understand the evolution of the universe, from the big bang to today. The observation of supernovae Ia explosions, the mapping of the CMB and the large redshift surveys have painted a rough picture of the universe (see figures 1.5 and 1.6).

Figure 1.6: Logarithmic map of the evolution of the universe from today (centre) to the big bang (outside). Artist impression by Pablo Carlos Budassi.

1.3 Open problems

Modern cosmology has – starting from Einstein's discovery of general relativity in 1915 and boosted by the wealth of observational probes in the last decades – undergone a spectacular transformation. The subject has been reshaped from a speculative branch of philosophy to an exact branch of physics. Along the way, science has provided an answer to some of the existential questions which characterize the subject. General relativity has shown us that spacetime is a dynamical entity which evolves according to the distribution and forms of matter it contains. Given the Friedman Lemaître Robertson Walker model, cosmological observations indicate that:

- Our universe evolves and is not in a steady-state;
- Our universe is approximately 13.8 billion years old;
- Our universe is currently expanding in an accelerated fashion and seems to turn into a de Sitter universe, in which all galaxies are pulled apart;
- Our current universe is close to being flat – in which Euclidean geometry is a good approximation – and consists of approximately 32% matter, 68% dark energy, and a negligible amount of radiation;
- Our universe contains many galaxies which are distributed in an intricate foam-like structure known as the cosmic-web;

Other problems have remained open:

- How did the universe originate?
- How do space and time emerge?
- What is dark energy and the cosmological constant or dark matter?
- How does the cosmic-web form and how do its different features interact with each other?
- How did life begin and did this only happen on earth?

In the last decades, there has been a lot of attention directed towards solving these problems. On the one hand, there are attempts to shed light on the fundamental side in early time cosmology. This includes among many, the inflationary paradigm developed since the '80s, and attempts to realize Lemaître's vision of the *primeval atom* by developing a theory of *quantum gravity* and *quantum cosmology*. However, so far the progress has been limited:

- The inflationary paradigm has led to a collection of models in which the early universe went through a phase of exponential expansion. This expansion sets the universe up in a flat state and allows for the creation of particles. The model is in particular able to match the observed Gaussian fluctuations in the CMB. However, the theory is often criticized for not being predictive, as the various models are of an ad hoc nature. The dynamics do not follow from a fundamental theory of physics, tend to be fine-tuned and can be tweaked to match many observation. Most models in the inflationary paradigm differ from Lemaître's early description of the beginning, in that they treat general relativity and in particular the background spacetime classically. They do not explain the origin and the emergence of classical spacetime. The current state of the theory of inflation, in particular, does not seem to answer the existential question about the creation of the cosmos.

- Quantum gravity and in particular quantum cosmology attempts to treat spacetime quantum-mechanically and, in this way, comes closer to Lemaître's early vision. However, there is at the moment no universal theory of quantum gravity. The existing approaches range from theories inspired by the geometric nature of general relativity, such as quantum geometrodynamics, to approaches based on particle physics, such as string theory. All models to a certain extent still suffer from conceptual problems. The formalisms that come closest to describing the big bang are probably Hartle and Hawking's no-boundary proposal [194, 188] and Vilenkin's tunneling proposal [325, 326, 323]. In the first part of this thesis, we will describe these approaches in detail.

In late time cosmology, there is a lot of attention for existential questions on the complexity side. Both galaxy redshift surveys and numerical experiments with N-body simulations have played an important role in studying the cosmic-web. Our knowledge of the structure has advanced, but our understanding of how it emerged from simple Gaussian fluctuations and how the different components interrelate is still limited. A better understanding will most probably require a more analytic derivation of the geometry of the cosmic-web to complement the redshift surveys and N-body simulations. The question on

the beginning of life is even more complex and studied in biology rather than mainstream cosmology.

1.4 Philosophy of this thesis

Abstract mathematics and theoretical physics have a curious relation. Logical reasoning, in the language of mathematics, forms the bridge between physical assumptions and falsifiable predictions. It was the introduction of Riemannian geometry, developed in the 19th century, which enabled Einstein to turn the *universal speed of light in vacuum* and the *equivalence principle* into a theory about the curvature of spacetime. Without the use of existing mathematics, it would have been much more difficult to formulate the gravitational bending of light and modern cosmology. The same can be said for the use of matrix algebra in quantum mechanics and group theory in gauge theories such as the standard model of particle physics. Mathematics, on the other hand, often draws new insights from theoretical physics. This motivated Paul Dirac to speculate about the two fields merging [106].

The interplay between the two subjects is however, initially, often far from clean. The initial application of mathematical techniques can be messy and inefficient. For example, compare Einstein's yearlong quest for general relativity with Hilbert's elegant derivation using the principle of least action. Moreover, theoretical physicists sometimes make progress by – for the time being – replacing mathematical rigor with physical intuition. A good example is the contemporary definition of the Lorentzian path integral. Nevertheless, we believe that abstract mathematics is a powerful tool in the development of physical understanding and that physical theories should ultimately be constructed on a rigorous mathematical foundation.

While our physical understanding of the cosmos has evolved, the use of mathematical tools has largely been limited to differential geometry and quantum field theory on curved spacetime. By using other advanced forms of mathematics, we can increase our understanding of the cosmos. In this thesis, we introduce and extend the use of Picard-Lefschetz theory, catastrophe theory, and infinite dimensional measure theory, to a range of problems in cosmology. In addition, we generalize the weak-value interpretation for quantum mechanics, developed by Ahronov et al. [5], to relativistic systems and apply the interpretation to quantum cosmology.

However beautiful they are, mathematical techniques in physics need to be in service of physical predictions and should be developed with the knowledge of our universe and

an eye on observations. We should in the study of new problems, stick to established physical principles as much as possible; a practice referred to as *radical conservatism* by John Wheeler. We should furthermore strive for models with a minimal number of free parameters. As Freeman Dyson recalled in 2004 [121]:

> By 1951 [...] we decided to explore the strong nuclear forces. [...] By 1953, after heroic efforts, we had plotted theoretical graphs of meson-proton scattering. We joyfully observed that our calculated numbers agreed pretty well with Fermi's measured numbers. So I made an appointment to meet with Fermi and show him our results. [...] When I arrived in Fermi's office, I handed the graphs to Fermi, but he hardly glanced at them. [...] Then he delivered his verdict in a quiet, even voice. 'There are two ways of doing physics', he said. 'One way, and this is the way I prefer, is to have a clear physical picture of the process that you are calculating. The other way is to have a precise and self-consistent mathematical formalism. You have neither'. [...] In desperation I asked Fermi whether he was not impressed by the agreement between our calculated numbers and his measured numbers. He replied, 'How many arbitrary parameters did you use for your calculations?', I thought for a moment about our cut-off procedures and said, 'Four'. He said, 'I remember my friend Johnny von Neumann used to say with four parameters I can fit an elephant, and with five I can make him wiggle his trunk'

In response to this claim, Mayer et al. (2010) [247] fitted an elephant with five complex parameters (see Figure 1.7). In Richard Feynman's words, "The first principle is that you must not fool yourself – and you are the easiest person to fool."

Theoretical physics – while in the process of development – is a confusing subject, with many competing principles and theories. However, the established theories, such as electromagnetism, special and general relativity and quantum mechanics have the strange quality of being interrelated in unexpected ways. Good examples are Maxwell's relation between electro-magnetism and optics, and Einstein's clock-in-the-box paradox, that demonstrates that quantum mechanics secretly implies the equivalence principle and is thus only consistent with Einstein's theory of general relativity (see Figure 1.8) [5]. Once formulated, quantum mechanics explains more than it assumes and, more importantly, provides a new perspective on physics. We believe that the ultimate theory of quantum gravity should share this surprising quality. My hope is that as with Andrew Wiles' experience when proving Fermat's last theorem in mathematics, the solution will be clear once we find it.

> Perhaps I can best describe my experience of doing mathematics in terms of entering a dark mansion. One goes into the first room, it is dark, one stumbles

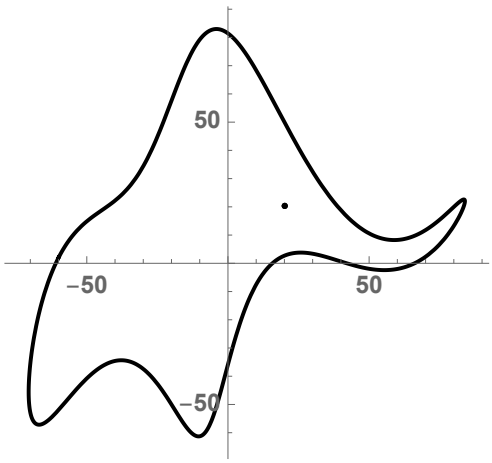


Figure 1.7: "With four parameters I can fit an elephant [or woolly mammoth], and with five I can make him wiggle his trunk." (Von Neumann)

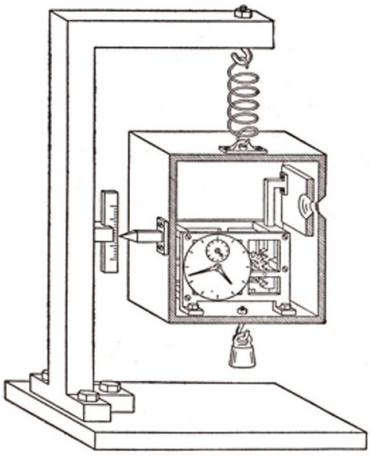


Figure 1.8: Sketch of Bohr's setup of Einstein's clock-in-box-experiment showing the intricate relation between quantum mechanics and general relativity. From *Albert Einstein: Philosopher-Scientist* [291].

around bumping into the furniture. Gradually you learn where each piece of furniture is. And finally after six months or so you find the light switch. You turn it on and suddenly it is all illuminated. You can see exactly where you were.

1.5 Outline

Our understanding of cosmology has, during the last century, dramatically advanced, but the mathematical techniques used in cosmology have remained limited to the analysis of the Friedmann equations and the dynamics of classical and quantum mechanical perturbations in an expanding universe. In this thesis, we introduce and extend the use of advanced mathematical techniques such as Picard-Lefschetz theory, catastrophe theory and infinite dimensional measure theory to the study of cosmology, in an attempt to shed new light on the open existential questions. In addition, we will extend weak-value theory to relativistic systems in an attempt to develop a new interpretation for quantum gravity.

Picard-Lefschetz, catastrophe and infinite-dimensional measure theory might at first glance seem unrelated. Picard-Lefschetz theory is an application of Cauchy's integral theorem and can be used to define and evaluate oscillatory integrals. Catastrophe theory concerns the classification of degenerate critical points in multi-dimensional functions. Infinite-dimensional measure theory is the study of stochastic processes such as Brownian motion. We have found these different branches of mathematics to be strongly interrelated, the combined framework being more powerful than the sum of its parts.

This thesis consists of two parts. The first part focuses on problems in the early universe, and in particular quantum cosmology. Chapter 2 is an introduction to quantum geometrodynamics as formulated by John Wheeler and Bryce DeWitt. We describe the original formulation, discuss the development of the Wheeler-DeWitt equation and the path integral for gravity, and summarize the original formulation of the Hartle-Hawking no-boundary proposal and the Vilenkin tunneling proposal for the beginning of the universe. In chapters 3, and 4, we study the no-boundary proposal and the tunneling proposal in the Lorentzian path integral formulation using Picard-Lefschetz theory. Our calculations demonstrate that the Lorentzian path integral selects Vilenkin's tunneling instanton. The fluctuations in this proposal are predicted to be unstable, indicating that our universe might not have started in a Euclidean phase of de Sitter spacetime. In response to the corresponding papers [139, 140], Diaz Dorronsoro, Halliwell, Hertog, Janssen, and Vreys [104, 117], developed two reformulations of the Lorentzian no-boundary proposal. In response to our work, also Vilenkin and Yamada [330], developed a new formulation

of the tunneling proposal. In chapters 5 and 6, we study the new proposals and illustrate a series of issues we find in these new formulations. More recently, Halliwell, Hartle, and Hertog [181] developed a new definition of the no-boundary proposal in terms of the Wheeler-DeWitt equation without reference to the path integral. Since this new formulation of the no-boundary proposal does not rely on the path integral for gravity, it is clear that the problems we found can be avoided. It should, however, be noted that such a definition in terms of the Wheeler-DeWitt equation requires the introduction of boundary conditions. In this thesis, we do not study this new formulation of the tunneling proposal. In response, Vilenkin and Yamada refined their tunneling proposal with the introduction of special boundary conditions [324]. It should be noted that the above discussion has inspired several cosmologists to develop new models for quantum cosmology to avoid the unstable fluctuations [48, 54, 27, 103]. We do not comment on these proposals here. The debate was moreover captured in several public articles [209, 345].

In chapter 7, we extend weak-value theory to relativistic systems. We investigate the world-line quantization of quantum field theory and study the creation of a particle-antiparticle pair in an electric field in 1+1-dimensional spacetime. Using these methods we analyze the particle while it tunnels in the Schwinger effect and evaluate the back-reaction on the electric field. We develop these methods in an attempt to formulate a new interpretation of quantum gravity.

In chapter 8, we use functional Picard-Lefschetz theory in an attempt to find a rigorous definition of the real-time path integral. Path integrals are usually defined through lattice regularization. We demonstrate why this definition is not mathematically satisfactory. Instead we propose a continuous-time regularization scheme to define the the real-time path integral. Unlike earlier continuous-time regularization schemes [222], we use Picard-Lefschetz theory define the integral by first deforming the integration domain to the Lefschetz thimble, rendering the integral absolutely convergent. We subsequently give an explicit definition of the σ-measure on the Lefschetz thimble for a range of toy-models and describe how these results can be extended to more complicated path integrals. It is in particular instructive to observe how the Lorentzian theory can differ from the perturbative theory on the Lefschetz thimble of the free theory.

In the second part of this thesis, we study problems in late time cosmology. In chapter 9, we numerically implementation Picard-Lefschetz theory for finite-dimensional oscillatory integrals. We use this new method to evaluate the interference patterns of two-dimensional lenses near caustics. The method is able to evaluate the Kirchhoff-Fresnel lensing integral in both the mildly oscillatory and geometric optics regime. Caustics in the images of astrophysical plasma lenses has been proposed as a mechanism causing the recently observed Fast Radio Bursts (FRBs) [83]. In chapter 9, we use the proposed method to describe

which plasma lenses are most likely to lead to FRBs.

In chapter 10, we study the application of catastrophe theory to large-scale structure formation. We prove that the caustic structure of the cosmic-web is completely determined by both the eigenvalue and the eigenvector fields of the deformation tensor of the evolution. We specifically describe the occurrence of different types of caustics by means of the caustic conditions. Note that the eigenvector fields are normally ignored. When the caustic conditions are applied to the Zel'dovich approximation of structure formation, they give rise to a three-dimensional caustic skeleton in terms of the initial conditions. This skeleton seems to closely resemble the geometric structure of the current cosmic-web, and can be constructed in a fraction of the time taken by N-body simulations of the cosmic-web.

Part I

Quantum cosmology

Chapter 2

Geometrodynamics

> Perhaps the most impressive fact which emerges from the study of the quantum theory of gravity is that it is an extraordinary economical theory. It gives one just exactly what is needed in order to analyze a particular physical situation, but not a bit more. [It] is an operational theory *par excellence*.
>
> Bryce DeWitt

2.1 Introduction

In 1915, Albert Einstein formulated general relativity, a theory about the curvature of spacetime. The theory was, in subsequent years, successfully applied to a range of problems. Notwithstanding its great achievements in early years, general relativity, due to its technical complexity, only entered mainstream theoretical physics and astrophysics between approximately 1960 and 1975 – the *golden age of general relativity*. In this period following the discovery of the CMB and the Penrose-Hawking singularity theorems, gravitational waves, black holes, and the big bang singularity were recognized as physical predictions of Einstein's general relativity theory.

Georges Lemaître's investigation on the expanding universe and his knowledge of quantum mechanics led him, as early as 1931, to speculate about the *primeval atom* described by a theory of quantum spacetime. However, it took till the golden age of general relativity for John Wheeler to take up the challenge of quantum gravity. In his 1961 paper, *Geometrodynamics and the problem of motion* [339]– where geometrodynamics refers to general relativity – he states that:

> General relativity and the quantum theory of the atom, both born in World War I, and surely destined someday to be married in high state, have grown at very different rates, and raised different degrees of expectation at the several stages of their careers. The quantum principle was overshadowed at first by the drama of relativity; however, it has steadily grown in power and scope and usefulness, until now there is no part of physics that does not acknowledge its suzerainty. Relativity, on the other hand, quickly and dramatically encompassed the description of gravitation, of gravitational waves, and of the three still better known and more easily testable predictions of general relativity. Thereupon productivity in the field began to languish. Only in the last few years has it once more become widely appreciated that in general relativity Einstein gave a kind of master theory of physics out of which many deep meanings and rich physical consequences are still to be read.

In this, Wheeler gives the geometric nature of the 'master theory' central stage in his investigations on quantum gravity.

In the subsequent years, Wheeler is said to have attempted to construct a theory of quantum geometrodynamics [100]. It was clear to him that the theory should describe the dynamics of the *wave function of the universe* defined on *three-geometries*, short for the stratified manifold of three-dimensional Riemannian manifolds later coined *superspace*. It was during a two-hour layover at Raleigh-Durham airport, North Carolina, that John Wheeler invited Bryce DeWitt to discuss quantum gravity. During this meeting, Bryce DeWitt found a mathematical equation for Wheeler's vision. Following this discussion, Bryce DeWitt, in 1967, published a set of three papers on quantum gravity. The investigation was submitted as a single paper but was, by the editor of *Physical Review*, divided into three papers, sometimes referred to as the to 'trilogy' [96, 97, 98]. Paper I concerns the canonical theory, paper II the manifestly covariant theory, and paper III the applications of the covariant theory to cosmology.

These three papers have ever since been milestones in the development of canonical quantum gravity. Paper I concerns the equation which governs the wave function on superspace, known as *the Wheeler-DeWitt equation*. In the process of deriving this result, DeWitt realized a few key features of quantum geometrodynamics:

> Perhaps the most impressive fact which emerges from the study of the quantum theory of gravity is that it is an extraordinary economical theory. It gives one just exactly what is needed in order to analyze a particular physical situation, but not a bit more. Thus it will say nothing about time unless a clock to

> measure time is provided, and it will say nothing about geometry unless a device (either a material object, gravitational waves, or some other form of radiation) is introduced to tell when and where the geometry is to be measured. In view of the strongly operational foundations of both the quantum theory and general relativity, this is to be expected. When the two theories are united the result is an operational theory *par excellence*.

This feature at the same time illustrates a big problem. Since predictions can only be obtained while modeling the measuring device, one needs to develop a new interpretation of quantum physics to make the theory predictive. That is to say, if one wants to derive the classical evolution of spacetime, one first needs to define the correct quantum clock. Paper II and III concern a covariant formulation, which does not rely on the time-slicing of spacetime. DeWitt, in particular, proposes the Feynman path integral as a definition of the theory.

Over the years there has been a lot of attention directed towards completing Wheeler's program. An unambiguous theory of quantum geometrodynamics has however not yet been realized. This can mainly be attributed to a lack of a natural time parameter. In Wheeler's words, "so today it is called the Wheeler-DeWitt equation. But it is one thing to have an equation, another to solve it and still another to interpret the solution." DeWitt was more dismissive in his views. He in later years referred to the Wheeler-DeWitt equation, as "that damned equation" which according to him 'should be confined to the dustbin of history' [99]. He continuous by motivating his viewpoint:

> It has played a useful role in getting physicists to frame important and fundamental questions, but otherwise, I think it is a bad equation, for the following reasons:
>
> 1. By focussing on time slices it violates the very spirit of relativity.
> 2. Scores of man-years have been wasted by researchers trying to extract from it a natural time parameter.
> 3. Since good path integral techniques exist for basing Quantum Theory on gauge invariant observables only, it seems a pity to drag in the paraphernalia of constrained Hamiltonian systems.
>
> I subscribe 100% to the modern view that the quantum theory should be defined by the path integral. I am going to show you how the path integral can be used both to resolve the conceptual issues and to yield gauge invariant transition

> amplitudes that are operationally well defined. Except in special cases, these amplitudes do not satisfy any local differential equation. They satisfy the Wheeler-DeWitt equation only approximately. This means that, generically, the Wheeler-DeWitt equation is wrong, even assuming that the difficult issues of quantum gravity's perturbative nonrenormalizability can be resolved, via string theory or whatever. One may legitimately use the Wheeler-DeWitt equation, and the WKB approximations to its solution, in analyzing such things as the role of quantum fluctuations in the early universe. But it is wrong to use it as a definition of quantum gravity or as a basis for refined and detailed analyses.

DeWitt and Wheeler's initial work motivated many scientists to join the quest to quantize gravity. Over the last decades, there have been numerous proposals, including causal set theory, loop quantum gravity, and string theory. Even though quantum geometrodynamics has not yet delivered Wheeler's vision, the Wheeler-DeWitt equation remains central to many of the approaches. This follows from the fact that, in essence, the equation is nothing more than the quantum mechanical realization of the diffeomorphism invariance of general relativity. Any quantum gravity theory built on the geometric nature of general relativity will relate in some way to the equation. Notwithstanding Wheeler and DeWitt's criticisms, it remains important to better understand this 'damned equation'.

This thesis discusses the application of mathematical techniques to cosmology, a considerable part of it concerns the development of Wheeler's quantum geometrodynamics. In this chapter, we summarize general relativity and cosmology and give an introduction and overview of quantum geometrodynamics. The first part of the discussion on quantum geometrodynamics draws inspiration from Wheeler's paper *Superspace and the nature of quantum geometrodynamics* (1987) [340]. The second part is based on subsequent work by James Hartle, Stephen Hawking, Alexander Vilenkin and Claudio Bunster (originally known as Claudio Teitelboim).

2.2 General relativity

In (pseudo-)Riemannian geometry, the metric $g_{\mu\nu}$ encodes the distance between points via the infinitesimal line element

$$\mathrm{d}s^2 = \sum_{\mu,\nu} g_{\mu,\nu} \mathrm{d}x^\mu \mathrm{d}x^\nu \,. \tag{2.1}$$

In the subsequent discussions we always assume the Einstein summation convention and drop the summation symbol. A metric $g_{\mu\nu}$ describes a spacetime in general relativity when it extremizes the action

$$S = \int \left[\mathcal{L}_{EH} + \mathcal{L}\right] \sqrt{-g}\, \mathrm{d}^4x \tag{2.2}$$

with the determinant of the metric tensor $g = \det(g_{\mu\nu})$, the matter Lagrangian $\mathcal{L}$ and the Einstein-Hilbert Lagrangian

$$\mathcal{L}_{EH} = \frac{1}{2\kappa}(R - 2\Lambda)\,, \tag{2.3}$$

which describes the geometry of spacetime in terms of the Ricci scalar R, cosmological constant Λ and Einstein constant $\kappa = 8\pi G$ in terms of Newton's constant G. Upon variation with respect to the metric, we obtain the Einstein field equations which govern the dynamics of spacetime,

$$R_{\mu\nu} - \frac{1}{2}R g_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4} T_{\mu\nu}\,. \tag{2.4}$$

The left hand side describes the geometry of spacetime, with the Ricci curvature tensor $R_{\mu\nu}$. The right hand side describes the energy content in terms of the energy-momentum tensor

$$T_{\mu\nu} = 2\frac{\delta\mathcal{L}}{\delta g^{\mu\nu}} + g_{\mu\nu}\mathcal{L}\,. \tag{2.5}$$

Matter and radiation, described by the stress-tensor $T_{\mu\nu}$, experience gravity through the curvature of spacetime; a free falling particle follows the geodesic described by

$$\frac{\mathrm{d}U^\mu}{\mathrm{d}s} + \Gamma^\mu{}_{\alpha\beta} U^\alpha U^\beta = 0 \tag{2.6}$$

with the four-velocity of the particle U^μ and the Christoffel symbol

$$\Gamma^\mu{}_{\alpha\beta} = \frac{1}{2} g^{\mu\lambda}\left(\partial_\alpha g_{\beta\lambda} + \partial_\beta g_{\alpha\lambda} - \partial_\lambda g_{\alpha\beta}\right)\,. \tag{2.7}$$

In John Wheeler's words: "matter [and radiation] tells spacetime how to curve and curved spacetime tells matter how to move".

In Einstein's vision, spacetime is radically different from what came before. It is a four-dimensional entity describing our universe, in which space and time are 'rotated' into

each other by the Lorentz transformations. There generally does not exist a notion of simultaneity or a preferred time-foliation. Moreover, the familiar conservation laws – such as conservation of energy and momentum – only exist when spacetime exhibits a special symmetry, *i.e.*, the spacetime has a Killing vector field. A Killing vector field ξ is a four-dimensional vector field which preserves the metric, *i.e.*,

$$0 = \mathcal{L}_\xi g_{\mu\nu} = \xi^\sigma \partial_\sigma g_{\mu\nu} + h_{\sigma\nu}\partial_\mu \xi^\sigma + g_{\mu\nu}\partial_\sigma \xi^\sigma \,, \tag{2.8}$$

with $\mathcal{L}_\xi$ the Lie derivative in the direction ξ. These properties make general relativity, more than a hundred years after its inception, an active field of investigation. For more details see [335, 216].

2.3 Cosmology

In cosmology, the Einstein field equations are simplified by assuming the universe to be spatially homogeneous. Spatial homogeneous spacetimes are classified by *Bianchi's classification of Lie algebras* [44, 253, 129, 243, 333, 279, 216]. Consider the class of spacetimes which allow a time-foliation

$$\mathrm{d}s^2 = -\mathrm{d}t^2 + h_{ij}\mathrm{d}x^i \mathrm{d}x^j \,, \tag{2.9}$$

with h_{ij} the space metric. When the space is homogeneous, there exist three linearly independent spatial Killing vector fields ξ_A, $A = 1, 2, 3$, satisfying the relation

$$\mathcal{L}_{\xi_A} h_{ij} = 0 \,, \tag{2.10}$$

with $\mathcal{L}_v$ the Lie derivative in the direction v. The commutator of two of the three Killing vector fields

$$[\xi_A, \xi_B]^j = \xi_A^i \partial_i \xi_B^j - \xi_B^i \partial_i \xi_A^j \,, \tag{2.11}$$

is again a Killing vector field, *i.e.*,

$$[\xi_A, \xi_B] = C_{AB}^D \xi_D \tag{2.12}$$

with C_{AB}^D a set of numbers known as the structure constants. Formally, the commutator defines a Lie algebra. The Bianchi classification of Lie algebras amounts to classifying inequivalent algebras, characterized by the structure constants C_{AB}^D. For three-dimensional homogeneous space, the classification consists of eight discrete classes (known as types I, II, IV, V, VI_0, VII_0, VIII and IX) and two continuous families (type VI_h and VII_h). Each type describes an inequivalent homogeneous cosmology.

2.3.1 Isotropic homogeneous cosmologies

An expanding homogeneous universe tends towards isotropy [334]. As a result, the current universe is both highly isotropic and homogeneous. The isotropic and homogeneous cosmologies (type I, VII_0, V, VII_h and IX) are described by the Roberson-Walker metric

$$\mathrm{d}s^2 = \mathrm{d}t^2 - a^2(t) h_{ij} \mathrm{d}x^i \mathrm{d}x^j \tag{2.13}$$

with the scale factor of the universe $a(t)$ and the metric on three-space h_{ij} given by

$$h_{ij} = \delta_{ij} + \frac{k x_i x_j}{1 - k(x_k x^k)} \,. \tag{2.14}$$

The index $k = -1, 0, 1$ represents the curvature of space, with $k = 0$ for Euclidean flat space, $k = +1$ for spherically curved and $k = -1$ for hyperbolically curved space.

When the energy content consists of a isotropic and homogeneous perfect fluid with the energy-momentum tensor

$$T_{\mu\nu} = (\rho + P) U_\mu U_\nu - P g_{\mu\nu} \tag{2.15}$$

with the density ρ, the pressure P and four-velocity U_μ, the Einstein equations reduce to the Friedmann equations

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{k}{a^2} \,, \tag{2.16}$$

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3P) \,. \tag{2.17}$$

The first Friedmann equation describes the evolution of the scale factor, or equivalently the Hubble parameter $H = \frac{\dot{a}}{a}$, as a function of time. The second Friedmann equation describes the evolution of the energy content as the universe evolves.

When the fluid is characterized by the equation of state $P = \omega\rho$, with $\omega = 0$ for matter, $\omega = \frac{1}{3}$ for radiation and $\omega = -1$ for dark energy, the second Friedmann equation implies that the density of the fluid evolves as

$$\rho \propto a^{-3(1+\omega)} \,. \tag{2.18}$$

Using these scaling relations, and normalizing the scale factor such that $a(t_0) = 1$ with t_0 the age of the universe, the first Friedmann equation can be written as

$$H^2 = H_0^2 \left[\Omega_r a^{-4} + \Omega_m a^{-3} + \Omega_k a^{-2} + \Omega_\Lambda\right] \tag{2.19}$$

Type of universe	ρ	ω	$a(t)$
Flat radiation-dominated	a^{-4}	$\frac{1}{3}$	$t^{1/2}$
Flat matter-dominated	a^{-3}	0	$t^{2/3}$
Flat cosmological-constant	a^0	-1	e^{Ht}

Table 2.1: The behaviour of the radiation-, matter- and cosmological constant-dominated flat $k = 0$ FLRW universe.

where the evolution of the Hubble parameter H is expressed in terms of the current Hubble parameter H_0 and the dimensionless density parameters Ω_I with $I = r, m, k, \Lambda$ referring to radiation, mater, curvature and the cosmological constant, at the current time. The dimensionless density parameters are defined by

$$\Omega_I = \frac{\rho_{I,0}}{\rho_{crit}} \tag{2.20}$$

with the current density $\rho_{I,0}$ and the critical density

$$\rho_{crit} = \frac{3H_0^2}{8\pi G}\,. \tag{2.21}$$

The critical density ρ_{crit} is the density required for the universe to be spatially flat. In this equation the "curvature" density parameter is given by

$$\Omega_k = -\frac{k}{H_0^2}\,, \tag{2.22}$$

and the different energy sources are defined to sum to unity

$$\Omega_r + \Omega_m + \Omega_k + \Omega_\Lambda = 1\,. \tag{2.23}$$

A general closed-form solution for the first Friedmann equation does not exist. When the energy content is dominated by one or two components, it is often possible to find exact solutions revealing the characteristic behavior of the universe. In particular, for the flat radiation-dominated, matter-dominated and cosmological constant-dominated universe, *i.e.*, one of the three $\Omega_r, \Omega_m, \Omega_\Lambda$ equals one and $\Omega_k = 0$, the solutions are described in Table 2.1. The range of different isotropic and homogeneous cosmologies are illustrated in figure 2.1.

Observations suggest that we life in a ΛCDM universe with currently $\Omega_r \approx 0$, $\Omega_m \approx 0.32$, $\Omega_k \approx 0$ and $\Omega_\Lambda \approx 0.68$. Starting from the big bang, at scale factor $a = 0$, in

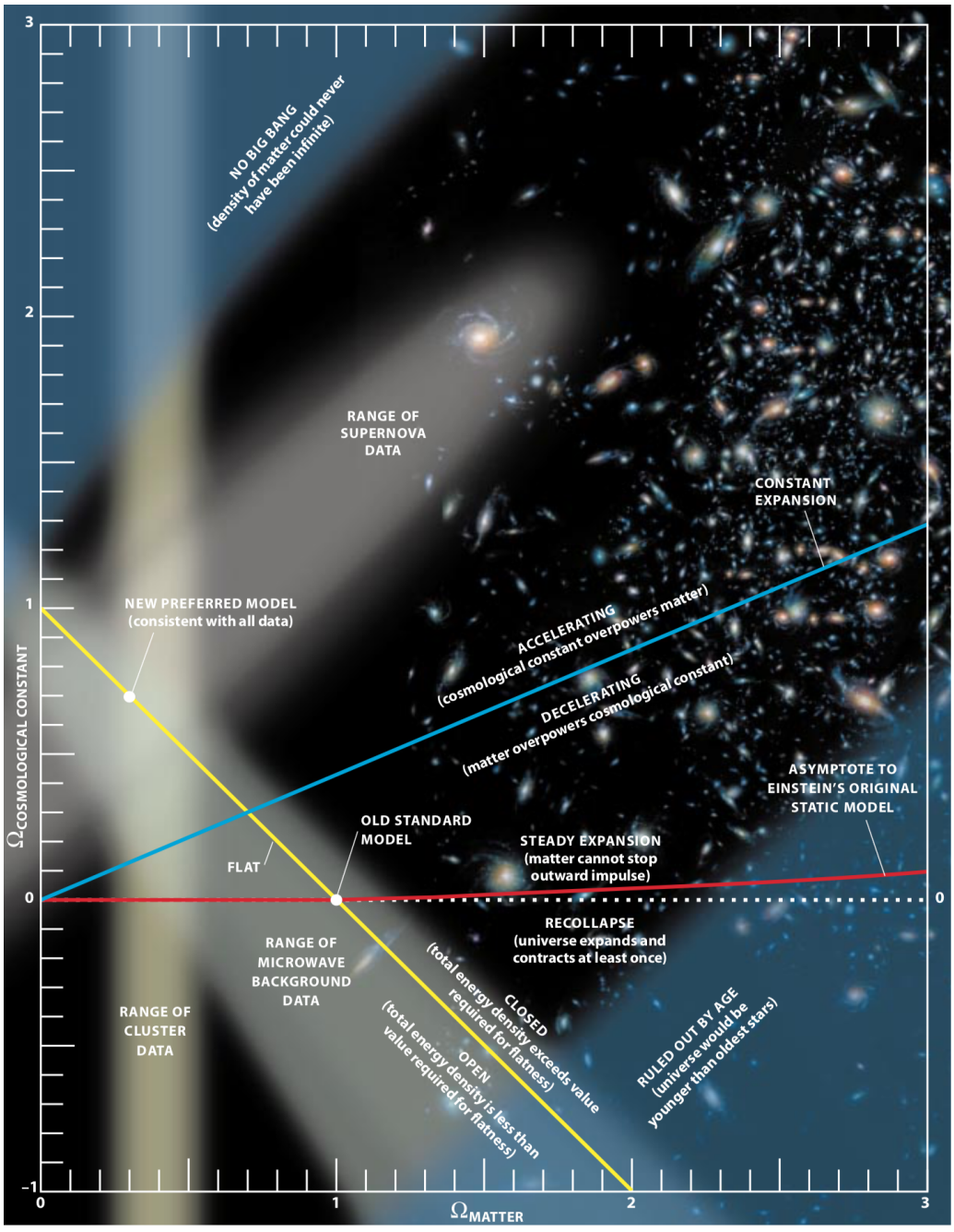


Figure 2.1: The parameter space of FLRW universes as a function of Ω_m and Ω_Λ. It is assumed that $\Omega_r = 0$. Figure from [227].

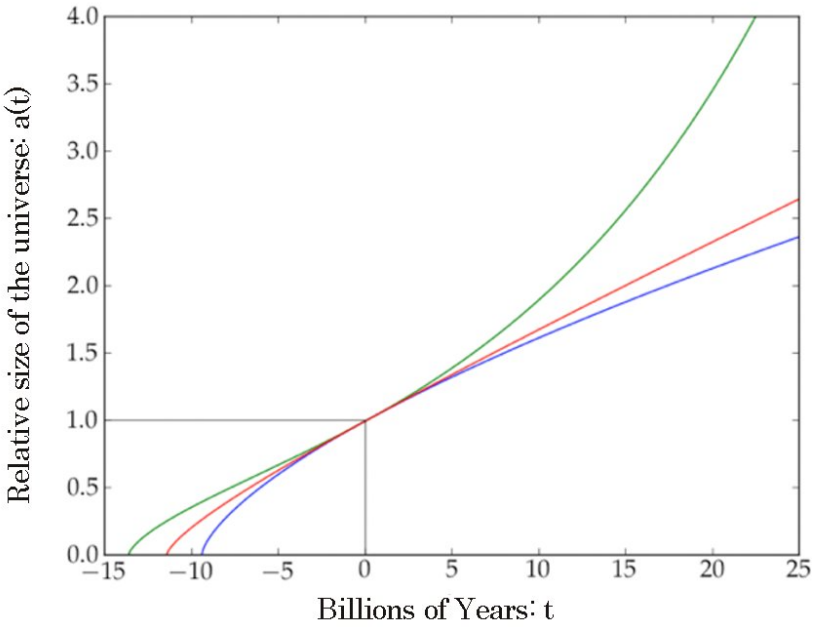


Figure 2.2: Three solutions of the Friedmann equations. The blue curve represents flat matter dominated Einstein-de Sitter universe $\Omega_m = 1$, and $\Omega_\Lambda = 0$. The red curve represents the lighter hyperbolically curved $\Omega_m = 0.3$, and $\Omega_\Lambda = 0$. The green curve is the ΛCDM universe with $\Omega_m = 0.3$, and $\Omega_\Lambda = 0.7$.

the flat case $k = 0$, the early universe is dominated by the radiation. In this phase, the universe expands like $a \propto t^{1/2}$. When the matter content takes over from the radiation, the expansion decelerates to $a \propto t^{2/3}$. When the energy content of the universe includes dark energy or a cosmological constant, Λ, the dark energy will eventually take over from the matter leading to an exponential expansion $a(t) \propto e^{Ht}$. The typical evolution is plotted in Figure 2.2.

2.3.2 Anisotropic homogeneous cosmologies

While the anisotropies are negligible in the late time universe, they are significant at the beginning. They for this reason often feature in studies of quantum cosmology attempting to study the big bang singularity. It is common to describe anisotropic homogeneous solutions of the Einstein field equations, by separating the overall volume expansion of space from the anisotropic effects. We consider for the time-foliation, for which the normal

vector n_ν to the spatial hypersurfaces is a geodesic, irrotational congruence, *i.e.*,

$$\nabla_\mu n_\nu = \sigma_{\mu\nu} + H h_{\mu\nu} \,, \tag{2.24}$$

with the covariant derivative ∇_μ with respect to the spatial metric $h_{\mu\nu}$, the Hubble parameter for the isotropic expansion H, and the symmetric trace-free shear tensor representing volume deformation $\sigma_{\mu\nu}$. In isotropic cosmologies, the deformation tensor $\sigma_{\mu\nu}$ vanishes.

For an anisotropic universe consisting of a perfect fluid, the Friedmann equation generalizes to

$$H^2 = \frac{1}{3}\left[8\pi G\rho + \sigma^2 - {}^3R/2 + \Lambda\right] \,, \tag{2.25}$$

with scalar shear $\sigma^2 = \sigma_{\mu\nu}\sigma^{\mu\nu}/2$ and ${}^3R$ the scalar Ricci 3-curvature. The scalar shear term sources the isotropic expansion rate of the universe. The Einstein field equations also describe how the shear $\sigma_{\mu\nu}$ is affected by the isotropic expansion,

$$h^\sigma_\mu h^\tau_\nu \dot{\sigma}_{\sigma\tau} = -3H\sigma_{\mu\nu} - {}^3S_{\mu\nu} + \pi_{\mu\nu} \,, \tag{2.26}$$

where the dot denotes the covariant derivative along the time like direction n^μ, the trace-free part of the spatial Ricci tensor ${}^3S_{\mu\nu}$ and the trace-free spatial part of the energy-momentum tensor $\pi_{\mu\nu}$.

Anisotropic Bianchi cosmologies, described by equations (2.25) and (2.26) can exhibit complicated behaviour. Consider in particular Misner's (1969) [253] mixmaster universe with many sharp bounces and exhibiting quantum chaos. Between the bounces, it is often argued that there is little time for information to propagate. In this regime, the curvature term can be ignored simplifying the equations. The corresponding behavior is known as a Kasner phase.

For a detailed analysis of anisotropic homogeneous cosmological solutions of the Einstein field equations see [333].

2.4 Quantum geometrodynamics

Wheeler developed quantum geometrodynamics by extending his geometric understanding of quantum mechanics, in terms of constructive interference of oscillatory integrals, to general relativity. We first discuss the emergence of classical trajectories in quantum mechanics. We subsequently describe Wheeler's vision on the quantization of gravity, in which classical spacetime emerges from a wave function on superspace (following Wheeler [340]). We describe the Wheeler-DeWitt equation and the path integral for gravity, which can be interpreted as two independent attempts to realize Wheeler's vision.

2.4.1 Time evolution in quantum mechanics

The world-line of a classical particle $x(t)$ completely describes its evolution. A non-relativistic particle moving in the potential $V(x)$ can be studied with the Hamilton-Jacobi equation

$$-\frac{\partial S}{\partial t} = H\left[x, \frac{\partial S}{\partial x}\right], \tag{2.27}$$

with the canonical momentum

$$p = \frac{\partial S}{\partial x}, \tag{2.28}$$

where the Hamiltonian takes the simple form

$$H[x,p] = \frac{p^2}{2m} + V(x), \tag{2.29}$$

with the canonical momentum p and the Hamilton-Jacobi function or classical action S, *i.e.*,

$$S = \int \left[\frac{1}{2}m\dot{x}^2(t) - V[x(t)]\right] \mathrm{d}t\,. \tag{2.30}$$

When entering the quantum world, the wave function ψ replaces the world-line $x(t)$. Many quantum mechanical problems are answered using the time-independent Schrödinger equation

$$\hat{H}\psi_E(x,t) = E\psi_E(x,t), \tag{2.31}$$

with the energy E and the Hamiltonian operator

$$\hat{H} = \frac{-\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x), \tag{2.32}$$

obtained from the classical Hamiltonian using the substitution $p \mapsto -i\hbar\frac{\partial}{\partial x}$. In the semi-classical approximation, the wave function takes the form

$$\psi_E(x,t) = (\text{slowly varying amplitude}) \times e^{\frac{i}{\hbar}S_E(x,t)} \tag{2.33}$$

with – to leading order in $\hbar$ – the Hamilton-Jacobi function S_E at energy E, satisfying

$$-\frac{\partial S_E}{\partial t} = H\left[x, \frac{\partial S}{\partial x}\right] \tag{2.34}$$

$$= \frac{1}{2m}\left(\frac{\partial S}{\partial x}\right)^2 + V(x) \tag{2.35}$$

$$= E\,. \tag{2.36}$$

For scattering states there exist continuous family of solutions as a function of the energy E. For bound states there exists a discrete spectrum of solutions as a function of the energy E.

This equation is solved by

$$S_E(x,t) = -Et \pm \int^x \sqrt{2m(E - V(x))}\mathrm{d}x\,, \tag{2.37}$$

where the sign of the square root corresponds corresponds to the direction of the momentum [1]. However, the wave function ψ_E does not describe the dynamics of the particle. There is no notion of causality or initial and final state. Complementarity forbids it! Since the energy E is specified, the particle is uniformly spread over all times as can be seen from S_E. We only recover a world-line $x(t)$ from the wave function ψ_E after superposing these energy eigenstates into a wave packet

$$\psi_t(x) = \int f(E)\psi_E(x,t)\mathrm{d}E \tag{2.38}$$

for some function $f(E)$ describing the localisation[2]. The resulting wave function satisfies the time-dependent Schrödinger equation

$$i\hbar\frac{\partial \psi_t(x)}{\partial t} = \hat{H}\psi_t(x) \tag{2.39}$$

$$= \left[\frac{-\hbar^2}{2m}\frac{\partial^2}{\partial x^2} + V(x)\right]\psi_t(x)\,. \tag{2.40}$$

Instead of definite knowledge, we obtain a probability density for the position

$$p_t(x) = |\psi_t(x)|^2\,, \tag{2.41}$$

[1]The equation $E = \frac{p^2}{2m} + V(x)$ has two solutions for p, since the energy is independent of the direction of the momentum. The energy states are thus degenerate.

[2]When the energy spectrum includes a discrete part, the integral over these energies is interpreted as a sum.

at time t. The world-line emerges in the classical limit[3] $\hbar \to 0$ – by constructive interference of the highly oscillatory integral (2.38) – at a critical point of the Hamilton-Jacobi function

$$\frac{\mathrm{d}}{\mathrm{d}E}\left[S_E(x,t) - i\hbar \log(f(E))\right] - t = 0. \tag{2.42}$$

For finite $\hbar$, we obtain an approximate notion of a trajectory (see Figure 2.3); the probability density $p_t(x)$ is blurred around the classical trajectory $x(t)$.

The same phenomenon is observed in the propagator formulation. In chapter 4 of Feynman and Hibbs [150], they define the wave function at time $t > 0$ as

$$\psi_t(x) = \int_{-\infty}^{\infty} G[x_1; x_0; t]\psi_0(x_0)\mathrm{d}x_0\,, \tag{2.43}$$

in terms of the wave function ψ_0 at time $t = 0$, where $G[x_1; x_0; t]$ is the Feynman propagator expressed by the path integral

$$G[x_1; x_0; t] = \Theta(t)\langle x_1|e^{-i\hat{H}t/\hbar}|x_0\rangle \tag{2.44}$$

$$= \Theta(t)\int_{x(0)=x_0}^{x(t)=x} e^{\frac{i}{\hbar}S[x(t)]}\mathcal{D}x \tag{2.45}$$

with S the action of the particle, and Θ the Heaviside step function. The propagator is a Green's function of the Schrödinger operator, *i.e.*,

$$\left[i\hbar\frac{\partial}{\partial t} - \hat{H}\right] G[x_1; x_0; t] = i\delta(x_0 - x_1)\delta(t)\,, \tag{2.46}$$

which Feynman with abuse of notation also calls the Schrödinger equation. The propagator is in addition defined by the boundary condition

$$\lim_{t\to 0} G[x_1; x_0; t] = \delta(x_0 - x_1)\,. \tag{2.47}$$

The Schrödinger equation illustrates the fact that the time t is really the proper-time of the particle[4], *i.e.*, the propagator represents the amplitude the particle to travel from the initial point x_0 to the final point x_1. There exist no classical paths going from x_0 to x_1,

[3]The Planck constant is a dimensionful constant. The classical limit $\hbar \to 0$, should always be interpreted in terms of some fixed units of the system. In many cases it means that Planck constant $\hbar$ is small compared to the time integral of the potential along the classical path $\int V[x(t)]\mathrm{d}t$.

[4]The proper-time and coordinate time of a particle coincide in the classical limit. The relativistic covariant Dirac delta function $\delta(x_0 - x_1)\delta(t)$ is inherited from the relativistic quantum particle.

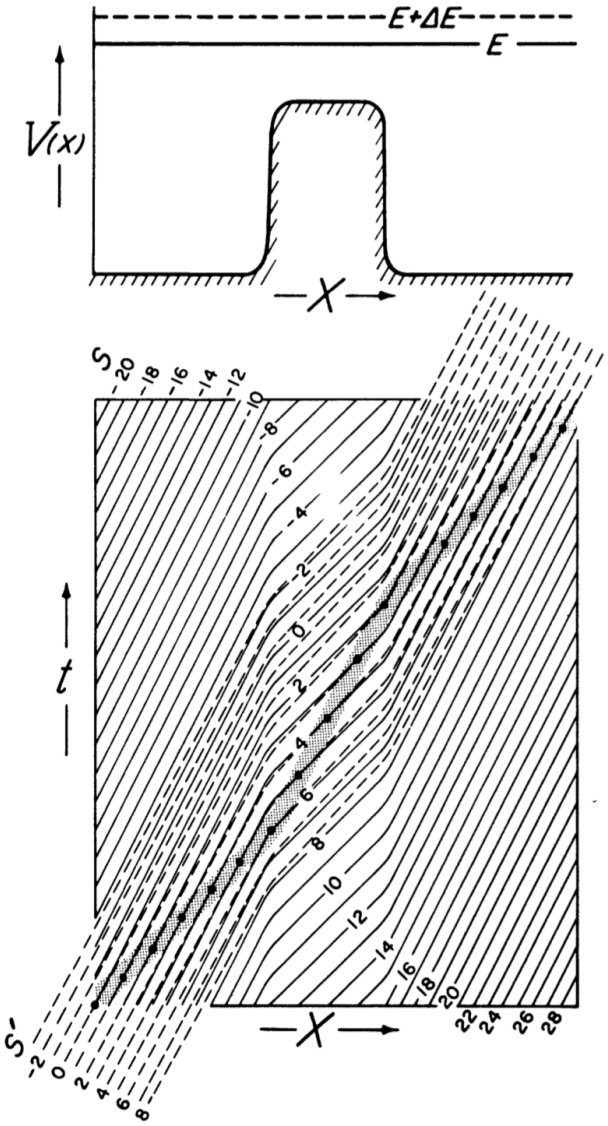
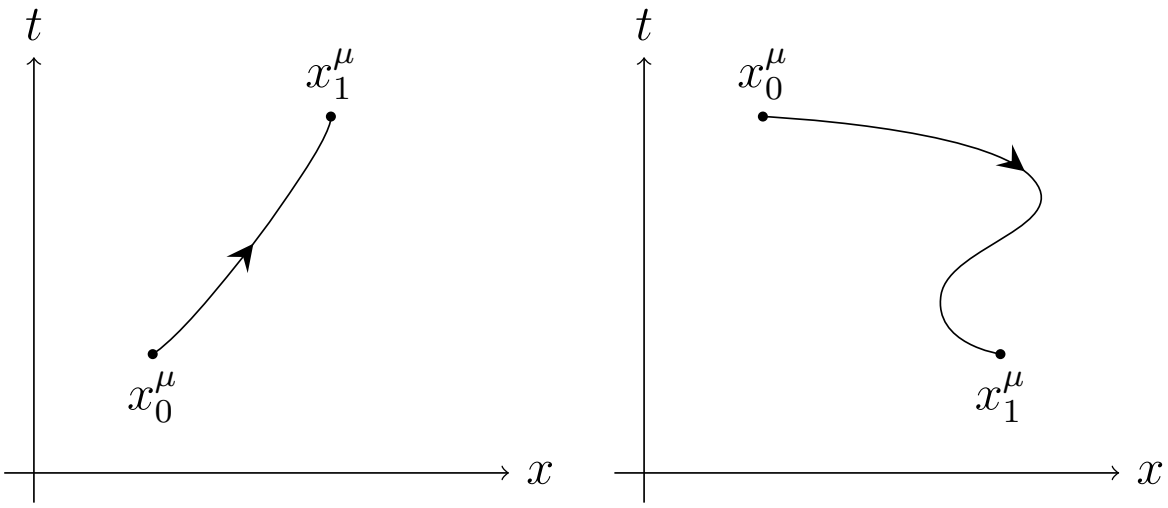


Figure 2.3: The emergence of a classical world-line from a wave packet in the classical limit. The upper figure is the potential V, in which the energy E of the particle is plotted. The lower figure represents the wavefronts as a function of the energy. In this figure, the marked region represents the classical trajectory for which the Hamilton-Jacobi function S_E is stationary. Since the wave packed is not localized with unlimited sharpness, we only obtain an approximate world-line. Figure from [340].



(a) A particle

(b) An antiparticle

Figure 2.4: A forward (left panel) and a backward (right panel) propagating particle. In classical physics the proper time of a particle is identified with the coordinate time and the particles cannot turn around in coordinate time. It for this reason suffices to work with either only forward or only backwards propagating particles.

with negative proper-time. Note that in the relativistic theory it is possible for a particle to travel backwards in coordinate time. Such a particle is identified as an antiparticle (see Figure 2.4). The vanishing propagator for negative proper-time, is in particular a useful property in perturbation theory, where a particle while propagating from x_0 to x_1 can scatter of the potential V at a number of positions y_i for $i = 1, 2, \ldots, N$ which need to be ordered in time to ensure that the particle always moves forward in coordinate time (see chapter 6 of Feynman and Hibbs [150]).

In terms of energy eigenstates, the propagator reads

$$G[x_1; x_0; t] = \Theta(t) \sum_n \phi_n(x_1)\phi_n(x_0)e^{-iE_n t/\hbar}, \tag{2.48}$$

where ϕ_n are solutions of the time-independent Schrödinger equation, $\hat{H}\phi_n = E_n\phi_n$, and the sum ranges over the energy eigenstates of the Hamiltonian $\hat{H}$. Here $\phi_n(x) = \langle x|E_n\rangle$ is the amplitude that if we are in energy state E_n, we are at position x, and $\phi_n^*(x) = \langle E_n|x\rangle$ is the amplitude that if we are at x, we are in energy state E_n. This sum is a proper path integral since:

1. $\phi_n^*(x_0)$ is the probability that if we are at x_0 we are in the energy state n,
2. $e^{-iE_nt/\hbar}$ is the amplitude to travel in energy state n at time t given that we where in energy state E_n at time $t = 0$,
3. $\phi_n(x_1)$ is the amplitude that we are found at x_1 when we are at energy state n

where we sum over the energy eigenstates (see Feynman and Hibbs chapter 5 [150]).

Now note that in the propagator formulation, the wave function no longer solves the Schrödinger equation

$$\left[i\hbar\frac{\partial}{\partial t} - \hat{H}\right]\psi_t(x) = -i\int_{-\infty}^{\infty}\delta(x-x_0)\delta(t)\psi_0(x_0)\mathrm{d}x_0 \tag{2.49}$$

$$= -i\psi_0(x)\delta(t)\,, \tag{2.50}$$

giving back the initial state at the time slice $t = 0$. Surprisingly, this property is not mentioned in Feynman and Hibbs [150]. This is normally not considered to be an important point as classical particles always travel forward in time. However, for relativistic theories we will show that this does play a key role in the formalism (see chapter 7).

Finally note that in the propagator formation, the path integral interferes constructively for paths close to the stationary point of the integrand, *i.e.*,

$$\frac{\delta S}{\delta x} = 0\,, \quad \frac{\partial}{\partial x_0}\left[\frac{i}{\hbar}S[x(t)] + \ln\psi_0(x_0)\right] = 0\,. \tag{2.51}$$

The stationary points are classical solution, as the first variation leads to the equation of motion

$$m\ddot{x}(t) = -V'[x(t)]\,, \tag{2.52}$$

where the prime a derivative with respect to position, and second variation specifies the initial conditions of the particle, *e.g.* for a Gaussian initial state

$$\psi_0(x) = \frac{1}{\sqrt[4]{2\pi\sigma^2}}e^{-\frac{(x-x_i)^2}{4\sigma^2}+i\frac{(x-x_i)p_i}{\hbar}} \tag{2.53}$$

centred at position x_i with momentum p_i this implements the boundary conditions

$$x(0) = x_i\,, \quad m\dot{x}(0) = p_i\,. \tag{2.54}$$

Note that the classical solution, following from the path integral, does not need to be real-valued. When no real-valued solution, interpolating between $x(0) = x_0$ and $x(1) = x_1$ exists, there generically will exist at least one complex classical trajectory describing a tunnelling phenomenon. This is extensively discussed in chapter 7.

2.4.2 Wheeler's vision on quantum geometrodynamics

In order to construct a theory of quantum spacetime, we compare the classical particle with classical spacetime (see Table 2.2 for an overview of the comparison). The world-line of a particle $x(t)$ in spacetime M consists of a succession of spacetime points $(t, x) \in M$, *i.e.*, a path can be interpreted as the set

$$\{(t, x') | x(t) = x'\} \subset M\,. \tag{2.55}$$

Note that the coordinate time guarantees the reconstruction of the world-line trajectory from the set of spacetime points.

Spacetime itself, consists – given a foliation – of the succession of three-dimensional space manifolds, known as *three-geometries* ${}^3\mathcal{G}$. The collection of all possible three-dimensional Riemannian spaces forms a stratified manifold known a superspace $\mathcal{S}$. The relation between the three-geometries and the spacetime manifold is however more intricate. A spacetime manifold allows many spacetime foliations with their unique set of three-geometries. Instead of an ordered set of three-geometries, corresponding to one specific foliation, we interpret the spacetime manifold M as the collection of all its three-dimensional space-like slices

$$\{{}^3\mathcal{G} |\text{ the three-geometry } {}^3\mathcal{G} \text{ is a slice of the spacetime manifold } M\} \subset \mathcal{S}\,. \tag{2.56}$$

Since this set does not have a natural ordering, it might at first seem impossible to reconstruct the spacetime manifold M. However, the three-geometries of different foliations interlock, guaranteeing the reconstruction of the spacetime manifold M.

Remark that, whereas Einstein's formulation of general relativity emphasizes the four-dimensional pseudo-Riemannian manifolds, the geometrodynamics formulation places emphasis on sets of three-dimensional Riemannian manifolds. The superspace formulation of general relativity might, at first sight, seem hopelessly overcomplicated. However, its similarity with classical mechanics makes it a fruitful perspective for the development of quantum gravity.

The wave function of the particle $\psi_t(x)$ is defined on the space manifold. It is natural to assume the theory of quantum gravity to mirror this by defining *the wave function of the universe* $\psi({}^3\mathcal{G})$ on the superspace stratified manifold $\mathcal{S}$. Whereas $|\psi_t(x)|^2$, describes the likelihood for the particle to be at (t, x), we postulate that $|\psi({}^3\mathcal{G})|^2$ describes the likelihood of the three-geometry ${}^3\mathcal{G}$ to be part of the evolution of the universe. In quantum mechanics, the world-line is not fundamental and emerges from constructive interference. The same

	Quantum particle	Quantum universe
Domain	spacetime: M	superspace: $\mathcal{S}$
Point	spacetime event: $x^\mu = (t, x)$	three-geometry: ${}^3\mathcal{G}$
wave function	of trajectory: $\psi(x^\mu)$	of universe: $\psi({}^3\mathcal{G})$
Ordering	ordering of space points x	interlocking foliations
Dynamics	Schrödinger equation	Wheeler-DeWitt equation

Table 2.2: Comparison between the relativistic quantum particle and the quantum universe.

should apply to spacetime in quantum gravity. While the interlocking of three-geometries guarantees the reconstruction of the spacetime manifold in classical geometrodynamics, we can in the quantum theory at best obtain an approximate notion of spacetime since the wave packet does not have an unlimited sharpness (see Figure 2.5). The predicted fluctuations are quantum fluctuations of the geometry of spacetime. This closely follows Lemaître's early vision on quantum spacetime in which space and time are only defined in the classical realm.

2.4.3 The Hamiltonian formulation of general relativity

This must have roughly been the state of Wheeler's vision on quantum geometrodynamics at the start of John Wheeler and Bryce DeWitt's meeting at Raleigh-Durham airport in North Carolina in 1967. During this meeting, Bryce DeWitt used the Hamiltonian formulation of general relativity to propose an equation governing the wave function on superspace. This equation is known as the Wheeler-DeWitt equation. In the subsequent years, he extended the treatment by proposing a covariant formulation in terms of Feynman's sum over histories formulation of quantum physics [96, 97, 98].

The Hamiltonian formulation of general relativity was first developed by Paul Dirac [113]. In 1958, he realized that while the Lagrangian formalization by Einstein and Hilbert empathizes four-dimensional pseudo-Riemannian spacetime with the Lorentz symmetries, the Hamiltonian formulation suggests the importance of three-dimensional Riemannian space. This turned out to be remarkably close to Wheeler's intuitive view of geometrodynamics.

The Lagrangian formulation of classical physics is global in time and is formulated in terms of the trajectories $x(t)$. Space and time are placed on the same footing. The Hamiltonian formulation is local in time and describes the evolution of the particle step

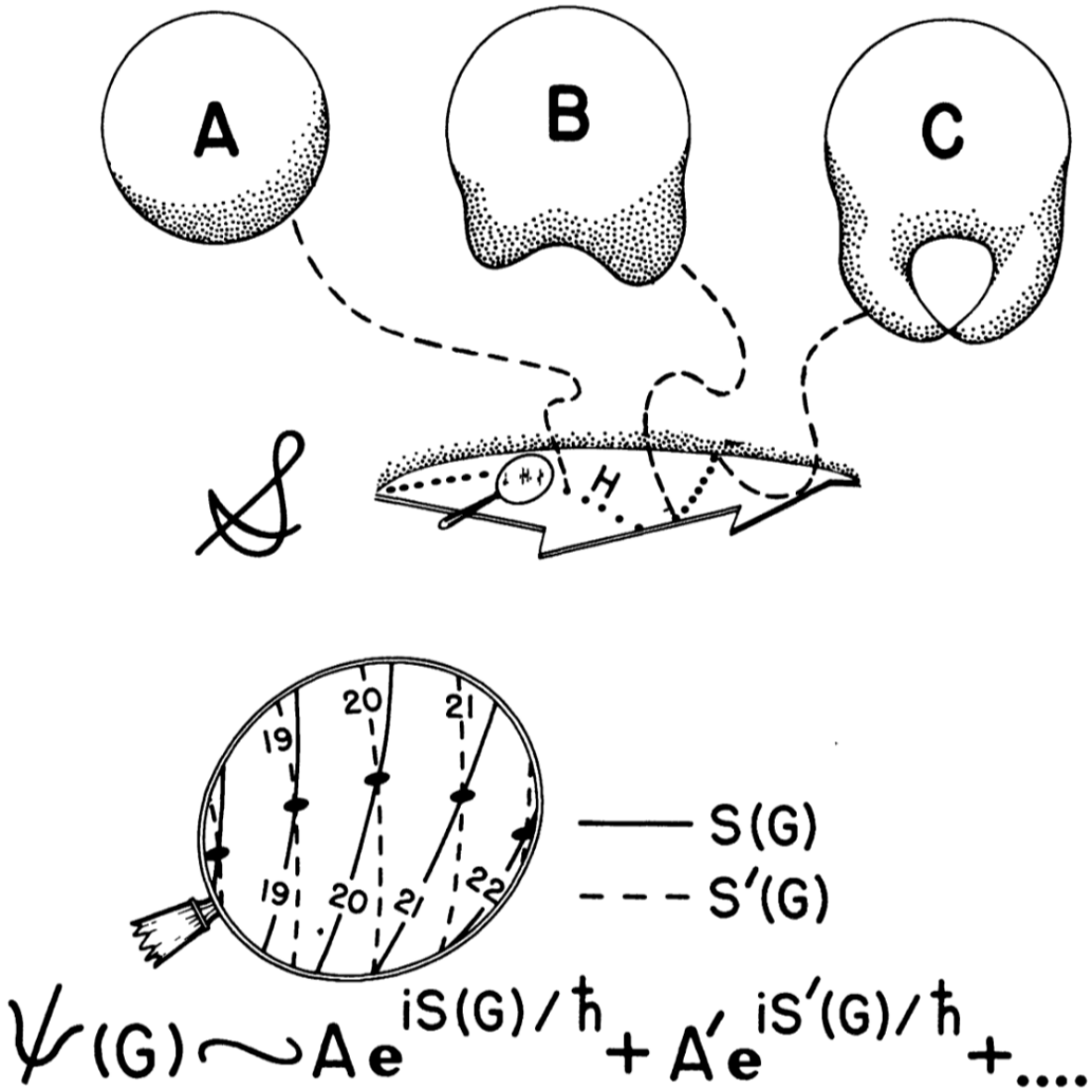


Figure 2.5: A section of superspace, in which A, B and C are three examples of three-geometries which are part of the sub-manifold H consisting of all those spacelike three-geometries that can be obtained as spacelike sections through a particular spacetime manifold. In quantum geometrodynamics, the wave function of the universe ψ is enhanced at the classical manifold H. However, since in quantum mechanics the wave packet is not localized with unlimited sharpness, the amplitude ψ is significant in a small region around H. These are the quantum fluctuations in the geometry of spacetime. Figure from [340].

by step. By analogy, the Hamiltonian formulation of general relativity is naturally defined on a foliation of spacetime.

In deriving the Hamiltonian formulation of general relativity, we start with the ADM metric (named after Richard Arnowitt, Stanley Deser and Charles Misner),

$$\mathrm{d}s^2 = -((h^{1/2}N^0)^2 - N_iN^i)\mathrm{d}\tau^2 + 2N_i\mathrm{d}x^i\mathrm{d}t + h_{ij}\mathrm{d}x^i\mathrm{d}x^j\,, \tag{2.57}$$

with the three-dimensional metric h_{ij} [24]. The lapse N^0 and the shift vector N^i parametrize the foliation. By varying the Einstein-Hilbert Lagrangian, expressed in terms of the ADM metric, with respect to $\dot{h}_{ij}$ (with the over dot a derivative with respect to τ) we obtain the canonical momentum

$$\pi^{ij} = \frac{\partial\mathcal{L}_{EH}}{\partial\dot{h}_{ij}} \tag{2.58}$$

$$= \frac{N_{i|j} + N_{j|i} - h_{ij,0}}{2\sqrt{h}N^0}\,, \tag{2.59}$$

where the $|$ denotes the covariant derivative with respect to the three-metric h_{ij}. Varying the Lagrangian with respect to $\dot{N}^0$ and $\dot{N}^i$ shows that the lapse N^0 and the shift vector N^i are Lagrange multipliers. Using the Legendre transformation, with respect to the canonical pair (h_{ij}, π^{ij}), we obtain the Hamiltonian density

$$\mathcal{H} = N^0\mathcal{H}_0 + N^i\mathcal{H}_i \tag{2.60}$$

where the Hamiltonians associated with the temporal and spatial directions are given by

$$\mathcal{H}_0 = (2\kappa)G_{ijkl}\pi^{ij}\pi^{kl} + (2\kappa)^{-1}h(2\Lambda - {}^3R)\,, \tag{2.61}$$

$$\mathcal{H}_i = -2h_{ij}\pi^{jk}{}_{|k}\,, \tag{2.62}$$

with the spatial Ricci scalar ${}^3R$ and *the metric tensor on superspace*

$$G_{ijkl} = \frac{1}{2}\left[h_{il}h_{jk} + h_{ik}h_{jl} - h_{ij}h_{kl}\right]\,. \tag{2.63}$$

The phase-space formulation of the Einstein-Hilbert action takes the simple form

$$S_{EH} = \int \mathcal{L}_{EH}\sqrt{-g}\,\mathrm{d}^4x \tag{2.64}$$

$$= \int_{\tau_0}^{\tau_1}\int \left[\pi^{ij}\dot{h}_{ij} - N^0\mathcal{H}_0 - N^i\mathcal{H}_i\right]\mathrm{d}^3x\,\mathrm{d}\tau\,. \tag{2.65}$$

This formulation of the Einstein-Hilbert action shows that the lapse N^0 and shift vector N^i are indeed Lagrange multipliers, enforcing the four constraints

$$\mathcal{H}_0 = 0\,, \quad \mathcal{H}_i = 0\,. \tag{2.66}$$

The constraint $\mathcal{H}_0 = 0$, known as the Hamiltonian constraint, is similar to the Hamiltonian of a relativistic particle on curved spacetime. The constraint is quadratic in momentum and couples to the three-metric via the metric on superspace. The three-metric h_{ij} consists of six components, which can be labelled as h_A with $A \in \{11, 12, 13, 22, 23, 33\}$. In this notation, the metric on superspace G_{AB} is six-dimensional, consisting of one time-like and five space-like directions (the metric as the signature $(-,+,+,+,+,+)$). The second term in the constraint can be viewed as a potential term. The constraint $\mathcal{H}_i = 0$, known as the momentum constraint, is of a different form, being only linear in momentum. The total Hamiltonian $\mathcal{H}$ vanishes for physical spacetimes satisfying the Einstein field equations.

The observation that the Einstein field equations are equivalent to a set of Hamiltonian constraints can be traced back to the geometric nature of the general relativity. The Lagrangian formulation is manifestly diffeomorphism invariant; the Einstein-Hilbert action is invariant of the choice of coordinates. The Hamiltonian formalism, on the other hand, requires a choice of time-slicing. The redundancy of this choice translates into the four constraints corresponding to the four dimensions of spacetime. The same can be observed in the Hamiltonian formulation of gauge theories such as Maxwell or Yang-Mills theory.

Diffeomorphism invariance of general relativity leads to the Hamiltonian and momentum constraints. Conversely, the constraints generate the time and space diffeomorphisms. Under an infinitesimal translation $\epsilon^\mu(x, \tau)$ in spacetime – with ϵ^0 corresponding to the normal and ϵ^i corresponding to the tangent directions, as illustrated in Figure 2.6 – the conjugate pair (h_{ij}, π^{ij}) transforms as

$$\delta h_{ij} = \{h_{ij}, H[\epsilon^\mu]\}\,, \quad \delta\pi^{ij} = \{\pi^{ij}, H[\epsilon^\mu]\}\,, \tag{2.67}$$

with the Poisson bracket $\{\cdot,\cdot\}$ and the Hamiltonian

$$H[\epsilon^\mu] = \int \mathrm{d}^3x \left[\epsilon^0\mathcal{H}_0 + \epsilon^i\mathcal{H}_i\right]\,. \tag{2.68}$$

Since the constraints generate diffeomorphisms, the Einstein-Hilbert action is invariant under the transformation provided that the transformation leaves the the initial and final three-geometry unchanged, *i.e.*, $\epsilon^0(x, \tau_0) = \epsilon^0(x, \tau_1) = 0$.

The application of the Poisson bracket enables us to move a three-geometry in the spacetime manifold. However, two translations do not generally commute, *i.e.*, the three-geometry obtained by successively translating by ϵ_1 and ϵ_2 may differ from the one obtained

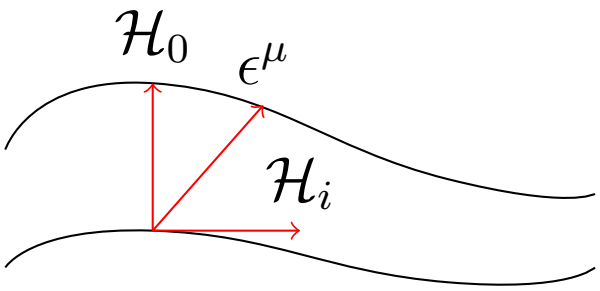


Figure 2.6: The Hamiltonian and momentum constraints respectively generate temporal and spatial translations parametrized by ϵ^μ.

by translating by ϵ_2 and ϵ_1. This can geometrically be seen by the change in the normal and tangent directions after a translation by ϵ_1 or ϵ_2. The deficit between the two translations is described by the constraint algebra,

$$\{\mathcal{H}_0(x), \mathcal{H}_0(x')\} = \left[(hh^{ij}(x)\mathcal{H}_i(x) + (hh^{ij}(x')\mathcal{H}_i(x')\right] \delta_{,j}(x, x') \,, \tag{2.69}$$

$$\{\mathcal{H}_0(x), \mathcal{H}_i(x')\} = [\mathcal{H}_0(x) + \mathcal{H}_0(x')] \, \delta_{,i}(x, x') \,, \tag{2.70}$$

$$\{\mathcal{H}_i(x), \mathcal{H}_j(x')\} = \mathcal{H}_i(x')\delta_{,j}(x, x') + \mathcal{H}_j(x)\delta_{,i}(x, x') \,, \tag{2.71}$$

first derived algebraically by Paul Dirac [114], and later Bryce DeWitt [96]. The constraint algebra is completely determined by and completely determines the embedding of space in spacetime (for the geometric derivation of the constraint algebra and the inverse construction of spacetime see [229] and [206]).

The discussion so far focused on empty spacetime. When including matter (or radiation) – which preserves the diffeomorphism invariance of the total theory – an analogous set of Hamiltonian and momentum constraints can be derived. The Hamiltonian constraint will generally consist of the Hamiltonian constraint of pure general relativity plus the Hamiltonian of the added matter (or radiation). The momentum constraint will generally include a term linear in the momentum of the added degrees of freedom.

After Dirac obtained the Hamiltonian of general relativity [113], he states that the superspace formulation might be more fundamental to the theory:

> One starts with ten degrees of freedom for each point in space, corresponding to the ten $g_{\mu\nu}$, but one finds with the method here followed that some dropout, leaving only six, corresponding to the six g_{rs}. This is a substantial simplification, but it can be obtained only at the expense of giving up four-dimensional

symmetry. I am inclined to believe from this that four-dimensional symmetry is not a fundamental property of the physical world".

Later he adds: "It would be permissible to look upon the Hamiltonian form as the fundamental one, and there would then be no fundamental four-dimensional symmetry in the theory." Along these lines, Wheeler and DeWitt further developed quantum geometrodynamics.

2.4.4 The Wheeler-DeWitt equation

So far, we discussed the concept of a wave function on superspace. The Hamiltonian formulation of general relativity enables us to develop the dynamics. Since the momentum constraint $\mathcal{H}_i$ describes the trivial space translations, we will restrict attention to the Hamiltonian constraint $\mathcal{H}_0$. The Einstein field equations in Hamilton-Jacobi form, known as he Hamilton-Jacobi-Einstein equation (see [271]), is given by

$$\mathcal{H}_0 \left[h_{ij}, \frac{\delta S}{\delta h_{ij}} \right] = (2\kappa) G_{ijkl} \frac{\delta S}{\delta h_{ij}} \frac{\delta S}{\delta h_{kl}} + (2\kappa)^{-1} h (2\Lambda - {}^3R) = 0 \,, \tag{2.72}$$

with $S[{}^3\mathcal{G}]$ the Hamilton-Jacobi function or action of general relativity. Note that this is like the Hamilton-Jacobi equation (2.36) at zero energy.

Dirac's scheme for the quantization of constraint systems results in the Wheeler-DeWitt equation (DeWitt [96])

$$\hat{\mathcal{H}}_0 \psi({}^3\mathcal{G}) = 0 \tag{2.73}$$

where the operator $\hat{\mathcal{H}}_0$ is obtained from the constraint using the substitution

$$\pi^{ij} \mapsto \frac{\delta}{i \delta h_{ij}} \,. \tag{2.74}$$

This equation is formally similar to the Klein-Gordon equation for the relativistic particle. Concretely, the Wheeler-DeWitt equation takes the Klein-Gordon-like form

$$\left[(2\kappa) G_{ijkl} \frac{\delta}{\delta h_{ij}} \frac{\delta}{\delta h_{kl}} + (2\kappa)^{-1} h ({}^3R - 2\Lambda) \right] \psi[{}^3\mathcal{G}] = 0 \,. \tag{2.75}$$

However, this is equation is not invariant under coordinate transformations. Since the canonical quantization procedure favors Cartesian coordinates, we interpret the kinetic term as the Laplace-Beltrami operator with respect to the metric on superspace,

$$\Delta^{(6)}\psi[{}^3\mathcal{G}] = \frac{1}{\sqrt{|G|}}\frac{\delta}{\delta h_{ij}}\left[\sqrt{|G|}G^{ijkl}\frac{\delta}{\delta h_{kl}}\psi[{}^3\mathcal{G}]\right], \tag{2.76}$$

where G is the determinant of the metric on superspace G_{AB} with A and B running over the independent components $\{11, 12, 13, 22, 23, 33\}$. In our conventions of the ADM-decomposition, the determinant can be expressed in terms of the determinant of the metric on space $G = -\frac{1}{16}h^4$. For flat Friedmann universes, with $h_{ij} = a^2\delta_{ij}$, the determinant of the metric on space and superspace are $h = a^6$ and $G = -\frac{1}{16}a^{24}$.

The Wheeler-DeWitt equation with the Laplace-Beltrami operator is covariant under coordinate transformations. However, it still suffers from an ordering problem. We thus write the Wheeler-DeWitt equation as

$$\left[(2\kappa)\Delta^{(6)} + (2\kappa)^{-1}h({}^3R - 2\Lambda) + \xi\,{}^6R\right]\psi[{}^3\mathcal{G}] = 0\,, \tag{2.77}$$

with ${}^6R$ the Rici curvature scalar corresponding to the metric on superspace G_{AB} and the ordering parameter ξ. The ordering problem also occurs in the path integral formulation [230].

In the semi-classical limit, the solution of the Wheeler-DeWitt equation can be approximated using the Hamilton-Jacobi function

$$\psi({}^3\mathcal{G}) = (\text{slowly varying amplitude}) \times e^{\frac{i}{\hbar}S[{}^3\mathcal{G}]}\,. \tag{2.78}$$

Analogous to the solutions of the time-independent Schrödinger equation, this wave function does not allow us to recover the dynamics of space in the form of classical spacetime. The wave function is spread over superspace! We can only hope to recover classical spacetime by superposing solutions of the Wheeler-DeWitt equation to constructing a localized wave packet in superspace. This wave function would imply an approximate notion of classical spacetime, including the quantum fluctuations of the geometry.

Even though this is a well-defined construction, it is difficult to implement and interpret. It is for example not clear what wave packet to select and how to translate the wave packet into physical predictions. Moreover, unlike the mathematically similar Klein-Gordon equation – which is a quantum constraint equation of similar form – the Wheeler-DeWitt equation does not include a time derivative by which it is difficult to formulate an unambiguous interpretation. These problems have been extensively studied. See the reviews [252, 25, 316, 211, 68, 231, 67]

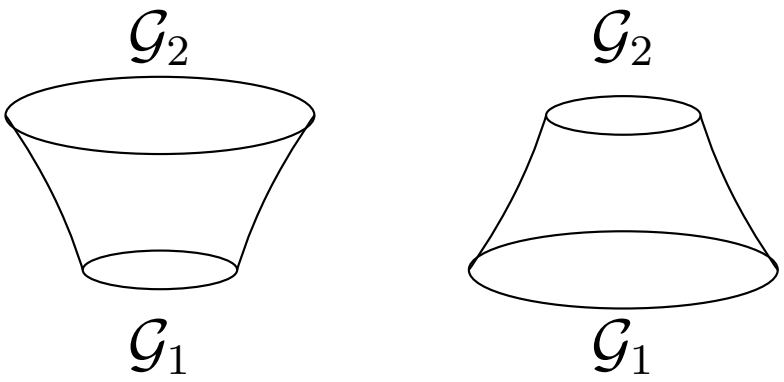


Figure 2.7: A pictorial description of the Feynman propagator, with $\mathcal{G}_1$ and $\mathcal{G}_2$ the initial and final three-geometry. *Left:* an expanding phase. *Right:* a contracting phase.

2.4.5 The path integral quantization

The Feynman path integral formulation of geometrodynamics is a natural extension of Hamilton-Jacobi construction discussed above. The amplitude to propagate between two three-geometries ${}^3\mathcal{G}_0, {}^3\mathcal{G}_1$ with the corresponding matter field configurations X_0, X_1 is colloquially given by the Feynman path integral over all interpolating spacetime manifolds and matter configurations (see Figure 2.7),

$$G[{}^3\mathcal{G}_1, X_1; {}^3\mathcal{G}_0, X_0] = \int_{{}^3\mathcal{G}_0, X_0}^{{}^3\mathcal{G}_1, X_1} \mathcal{D}g_{\mu\nu}\mathcal{D}X e^{\frac{i}{\hbar}S[g_{\mu\nu}, X]}, \tag{2.79}$$

with the action S consisting of the Einstein-Hilbert action plus the matter action,

$$S = \int (\mathcal{L}_{EH}(g_{\mu\nu}) + \mathcal{L}(g_{\mu\nu}, X))\sqrt{-g}\,\mathrm{d}^4x\,. \tag{2.80}$$

The advantage of the path integral over the canonical formalism is that the theory is covariant, *i.e.*, it does not rely on a spacetime foliation. Note that the factor ordering problem is still present [230].

The naive path integral prescription (2.79) does, however, not suffice, as the path integral over four metrics $g_{\mu\nu}$ over-counts spacetimes geometries. A similar problem occurs in the quantization of the relativistic particle and gauge theories in particle physics. In gauge theories, the constraints are linear in momentum (similar to the momentum constraints $\mathcal{H}_i$). The path integral can be made rigorous by adding the Faddeev-Popov ghosts, resulting from the BRST symmetry [133, 34, 315]. In general relativity, on the other hand, the

Hamiltonian constraint is quadratic in momenta, and the constraint algebra does not form a Lie algebra (as we say in the previous section). The structure constants are functions on superspace. In order to gauge-fix the path integral of a diffeomorphism invariant theory, we use the BFV quantization developed by Batalin, Fradkin, Vilkovisky [30] with corresponding BFV ghosts. The BFV quantization is an extension of the well-known BRST quantization for general constraints. Claudio Bunster (then known as Claudio Teitelboim) was the first to apply the BFV quantization to general relativity [307, 309, 308]. In the subsequent years, Jonathan Halliwell [184] obtained similar results for minisuperspace models. We here describe the resulting path integral formalism. For a detailed derivation, we refer to appendices 2.A, 2.B, and 2.C.

Given the constraint algebra of spacetime, the BFV quantization leads to the Feynman propagator of quantum geometrodynamics,

$$G[h_{ij}(\tau_1), X(\tau_1); h_{ij}(\tau_0), X(\tau_0)] = \int_{N(x)=0}^{N(x)=\infty} G[h_{ij}(\tau_1), X(\tau_1); h_{ij}(\tau_0), X(\tau_0); N(\tau_1-\tau_0)] \times \prod_x \mathrm{d}[\ln N(x)(\tau_1-\tau_0)]\,, \tag{2.81}$$

where the propagator in intrinsic parameter time $N(\tau_1-\tau_0)$ is given by the path integral

$$G[h_{ij}(\tau_1), X(\tau_1); h_{ij}(\tau_0), X(\tau_0); N(\tau_1-\tau_0)] = \int e^{iS} \prod_{x,\tau} \frac{\mathrm{d}h_{ij}\mathrm{d}\pi^{ij}\mathrm{d}C\mathrm{d}\bar{P}\mathrm{d}\bar{C}\mathrm{d}P}{2\pi} \mathcal{D}X\mathcal{D}\Pi\,, \tag{2.82}$$

with the phase-space action

$$S = \int_{\tau_1}^{\tau_2} \mathrm{d}\tau \int \mathrm{d}^3x \left[\dot{h}_{ij}\pi^{ij} + \dot{X}\Pi + i\dot{C}\bar{P} + i\dot{\bar{C}}P - N\left(\mathcal{H} + \mathcal{H}^{\mathrm{ghost}}\right)\right]. \tag{2.83}$$

The physical Hamiltonian $\mathcal{H}[h_{ij}, \pi^{ij}; X, \Pi] = \mathcal{H}_0[h_{ij}, \pi^{ij}] + \mathcal{H}_m[h_{ij}, X, \Pi]$ consists of the Einstein-Hilbert Hamiltonian $\mathcal{H}_0$ and the matter Hamiltonian $\mathcal{H}_m$. The ghost Hamiltonian

$$\mathcal{H}^{\mathrm{ghost}}[h_{ij}; C, \bar{P}; \bar{C}, P] = i[\bar{P}P - hh^{ij}\bar{C}_{,i}C_{,j}]\,, \tag{2.84}$$

is defined in terms of the fermionic ghosts fields $C, \bar{C}$ and their canonical momenta $\bar{P}, P$ satisfying the boundary conditions

$$C[x,\tau_1] = C[x,\tau_2] = 0\,, \quad \bar{C}[x,\tau_1] = \bar{C}[x,\tau_2] = 0\,. \tag{2.85}$$

For convenience, we will, without loss of generality, assume $\tau_0 = 0$ and $\tau_1 = 1$ to simplify the formulas.

Quantum cosmology

In the context of quantum cosmology, we can simplify the formalism by restricting the integral to three-geometries satisfying certain symmetries. The subset if known as minisuperspace.

For simplicity consider the minisuperspace, consisting of homogeneous and isotropic three-geometries. Dropping the spatial derivatives, the propagator factorizes into an infinite product over space

$$G = \prod_x \int_0^\infty \mathrm{d}\left[\ln N\right] \int e^{i\int_0^1 \left[\dot{a}p + \dot{X}\Pi + i\dot{C}\bar{P} + i\dot{\bar{C}}P - N(\mathcal{H}+\mathcal{H}^{\mathrm{ghost}})\right]\mathrm{d}\tau} \mathcal{D}a\mathcal{D}p\mathcal{D}C\mathcal{D}\bar{P}\mathcal{D}\bar{C}\mathcal{D}P\,. \tag{2.86}$$

The ghost Hamiltonian decouples from the metric, *i.e.*,

$$\mathcal{H}^{\mathrm{ghost}} = i\bar{P}P\,, \tag{2.87}$$

and the integral over the ghost can be evaluated exactly

$$\int e^{i\int_{-1/2}^{1/2}\left[i\dot{C}\bar{P} + i\dot{\bar{C}}P - iN\bar{P}P\right]\mathrm{d}\tau} \mathcal{D}C\mathcal{D}\bar{P}\mathcal{D}\bar{C}\mathcal{D}\bar{P} = \int e^{-\int_{-1/2}^{1/2} N\dot{C}\dot{\bar{C}}\mathrm{d}\tau} \mathcal{D}C\mathcal{D}\bar{C} = N\,. \tag{2.88}$$

The Feynman propagator for minisuperspace reduces to the path integral

$$G[a_2; a_1] = \int_0^\infty \mathrm{d}N \int_{a_1}^{a_2} \mathcal{D}a e^{\frac{i}{\hbar}\int_{-1/2}^{1/2} L[a,\dot{a};N]\mathrm{d}t}\,, \tag{2.89}$$

in accordance with Halliwell [184]. In terms of the canonical quantization, the propagator is a Green's function of the Wheeler-DeWitt equation

$$\hat{\mathcal{H}}_0 G[a_2; a_1] = -i\delta(a_1, a_2)\,, \tag{2.90}$$

with the Hamiltonian $\hat{\mathcal{H}}_0$ operating on either a_1 or a_2.

For anisotropic homogeneous Bianchi cosmologies, a similar path integral formulation can be obtained. After writing the Einstein action in terms of coordinates respecting the Killing vector fields, the action reduces to a functional dependent the lapse N_0 and a set of time-dependent variables (often called scale-factors). This action leads to a single quadratic constraint in the momenta. The constraint algebra is trivial, and the BFV quantization analogous to the quantization of the relativistic particle (see appendix 2.B). As a consequence equation, (2.89) applies path integrals over homogeneous minisuperspace.

2.5 The beginning of the universe

Cosmological surveys have painted a relatively simple picture of the early universe. The universe is spatially flat, homogeneous, and isotropic on large scales. The fluctuations on smaller scales, which source structure formation, are close to Gaussian with a conformal power spectrum. The cosmological concordance model manages to explain a range of observations in terms of only a few cosmological parameters. It is thus natural to study, how our universe came to be in such a simple initial state.

In the late '70s and early '80s, several mechanisms were proposed including the Bunch-Davies vacuum, the no-boundary proposal, and the tunneling proposal. These models describe the formation of quantum fluctuations where are amplified and rendered classical in a subsequent phase of cosmological inflation. In this inflationary phase, the universe expands exponentially smoothing the inhomogeneities. Even though it is normally reasoned that these models in combination with an inflationary phase are able to fit the data, it remains an active field of research, as the current models often assume more than they predict and are difficult to observationally falsify.

We will here shortly summarize these three proposals. Remark that whereas the Bunch-Davies vacuum is an application of quantum field theory on a classical spacetime, the no-boundary and tunneling proposals are based on the quantization of both the quantum field and the background spacetime.

2.5.1 The Bunch-Davies vacuum

In Minkowski spacetime, there exists a unique vacuum for a scalar field described by the Klein-Gordon equation. The state is the minimal excitation and is invariant under the isometries of the spacetime. The concept of a vacuum state does not naturally generalize to quantum field theory in curved spacetime. Unless the background metric tensor has a global timelike Killing vector, there is no way to define a vacuum or ground state, *i.e.*, the vacuum of one observer does generically not look like a vacuum state to another observer in another frame.

In maximally symmetric de Sitter spacetime, there exists a class of states which are invariant under the isometries. These are known as the α-vacua. In conventional inflation theory, it is usually reasoned that there exists a preferred α-vacuum known as the Bunch-Davies state [66]. The Gaussian fluctuations observed in the cosmic microwave background radiation field can be traced back to an early time at which they were sub-horizon modes. For these sub-horizon modes, the effect of gravity is negligible. It, for this reason, can

be argued that the vacuum state should at this time coincide with that of the Minkowski vacuum. As time evolves, these Minkowski modes are stretched outside of the horizon after which they freeze. The modes only start to evolve again, as classical density fluctuation, after inflation, when they re-enter the horizon to form the CMB. The exact expansion rate of the inflating spacetime, at the time when the Minkowski modes exist the horizon translate into the nearly conformal power-spectrum of the observed matter fluctuations. For an overview of quantum field theory on curved spacetime see [46, 154, 336, 257]

Note that since the analysis is based on the quantization of a field on a classical spacetime, the Bunch-Davies vacuum is not a model for the beginning of the universe in the sense of the primeval atom George Lemaître envisaged. We are currently investigating the Bunch-Davies vacuum on a quantum spacetime. Current results indicate that the Bunch-Davies state is not recovered on a quantum background [101].

2.5.2 The no-boundary proposal

The no-boundary proposal comes closer to the idea of the primeval atom. Stephen Hawking, at the 1981 conference at the Vatican, described his vision for a theory for the beginning of the universe. This presentation was in the subsequent year captured in the paper *The boundary conditions of the universe* [194]. As he states:

> This paper considers the question of what are the boundary conditions of the universe and where should they be imposed. It is difficult to define boundary conditions at the initial singularity and, even if one could, they would be insufficient to determine the evolution of the universe. In order to overcome this problem, it is suggested that one should adopt the Euclidean approach and evaluate the path integral for quantum gravity over positive definite metrics.

In this, he builds on the earlier work by DeWitt and Wheeler. In Hawking's words:

> If one took these metrics to be compact, one would avoid the need to specify any boundary conditions for the universe. There ought to be something very special about the boundary conditions of the universe and what can be more special than the condition that there is no-boundary.

This is an extremely elegant and beautiful idea, as it would remove the initial singularity of classical general relativity and at the same time describe the initial fluctuations without relying on additional particles and parameters.

He is moreover very clear about the rules one has to obey in such an endeavor.

> The aim of physics is to provide a mathematical model of the universe which will agree with all observations that have been made so far and which will predict the results of further observations. Our present models would consist of two parts
>
> 1. a set of differential equations that govern the variables in the theory. These are normally derived from an action principle;
> 2. boundary conditions for the differential equations, or for the fields that are considered in the action principle.
>
> One could conceive of models which did not have this division into field equations and boundary conditions but the separation of the two has been very valuable in enabling us to make progress by considering only a local region of the universe instead of having to try to devise a model which would account for the whole universe at one go. We have made a lot of progress on the first part of the problem in recent years and it now seems possible that we might find a fully unified field theory within the not-too-distant future. However, we shall not have a complete model of the universe until we can say more about the boundary conditions than that they must be whatever would produce what we observe.

I would like to emphasize the last point: we need to find a principled explanation of the beginning of the universe, since it is always possible to construct a fast range of ac hoc models which fit the data.

This integral over compact geometries can naturally be implemented in the Euclidean version of de Sitter space, as the analytic continuation consists of the compact four-sphere. In the question session of Hawking's talk the theoretical physicist, and his Ph.D. supervisor, Dennis Sciama probed Stephen Hawking on the Euclidean assumption of his proposal: "I understand that not all curved spaces of indefinite metric can be analytically continued into space with a positive definite metric. Are you depending essentially on the fact that it can be done for a Robertson-Walker metric?" after which Stephen Hawking details his motivation

> It is true that only rather special metrics, like the Schwarzschild and Robertson-Walker metrics, admit a Euclidean section on which the metric is real and positive definite. However, the idea is that a path integral over all positive definite metrics is equivalent to a path integral over all Lorentzian metrics. It is similar to what is done in ordinary field theory: one integrates over all fields

> that are real in Euclidean space and says that it is equivalent to integrating over fields that are real in Minkowski space.

It is thus safe to say that the no-boundary proposal was first envisaged as a path integral over all compact Lorentzian metrics. By considering spacetimes which do not have a boundary or singularity, the no-boundary proposal avoids the need for initial boundary conditions.

In 1983, James Hartle and Stephen Hawking formalized the proposal in their famous paper *Wave function of the Universe* [188]. The wavefunction of the universe is written as a Lorentzian path integral. It is subsequently stated that " The oscillatory integral [the Lorentzian propagator] is not well defined but can be made so by rotating the time to imaginary values. " The resulting Euclidean integral is no longer oscillatory but is however not sensible since the action is not positive definite. This is known as the *conformal factor problem*. As a consequence, the Wick-rotated integral does not converge.

> The [Euclidean] functional integrals [...] therefore require careful definition. One way of doing this is to break the integration up into an integral over conformal factors and over geometries in a given conformal equivalence class. By appropriate choice of the contour of integration of the conformal factor the integral can probably be made convergent. If this is the case a properly convergent functional integral can be constructed.

This is subsequently studied in a mini-superspace model consisting of a Λ-dominated universe. The remarkable thing is that we, in chapter 3, find that the original Lorentzian integral is, in fact, well-defined and that the Lefschetz thimble is the corresponding deformation of the integration domain in the complex plane. Following this initial paper, Halliwell and Hawking went beyond mini-superspace and analyzed the formation of the fluctuations observed in the cosmic microwave background radiation field [182, 178, 179]. The no-boundary proposal predicts the universe to start its evolution at to bottom of the inflaton potential with Gaussian fluctuations. The predictions can be made to fit cosmological observations using the anthropic principle.

More recently, Hartle, Halliwell, and Hertog have developed the proposal further by rejecting the path integral formulation and basing the theory on solutions of the Wheeler-DeWitt equations [181]. It is proposed that one can go beyond the semi-classical approximation using holography. It should, however, be remarked that this new proposal significantly differs from Hawking's original vision as it is a based on a different mathematical technique. The new definition, moreover, requires the boundary conditions to be selected to match observations, contrary to Hawking's original philosophy.

2.5.3 The tunneling proposal

Around the same time as the conception of the no-boundary proposal, Alexander Vilenkin developed the tunneling proposal, in which the early universe tunneled to a classical state from 'nothing' [325, 326, 323, 327, 328]. This is in spirit similar to the primeval atom. In his words:

> Modern cosmology gives an evolutionary picture of the Universe. It aims to describe how the Universe has evolved to its present state form a certain initial state. The differential equations describing the evolution are derived from known laws of physics; however, there are no laws determining the initial conditions. This seems to indicate that our understanding of the Universe is bound to be incomplete: we will have to say, in effect, that the Universe is what it is because it was what it was.
>
> I have recently suggested a cosmological model in which the Universe is created by quantum tunneling from 'nothing' to de Sitter space, whereby 'nothing' I mean a state with no classical spacetime. In this model, the initial state of the Universe is determined by the laws of physics, and no initial or boundary conditions are required.

The motivation of this work is thus rather similar to Hawking's motivation for the no-boundary proposal. Here the *tunnelling from nothing* should be compared to the pair creation of an electron, positron pair in an electric field in the Schwinger effect. The wave-function for the creation of the universe from nothing can be written as the Lorentzian path integral over compact Lorentzian four-geometries. This is thus analogous to the formulation of the no-boundary proposal. The proposal, however, differs in that Vilenkin proposes to include four-geometries with integrable singularities (and finite actions). The path integral is, by analogy with the Schwinger effect, dominated by the Euclidean instanton of de Sitter spacetime.

The prediction of the tunneling proposal, however, differs from the no-boundary proposal, as the tunneling proposal predicts the universe to start at the top of the inflaton potential, enabling the universe to expand exponentially. This proposal thus does not require the anthropic principle to match observations. The cosmological predictions of the proposal do rely on the details of the subsequent inflationary phase.

More recently Vilenkin et al. have developed the proposal further [329, 330, 324].

2.6 Summary

Wheeler developed the framework for quantum geometrodynamics, along the lines of quantum mechanics, in an attempt to marry quantum mechanics with general relativity. Quantum spacetime is described in terms of a wave function on superspace. Classical spacetime emerges from constructive interference. The framework is surprisingly close to Lemaître's early vision on the primeval atom, *i.e.*, 'the now, which has no yesterday because, yesterday, there was no space'.

In the subsequent years, several scientists attempted to formalize the theory, leading to the Wheeler-DeWitt equation and the path integral for gravity. Notwithstanding these achievements, quantum geometrodynamics has not yet achieved the status of a falsifiable theory in physics. This can mainly be attributed to the fact that we do not have an unambiguous interpretation for the solutions of the Wheeler-DeWitt equation and the amplitude to propagate between two three-geometries. The theory lacks a natural time parameter, making it difficult to describe the emergence of spacetime and the time evolution of physical phenomena. It was for this reason that John Wheeler and Bryce DeWitt eventually gave up on this line of investigation.

The central ideas of quantum geometrodynamics have however persisted. Many theories which attempt to describe the quantization of gravity, ranging from string theory to loop quantum gravity and causal set theory, relate in some way to either the Wheeler-DeWitt equation or the path integral of gravity. This follows from the fact that, in essence, the Wheeler-DeWitt equation is nothing more than the quantum mechanical realization of the diffeomorphism invariance of general relativity. It is for this reason, that both the Wheeler-DeWitt equation and the path integral formulation, remain active fields of investigation.

Moreover, Stephen Hawking and Alexander Vilenkin constructed two models for the beginning of spacetime, which are sill being investigated. These models rely on either the Wheeler-DeWitt equation of geometrodynamics or the path integral formulation developed by Claudio Bunster. The two proposals have opened the investigations of the beginning of the universe using quantum gravity. The two models, however, rely on Euclidean gravity, which is not without its complications and the two models give contradictory results. In the subsequent chapters, we study the two models using the Lorentzian path integral. We use Picard-Lefschetz theory to remove the oscillations by deforming the integration contour in the complex plane.

2.A The BFV quantization

The Batalin Fradkin Vilkovisky (BFV) quantization of constraint systems [30] forms an extension of the quantization by Faddeev and Popov [133, 132]. We here briefly illustrate the quantization procedure following the discussion by Halliwell and Hartle [180] and Teitelboim [307].

We start with the action

$$S_0 = \int_{\tau_1}^{\tau_2} \left[p_i \dot{q}^i - H_0 - \lambda^\alpha T_\alpha \right] \mathrm{d}\tau \,, \tag{2.91}$$

with the canonical Hamiltonian H_0, a set of first class constraints T_α and a corresponding set of Lagrange multipliers λ^α. The constraints satisfy the constraint algebra

$$\{T_\alpha, T_\beta\} = U^\gamma_{\alpha\beta} T_\gamma \,, \quad \{H_0, T_\alpha\} = V^\beta_\alpha T_\beta \,. \tag{2.92}$$

In Yang-Mills theory, the constraints generate a Lie algebra with structure constants $U^\gamma_{\alpha\beta}, V^\beta_\alpha$. In general $U^\gamma_{\alpha\beta}, V^\beta_\alpha$ can depend on p_i and q^i and are known as structure functions. The constraints generate a symmetry transformation

$$\delta q^i = \epsilon^\alpha \{q^i, T_\alpha\} \,, \quad \delta p_i = \epsilon^\alpha \{p_i, T_\alpha\} \,, \tag{2.93}$$

$$\delta \lambda^\alpha = \dot{\epsilon}^\alpha - U^\alpha_{\beta\gamma} \lambda^\beta \epsilon^\gamma - V^\alpha_\beta \epsilon^\beta \,, \tag{2.94}$$

for infinitesimal ϵ. Under this transformation the action transforms as

$$\delta S_0 = \left[\epsilon^\alpha \left[p_i \frac{\partial T_\alpha}{\partial p_i} - T_\alpha \right] \right]_{\tau_1}^{\tau_2} \,. \tag{2.95}$$

The transformation is a symmetry under the boundary condition

$$\epsilon^\alpha(\tau_1) = \epsilon^\alpha(\tau_2) = 0 \,, \tag{2.96}$$

for all constraints not strictly linear in momentum. The symmetry can be broken by adding the gauge fixing term

$$S_{GF} = \int_{\tau_1}^{\tau_2} \Pi_\alpha \left[\dot{\lambda}^\alpha - \chi^\alpha \right] \,. \tag{2.97}$$

imposing the condition $\dot{\lambda}^\alpha = \chi^\alpha(p_i, q^i, \lambda^\alpha)$. The action $S_0 + S_{GF}$ is however not invariant under the gauge fixing condition. We for this reason add the ghost action

$$S_{ghost} = \int_{\tau_1}^{\tau_2} \Big[\bar{\rho}_\alpha \dot{c}^\alpha + \rho^\alpha \dot{\bar{c}}_\alpha - \bar{\rho}_\alpha \rho^\alpha - \bar{c}_\alpha \{\chi^\alpha, T_\beta\} c^\beta - \bar{c}_\alpha \frac{\partial \chi^\alpha}{\partial \lambda^\beta} \rho^\beta \tag{2.98}$$

$$- \bar{\rho}_\alpha V^\alpha_\beta c^\beta - \bar{\rho}_\alpha U^\alpha_{\beta\gamma} \lambda^\beta c^\gamma - \frac{1}{2} \bar{c}_\alpha c^\gamma \{\chi^\alpha, U^\beta_{\gamma\sigma}\} \bar{\rho}_\beta c^\sigma \Big] \mathrm{d}\tau \,. \tag{2.99}$$

with the fermionic ghosts fields $c^\alpha, \bar{\rho}_\alpha, \bar{c}_\alpha, \rho^\alpha$ and the Lagrange multipliers $\lambda^\alpha, \Pi_\alpha$ satisfying the Poisson-brackets

$$\{\rho_\alpha, c^\beta\} = \{\rho^\alpha, \bar{c}_\beta\} = \{\lambda^\alpha, \Pi_\beta\} = \delta^\alpha_\beta \,. \tag{2.100}$$

The Poisson bracket for Grassmann variables differs by a sign. The total action

$$S_T = S_0 + S_{GF} + S_{ghost} \tag{2.101}$$

fixes the gauge $\dot{\lambda} = \chi$ and is independent of the gauge choice. In Yang-Mills theories, the procedure reduces to the Faddeev-Popov ghosts.

We here considered a constraint system with rank 1. This treatment suffices for general relativity. For theories with higher rank see Henneaux [198]. The BFV ghost action is often written more compactly. The pair (P_A, Q^A), defined by

$$Q^A = (q^i, \lambda^\alpha, c^\alpha, \bar{C}_\alpha) \,, \quad P_A = (p_i, \Pi_\alpha, \bar{\rho}_\alpha, \rho^\alpha), \tag{2.102}$$

is canonical since

$$\{Q^A, P_B\} = \delta^A_B \,. \tag{2.103}$$

The symmetry transformation of a functional F on the extended phase-space is given by

$$\delta F = \{F, \Lambda\Omega\} \tag{2.104}$$

with the BRST charge

$$\Omega = c^\alpha T_\alpha + \rho^\alpha \Omega_\alpha - \frac{1}{2} U^\alpha_{\beta\gamma} c^\beta c^\gamma \bar{\rho}_\alpha \,, \tag{2.105}$$

and the infinitesimal Grassmann parameter Λ. The BRST symmetry transformation is exact, *i.e.*,

$$\delta^2 F = \delta(\delta F) = \{\delta F, \Lambda_1 \Omega\} = \{\{F, \Lambda_2 \Omega\}, \Lambda_1 \Omega\} = 0 \,, \tag{2.106}$$

as can be shown using the Jacobi identity

$$0 = 2\Lambda_1\Lambda_2\{\{F,\Omega\},\Omega\} + \Lambda_2\Lambda_1\{\{\Omega,\Omega\},F\}\,, \tag{2.107}$$

where by construction $\{\Omega,\Omega\} = 0$.

The total action reads

$$S_T = S_0 + S_{GF} + S_{ghost} = \int_{\tau_1}^{\tau_2} \left[P_A\dot{Q}^A - H_0 - \bar{\rho}_\alpha V_\beta^\alpha c^\beta - \{\bar{\rho}_\alpha\lambda^\alpha + \bar{c}_\alpha\chi^\alpha,\Omega\}\right]\mathrm{d}\tau\,. \tag{2.108}$$

The BRST transformation is a symmetry when the additional phase-space variables satisfy the boundary conditions

$$q^i(\tau_1) = q_1^i\,, \quad q^i(\tau_2) = q_2^i \tag{2.109}$$

$$\Pi_\alpha(\tau_1) = \Pi_\alpha(\tau_2) = c^\alpha(\tau_1) = c^\alpha(\tau_2) = \bar{c}_\alpha(\tau_1) = \bar{c}_\alpha(\tau_2) = 0\,. \tag{2.110}$$

The BFV gauge fixed path integral is given by

$$G[q_2,q_1] = \int \mathcal{D}P_A\mathcal{D}Q^A e^{iS_T}\,. \tag{2.111}$$

The path integral is independent of the gauge fixing, since a coordinate transformation, which keeps the integral fixed, leads to the addition of a Poisson bracket to the action. Constructing a coordinate transformation that leads to the addition of $\{\bar{c}_\alpha(\chi'-\chi),\Omega\}$, changes the gauge choice from $\dot{\Lambda}^\alpha = \chi$ to $\dot{\Lambda}^\alpha = \chi'$ without changing the path integral. This is known as the Fradkin-Vilkovisky theorem.

2.B The BFV quantization of the relativistic particle

The action of a free relativistic particle of mass m is given by the proper length of its trajectory,

$$S[x(\tau)] = -m\int_{\tau_0}^{\tau_1}\sqrt{-\dot{x}^2}\mathrm{d}\tau \tag{2.112}$$

$$= \int_{\tau_0}^{\tau_1}\left[\frac{\dot{x}^2}{2N} - \frac{Nm^2}{2}\right]\mathrm{d}\tau\,, \tag{2.113}$$

with $x^\mu(\tau)$ the trajectory of the particle as a function of parameter time $\tau \in [\tau_0, \tau_1]$, and $\dot{x}^\mu = \frac{\mathrm{d}x^\mu}{\mathrm{d}\tau}$. The second equality is the Polyakov formulation with the Lagrange multiplier N. The equation of motion for N is

$$N = \frac{\sqrt{-\dot{x}^2}}{m}\,. \tag{2.114}$$

The two actions are classically equivalent since upon substitution in the Polyakov action we obtain the traditional action. We use the Polyakov action, since this formulation is closest to the Einstein-Hilbert action.

To illustrate the BFV quantization, we write the action in the phase-space formulation. The momentum canonical to x^μ is given by

$$p_\mu = \frac{\partial L}{\partial \dot{x}^\mu} = \frac{\dot{x}_\mu}{N}\,. \tag{2.115}$$

The corresponding Hamiltonian is

$$H = \dot{x}^\mu p_\mu - L = \frac{1}{2}N\left[p_\mu p^\mu + m^2\right] = \frac{1}{2}N\mathcal{H}\,, \tag{2.116}$$

with the physical Hamiltonian $\mathcal{H} = p_\mu p^\mu + m^2$. The action is given by

$$S[x(\tau)] = \int_{\tau_0}^{\tau_1} \left[\dot{x}^\mu p_\mu - N\mathcal{H}\right] \mathrm{d}\tau\,, \tag{2.117}$$

where we redefined the lapse $N \to 2N$. The equation of motion with respect to the Lagrange multiplier N gives the constraint

$$\mathcal{H} = 0\,, \tag{2.118}$$

which is the well known relation $E^2 - \mathbf{p}^2 = m^2$. Quantization of the constraint, according to the Dirac scheme, gives the free Klein-Gordon equation

$$\hat{H}\psi = \left[\partial^\mu \partial_\mu - m^2\right]\psi = 0\,. \tag{2.119}$$

The constraint is a consequence of reparametrization invariance. Consider trajectories $x(\tau)$ with the boundary conditions $x(\tau_0) = x_0$ and $x(\tau_1) = x_1$. The action of a relativistic particle is the proper length of the trajectory. Consequently, the action is invariant under the reparametrizations

$$\tau \to f(\tau) \tag{2.120}$$

for $f : [\tau_0, \tau_1] \to [\tau_0, \tau_1]$, as long as f holds the endpoints fixed

$$f(\tau_0) = \tau_0 \,, \quad f(\tau_1) = \tau_1 \,. \tag{2.121}$$

For infinitesimal transformations

$$f(\tau) = \tau + \epsilon(\tau) \,, \tag{2.122}$$

equation (2.121) gives the boundary condition

$$\epsilon(\tau_0) = \epsilon(\tau_1) = 0 \,. \tag{2.123}$$

The constraint $H = 0$ is a consequence of this symmetry, since the transformation corresponds to a redefinition of the lapse N. Conversely, the constraint generates the reparametrization symmetry, since under the transformation

$$\delta x^\mu = \epsilon\{x^\mu, \mathcal{H}\} \,, \quad \delta p_\mu = \epsilon\{p_\mu, \mathcal{H}\} \,, \quad \delta N = \dot{\epsilon} \,, \tag{2.124}$$

the action is invariant, *i.e.*,

$$\delta S = \left[\epsilon \left(p_\mu \frac{\partial \mathcal{H}}{\partial p_\mu} - \mathcal{H} \right) \right]_{\tau_0}^{\tau_1} \tag{2.125}$$

$$= \left[\epsilon \left(p_\mu p^\mu - m^2 \right) \right]_{\tau_0}^{\tau_1} \tag{2.126}$$

$$= 0 \,, \tag{2.127}$$

due to the boundary conditions (2.123).

The Hamiltonian formulation of the free relativistic particle consists of only one constraint. The constraint satisfies a trivial constraint algebra,

$$\{\mathcal{H}, \mathcal{H}\} = 0 \,. \tag{2.128}$$

The constraint algebra allows us to fix a gauge using the BFV procedure. We here consider the proper-time gauge

$$\dot{N} = 0 \,, \tag{2.129}$$

which is a valid gauge choice for the free relativistic particle, since every parametrization $N(\tau)$ can be brought to this gauge by the reparametrization (2.124). That is to say, applying the gauge condition to the transformation gives

$$\ddot{\epsilon}(\tau) = -\dot{N}(\tau) \,, \tag{2.130}$$

which has a unique solution satisfying the boundary condition (2.123). Applying the gauge fixing and ghost action we obtain the total action

$$S_T = \int_{\tau_0}^{\tau_1} \left[\dot{x}^\mu p_\mu - N\mathcal{H} + \dot{N}\Pi + \dot{c}\bar{\rho} + \dot{\bar{c}}\rho - \rho\bar{\rho} \right] \mathrm{d}\tau \,. \tag{2.131}$$

The Polyakov action, the gauge fixing term and the ghost term decouple. The BFV path integral can be evaluated exactly,

$$\begin{aligned} G[x_1; x_0] &= \int \mathcal{D}x^\mu \mathcal{D}p_\mu \mathcal{D}N \mathcal{D}\Pi \mathcal{D}c \mathcal{D}\bar{\rho} \mathcal{D}\bar{c} \mathcal{D}\rho \; e^{iS_T} && (2.132) \\ &= \int \mathrm{d}N \int \mathcal{D}x^\mu \mathcal{D}p_\mu \mathcal{D}c \mathcal{D}\bar{\rho} \mathcal{D}\bar{c} \mathcal{D}\rho \; e^{i\int_{\tau_0}^{\tau_1} \left[\dot{x}^\mu p_\mu - N\mathcal{H} + \dot{c}\bar{\rho} + \dot{\bar{c}}\rho - \rho\bar{\rho} \right] \mathrm{d}\tau} && (2.133) \\ &= \int \mathrm{d}N (\tau_1 - \tau_0) \int_{x_0}^{x_1} \mathcal{D}x^\mu \mathcal{D}p_\mu \; e^{i\int_{\tau_0}^{\tau_1} [\dot{x}^\mu p_\mu - N\mathcal{H}] \mathrm{d}\tau} \,, && (2.134) \end{aligned}$$

where we use the identities

$$\int \mathcal{D}N \mathcal{D}\Pi e^{i\int_{\tau_0}^{\tau_1} \dot{N}\Pi \mathrm{d}\tau} = \int \mathcal{D}N \delta[N] = \int \mathrm{d}N \,, \tag{2.135}$$

$$\int \mathcal{D}c \mathcal{D}\bar{\rho} \mathcal{D}\bar{c} \mathcal{D}\rho e^{i\int_{\tau_0}^{\tau_1} \left[\dot{c}\bar{\rho} + \dot{\bar{c}}\rho - \rho\bar{\rho} \right] \mathrm{d}\tau} = (\tau_1 - \tau_0) e^{\frac{(\bar{c}_1 - \bar{c}_0)(c_1 - c_0)}{\tau_2 - \tau_1}} = \tau_1 - \tau_0 \,. \tag{2.136}$$

Now assuming $\tau_0 = 0$ and $\tau_1 = 1$, the propagator can be written as

$$G[x_1; x_0] = \int_0^\infty \mathrm{d}N G[x_1; x_0; N] \,, \tag{2.137}$$

in terms of the propagator from x_0 to x_1 in Schwinger time N

$$G[x_1; x_0; N] = \int_{x_0}^{x_1} \mathcal{D}x^\mu \mathcal{D}p_\mu e^{i\int_0^1 [\dot{x}^\mu p_\mu - NH] \mathrm{d}\tau} \,. \tag{2.138}$$

This path integral is Gaussian and can be explicitly evaluated. First note that the Hamiltonian is independent of position. Integrating $\dot{x}^\mu p_\mu$ by parts gives

$$G[x_1; x_0; N] = \int_{x_0}^{x_1} \mathcal{D}x^\mu \mathcal{D}p_\mu e^{i\int_0^1 [-x^\mu \dot{p}_\mu - N\mathcal{H}] \mathrm{d}\tau + i x^\mu p_\mu |_0^1} \,. \tag{2.139}$$

The path integral over x gives a functional Dirac delta function $\delta[\dot{p}]$ which reduces the path integral over p_μ to an ordinary integral,

$$G[x_1; x_0; N] = \frac{1}{(2\pi)^3} \int \mathrm{d}p_\mu e^{i\left[-N(p^2 + m^2) + p_\mu (x_1^\mu - x_0^\mu) \right]} \,. \tag{2.140}$$

The Feynman propagator is the propagator of a free Klein-Gordon field, *i.e.*,

$$G[x_1; x_0] = \frac{1}{i(2\pi)^3} \int \mathrm{d}^4 p \frac{e^{ip\cdot(x_1 - x_0)}}{p^2 + m^2 - i\epsilon}. \tag{2.141}$$

The $i\epsilon$ regularization follows from Picard-Lefschetz theory. Note that there is no freedom in the integration contour. We are not allowed to freely move the poles of the integrand to obtain the other Klein-Gordon propagators.

2.C The BVF quantization for general relativity

Consider a spacetime, parametrized by parameter time τ, with the space-like boundaries ${}^3\mathcal{G}_0$ and ${}^3\mathcal{G}_1$ and corresponding matter fields X_0, X_1 at parameter times τ_0 and τ_1. The Einstein-Hilbert action is invariant under a translation ϵ^μ generated by the constraints equation (2.67), with a corresponding transformation of the lapse and shift functions

$$\delta N^{\rho''} = \dot{\epsilon}^{\rho''} - \kappa_{\mu\nu'}{}^{\rho''} N^\mu \epsilon^{\nu'}, \tag{2.142}$$

provided ϵ^μ satisfies the boundary condition

$$\epsilon^0(x, \tau_0) = \epsilon^0(x, \tau_1) = 0\,. \tag{2.143}$$

Given the constraint algebra and a gauge choice, the BFV scheme allows us to gauge fix the Feynman path integral. To remove any gauge freedom in the choice of coordinates on the boundary three-geometries ${}^3\mathcal{G}_0$, ${}^3\mathcal{G}_1$, we require

$$\epsilon^i(x, \tau_0) = \epsilon^i(x, \tau_1) = 0\,. \tag{2.144}$$

The proper-time gauge

$$\dot{N}^0(x, \tau) = 0\,, \quad N^i(x, \tau) = 0 \tag{2.145}$$

for all $x \in {}^3\mathcal{G}_0$ and $\tau_0 < \tau < \tau_1$, fixes the coordinates on the spacetime manifold. This gauge respects condition (2.143), but is too stringent to respect condition (2.144). That is to say, given an initial time-slice, we can choose a constant N^0 for every point in the three-geometry to reach the final time slice (with the correct proper time). The coordinates on the final slice are however completely fixed by the choice of coordinates on the initial slice, since $N^i = 0$.

In order to make the gauge more flexible, we divide the interval $[\tau_0, \tau_1]$ into two parts $[\tau_0, \tau']$ and $[\tau', \tau_1]$ for fixed $\tau_0 < \tau' < \tau_1$. In the first interval we move from the boundary ${}^3\mathcal{G}_0$ to ${}^3\mathcal{G}_1$ while keeping the spatial coordinates fixed. In the second interval we remain on the boundary ${}^3\mathcal{G}_1$ and adjust the spatial coordinates, *i.e.*,

$$\dot{N}^0(x,\tau) = 0, \quad N^i(x,\tau) = 0 \quad \text{for } \tau_0 \leq \tau \leq \tau' \,, \tag{2.146}$$

$$N^0(x,\tau) = 0, \quad \dot{N}^i(x,\tau) = 0 \quad \text{for } \tau' \leq \tau \leq \tau_1 \,, \tag{2.147}$$

for all $x \in {}^3\mathcal{G}_0$. This gauge is more flexible than (2.145) and respects both boundary conditions (2.143), (2.144). The gauge is of course far from unique and the intermediate parameter time τ' is arbitrary. The final propagator will not depend on τ'.

We now derive the BFV ghosts corresponding to this gauge choice. In general relativity we start with the Einstein-Hilbert action, in the phase-space form given by

$$S_0 = \int_{\tau_0}^{\tau_1} \mathrm{d}\tau \int \mathrm{d}^3x \left[\dot{h}_{ij}\pi^{ij} - N^\mu \mathcal{H}_\mu \right] . \tag{2.148}$$

The theory has the constrains $T_\mu = \mathcal{H}_\mu$, Lagrange multipliers $\lambda^\mu = N^\mu$ and no canonical Hamiltonian. The constraint algebra can be written as

$$\{\mathcal{H}_\mu(x), \mathcal{H}_\nu(x')\} = \int \mathrm{d}^3x'' \kappa_{\mu\nu}{}^\rho(x, x', x'') \mathcal{H}_\rho(x'') \,, \tag{2.149}$$

which in DeWitt's prime notation reads

$$\{\mathcal{H}_\mu, \mathcal{H}_{\nu'}\} = \kappa_{\mu\nu'}{}^{\rho''} \mathcal{H}_{\rho''} \,, \tag{2.150}$$

with primes indicating different points in space and where the Einstein summation convention is extended to integration over repeated dummy space variables. Note that in analogy with Lie algebras, the functions $\kappa_{\mu\nu'}{}^{\rho''}$ are known as structure functions.

The corresponding BRST charge is given by

$$\Omega = c^\mu \mathcal{H}_\mu + \rho^\mu \Pi_\mu - \frac{1}{2} \kappa_{\nu'\sigma''}{}^\mu c^{\nu'} c^{\sigma''} \bar{\rho}_\mu \,. \tag{2.151}$$

The gauge choice $\dot{N}^\mu = 0$ leads to the gauge fixing action

$$S_{GF} = \int_{\tau_0}^{\tau_1} \int \Pi_\mu \dot{N}^\mu \, \mathrm{d}^3x \mathrm{d}\tau \,. \tag{2.152}$$

The ghost action is

$$S_{ghost} = \int_{\tau_0}^{\tau_1} \mathrm{d}\tau \int \mathrm{d}^3x \left[\bar{\rho}_\mu \dot{c}^\mu + \rho^\mu \dot{\bar{c}}_\mu - \bar{\rho}_\mu \rho^\mu - \bar{\rho}_{\mu''} \kappa_{\nu\sigma'}{}^{\mu''} N^\nu c^{\sigma'} \right] . \tag{2.153}$$

In this gauge, the Lagrange multipliers Π_μ are decoupled from the ghosts. Evaluation of the corresponding path integrals reduces the path integrals over the lapse and shift functions to ordinary integrals.

The gauge choice $\dot{N}^\mu = 0$ suffices, but does not lead to particular simple ghosts. Teitelboim [307] instead proposed a gauge consisting of two parts. In the first interval $[\tau_0, \tau']$ we impose the gauge $\dot{N}^0 = 0, N^i = 0$, and in the second interval $[\tau', \tau_1]$ we impose $N^0 = 0, \dot{N}^i = 0$. The corresponding ghost action can be obtained by substituting $N^i = 0$ for the first interval and $N^0 = 0$ for the second interval. This gives for the first interval the ghost Hamiltonian

$$\mathcal{H}_I^{ghost} = \bar{\rho}_\mu \rho^\mu + \bar{\rho}_{i''} \kappa_{00'}{}^{i''} N^0 c^{0'} + \bar{\rho}_{0''} \kappa_{0i'}{}^{0''} N^0 c^{i'} , \tag{2.154}$$

and for the second interval

$$\mathcal{H}_{II}^{ghost} = \bar{\rho}_\mu \rho^\mu + \bar{\rho}_{0''} \kappa_{i0'}{}^{0''} N^i c^{0'} + \bar{\rho}_{k''} \kappa_{ij'}{}^{k''} N^i c^{j'} . \tag{2.155}$$

These ghosts can subsequently be combined to obtain the gauge fixed action

$$S = \int_{\tau_0}^{\tau_1} \mathrm{d}\tau \int \mathrm{d}^3x \left[\dot{h}_{ij} \pi^{ij} + \dot{X}\Pi + i\dot{C}\bar{P} + i\dot{\bar{C}}P - N \left(\mathcal{H} + \mathcal{H}^{\mathrm{ghost}} \right) \right] , \tag{2.156}$$

with a simple ghost Hamiltonian

$$\mathcal{H}^{\mathrm{ghost}}[h_{ij}; C, \bar{P}; \bar{C}, P] = i[\bar{P}P - hh^{ij} \bar{C}_{,i} C_{,j}] . \tag{2.157}$$

Here the ghosts fields $C, \bar{C}, \bar{P}, P$ satisfy the boundary conditions

$$C[x, \tau_0] = C[x, \tau_1] = \bar{C}[x, \tau_0] = \bar{C}[x, \tau_1] = 0 . \tag{2.158}$$

In the process, we obtain the Feynman propagator

$$\begin{aligned} G[h_{ij}(\tau_1), X(\tau_1); h_{ij}(\tau_0), X(\tau_0)] = \int_{N(x)=0}^{N(x)=\infty} & G[h_{ij}(\tau_1), X(\tau_1); h_{ij}(\tau_0), X(\tau_0); N(\tau_1 - \tau_0)] \\ & \times \prod_x \mathrm{d}[\ln N(x)(\tau_1 - \tau_0)] , \end{aligned} \tag{2.159}$$

with the propagator in intrinsic parameter time $N(\tau_1 - \tau_0)$ given by the path integral

$$G[h_{ij}(\tau_1), X(\tau_1); h_{ij}(\tau_0), X(\tau_0); N(\tau_1 - \tau_0)] = \int e^{iS} \prod_{x,\tau} \frac{\mathrm{d}h_{ij}\mathrm{d}\pi^{ij}\mathrm{d}C\mathrm{d}\bar{P}\mathrm{d}\bar{C}\mathrm{d}P}{2\pi} \mathcal{D}X\mathcal{D}\Pi\,. \tag{2.160}$$

Chapter 3

Lorentzian quantum cosmology

> So long as the universe had a beginning, we could suppose it had a creator. But if the universe is really completely self-contained, having no-boundary or edge, it would have neither beginning nor end: it would simply be. What place, then, for a creator?
>
> Stephen W. Hawking

Abstract

We argue that the Lorentzian path integral is a better starting point for quantum cosmology than the Euclidean version. In particular, we revisit the mini-superspace calculation of the Feynman path integral for quantum gravity with a positive cosmological constant. Instead of rotating to Euclidean time, we deform the contour of integration over metrics into the complex plane, exploiting Picard-Lefschetz theory to transform the path integral from a conditionally convergent integral into an absolutely convergent one. We show that this procedure unambiguously determines which semiclassical saddle point solutions are relevant to the quantum mechanical amplitude. Imposing "no-boundary" initial conditions, *i.e.*, restricting attention to regular, complex metrics with no initial boundary, we find that the dominant saddle contributes a semiclassical exponential factor which is precisely the *inverse* of the famous Hartle-Hawking result.

3.1 Introduction

Any theory of cosmology must provide both a successful description of the dynamics and an explanation of the initial state. Proposals for the initial quantum state include the "no-boundary" proposal of Hartle and Hawing [194, 188, 195, 189] and the tunneling wavefunction of Vilenkin [325, 326, 323, 327].

The "no-boundary" proposal is usually framed in terms of the path integral for the Euclidean (or Riemannian) version of general relativity. One motivation provided was that the Euclidean path integral would have nicer convergence properties, making it better defined, in analogy with Euclidean (Wick-rotated) quantum field theory [194]. Another was that the Euclidean approach would straightforwardly allow for the inclusion of topologically non-trivial manifolds [194]. However, the Euclidean approach to quantum gravity was only partly successful. An immediate problem is that the Euclidean action is unbounded below [159]. One well-known aspect of this is the conformal factor problem; gradients in the overall scale factor of the metric contribute negatively to the Euclidean action and typically render it unbounded below. Another aspect stems precisely from the inclusion of topologically non-trivial manifolds, whose action can be shown to be unbounded both above and below [159]. Unfortunately, this means that specifying a boundary condition, such as the "no-boundary" condition, is insufficient to uniquely define the Euclidean path integral. Additional input is required to determine the complex contour over which one should integrate Euclidean metrics [160]. In the absence of such a prescription, the Euclidean path integral for quantum gravity is ill-defined at best.

An alternate approach, followed in much of the literature, is based on solving the homogeneous Wheeler-DeWitt equation for the "wavefunction of the universe." However, there are also significant ambiguities in this approach, since there is in principle an infinite degree of freedom in the choice of boundary conditions on superspace. There are both real and complex solutions and different authors have taken different views about which of these should be taken into account. The tunneling wavefunction, for example, was developed as a particular choice for a complex solution, representing an outgoing, expanding universe only, while the "no-boundary" proposal was presented as a real solution representing a quantum superposition of an expanding universe with a collapsing one.

More recently, a holographic approach to quantum cosmology, based on the AdS/CFT correspondence, has also been advocated [200, 201]. The problem with this approach is that it requires boundary conditions which strongly influence the dynamics in the bulk. If the boundary conditions are AdS-invariant, the cosmological dynamics (including the Jeans instability on large scales) is eliminated altogether. More general boundary conditions

typically either allow instabilities which are difficult to resolve [187], or they affect the bulk dynamics so strongly that any connection to four-dimensional general relativistic cosmology is unclear [245]. More general holographic scenarios have also been proposed [249], based on analytically continuing Newton's constant and other fundamental parameters in order to obtain a well-defined dual 3d theory. While this procedure works formally for certain specific cosmological backgrounds, it does not so far seem to address more fundamental questions such as the likelihood of such backgrounds.

We here argue for what seems to us a much simpler, clearer and more general approach. We start from the Lorentzian path integral, defined as a functional integral over real Lorentzian metrics. This has several immediate advantages over the Euclidean path integral. There is no conformal factor problem. The Lorentzian path integral naturally incorporates notions of causality and unitarity, as well as boundary conditions specified in terms of the initial and final three-geometry, eliminating the corresponding ambiguities in the Wheeler-DeWitt wavefunction approach. The only disadvantage of the Lorentzian path integral is a technical, although very important one: it is a highly oscillatory integral, whose convergence is not obvious. In fact, from its definition as an infinite-dimensional integral over an infinite measure of the phase $e^{iS/\hbar}$, it is clearly *not* absolutely convergent. In simple examples, however, we find that it is *conditionally* convergent, for very simple reasons which are likely to extend to the general case. Conditionally convergent integrals dependent on the order of "summation," just as conditionally convergent series depend on the ordering of the sum, and for this reason they are more subtle to treat. A classic example is the Dirichlet integral $\int_{-\infty}^{\infty} \frac{\sin(x)}{x} \mathrm{d}x = \pi$.

Traditionally, convergence of the path integral in quantum field theory has been ensured either with Feynman's $i\epsilon$ regularization or, more rigorously, via a Wick rotation to imaginary time, the procedure which motivated the Euclidean path integral approach to quantum gravity. However, in flat space quantum field theory, Wick rotation is much more natural. It exploits the global symmetries of Minkowski spacetime to convert the Lorentz group into a compact rotation group. There are no such symmetries in a generic spacetime. Furthermore, in general relativity, the time coordinate is an arbitrary, unphysical quantity, so analytically continuing it into the complex plane, with no control over its range, has no comparable justification.

Picard-Lefschetz theory provides an alternative procedure to convert conditionally convergent integrals into absolutely convergent integrals. The idea is to deform the contour of integration from the real axis into the complex plane in such a way as to make the integral absolutely convergent. In the context of a theory in which spacetime itself is dynamical, this seems far more natural than trying to generalize the Wick rotation. One deforms the integral over physical quantities, *i.e.*, the four-geometries, which are being integrated

over, into the complex plane, while holding the three-geometry boundaries fixed. Formally, one can appeal to an infinite dimensional version of Cauchy's theorem to ensure that the result is equal to the original Lorentzian path integral. An elementary example of the Picard-Lefschetz approach was given in [164], where the Feynman propagator for a relativistic particle was derived from the path integral without any need for the $i\epsilon$ prescription or a Wick rotation. Here, we shall apply the same approach to the minisuperspace path integral for quantum gravity. This gives a well-defined answer while properly incorporating causality and the correct boundary conditions. As we shall see, Picard-Lefschetz theory neatly and unambiguously determines the combination of classical saddle point solutions contributing to semiclassical quantum gravity amplitudes. We have to admit we are puzzled as to why this method, which with hindsight seems by far the most natural and obvious one, has not, as far as we know, been previously advanced in the context of quantum cosmology.

We shall illustrate our approach in the simplest of toy cosmological models, namely a homogeneous, isotropic, closed universe with a cosmological constant Λ. We shall compute the quantum mechanical propagator with various boundary conditions, namely classical, non-classical and "no-boundary" boundary conditions. Our results for the "no-boundary" case differ from those of Hartle and Hawking, for an easily understandable reason.

The Lorentzian path integral is defined as an integral over a phase $e^{iS/\hbar}$, with the action S a real function of real dynamical variables. If one deforms the integration contour into the complex plane for these variables, running through a complex saddle point, as we shall show, one necessarily does so by "sliding down" a contour of steepest ascent from the saddle point which intersects the real axis. The real part of the exponent $\mathrm{Re}[iS_{cl}/\hbar]$, which determines the semiclassical factor in the quantum mechanical amplitude, necessarily decreases on the way down. Since the real part of the semiclassical exponent starts out zero on the real axis, *it must be negative at any relevant saddle point.* Such semiclassical factors, by this argument, can only suppress, and never enhance, a quantum mechanical amplitude.

As is well known, Euclidean quantum gravity yields a *positive* real part for the semiclassical exponent, in the case of our simple cosmology. The classical saddle point solution is just a four-sphere, continued at its equator to de Sitter spacetime. In units where $8\pi G = 1$, it yields a semiclassical factor $|e^{iS_{cl}/\hbar}| = e^{\mathrm{Re}[iS_{cl}/\hbar]} = e^{+12\pi^2/(\hbar\Lambda)}$. This is clearly inconsistent with our argument, so we can safely conclude that the Euclidean solution is not relevant to the Lorentzian path integral.

Instead, we find that there is a different classical solution, contributing precisely the inverse semiclassical factor, *i.e.*, a suppression. The reason is simply that the equations

of motion are real. If a complex solution exists, its complex conjugate must also be a solution. But the complex conjugate solution has the complex conjugate value for the classical action, so the real part of the semiclassical exponent, $\mathrm{Re}[iS_{cl}/\hbar]$, has the opposite sign. This complex conjugate saddle point therefore can be (and, we shall show, is) relevant to the Lorentzian path integral, and gives a semiclassical exponential factor of $e^{-12\pi^2/(\hbar\Lambda)}$, precisely the inverse of the Hartle-Hawking result. This is the crux of our argument, which the remainder of the chapter is devoted to fleshing out in detail.

The semiclassical factor we obtain agrees with Vilenkin's "tunneling" proposal, for this simple cosmology. Since the logic is quite different, however, it remains to be seen whether the correspondence persists for more general models. Note also that we are not employing an "inverse" Wick rotation, advocated by Linde [241]. That prescription is well known to be problematic since it leads to a divergent measure for the perturbations. Since we are always considering the Lorentzian path integral, we never perform a Wick rotation. The appropriate contours for the path integral are completely specified by requiring that a) they are continously deformable to contours running over real, Lorentzian spacetime metrics and b) they follow steepest descent contours, along which the path integral is absolutely convergent. These criteria are clearly the appropriate ones for considering semiclassical, Lorentzian amplitudes in general relativity: in our view, there are no good reasons for adopting (and many good reasons *not* to adopt) different criteria in quantum cosmology.

As our argument above already indicates, in these toy cosmologies (and, most likely, in general) the path integral over real Lorentzian metrics *cannot* be deformed to a Euclidean contour. Just to be clear, there *is* a saddle point of the Euclidean action – Hartle and Hawking's solution. And there *is* an integration contour running through this saddle point which gives a convergent result – the steepest descent contour through this saddle. However, this contour bears no relation either to the Lorentzian path integral, *or* to one taken over Euclidean metrics, which is a meaningless divergent integral. Instead, the steepest descent contour through the Euclidean (Hartle-Hawking) solution defines an intrinsically complex theory bearing no relation to quantum mechanics or the Lorentzian path integral, and from which, we claim, there is no reason to expect causal or unitary behavior to emerge.

In summary, the Lorentzian approach we take has several manifest advantages: (i) it starts from a theory with clear notions of causality and quantum-mechanical unitarity, (ii) it does not suffer from a conformal factor problem and (iii) with the Picard-Lefschetz approach to ensuring absolute convergence, it has a chance of being mathematically well defined. We emphasize that few of these detailed calculations are new or original. For the most part they recapitulate the analyses of earlier authors, some performed decades ago. Our sole claim to originality is to demonstrate that very minimal and well-founded

principles can and do substantially clarify the rules of the game.

This chapter is organised as follows. In section 3.2 we provide a simple introduction to Picard-Lefschetz theory. In section 3.3, we apply this method to the mini-superspace path integral for general relativity with a cosmological constant, for various boundary conditions. In particular, we check cases where the boundary conditions lead to purely classical evolution. Subsequently, we analyze the "no-boundary" initial condition in detail, as the main focus of this chapter. For completeness we also consider the case with non-classical boundary conditions, where both the initial and final scale factor are smaller than the waist of the de Sitter hyperboloid. In section 3.4 we relate the path integral description to the canonical formalism and the Wheeler-DeWitt equation. We summarize our findings, and point to future research directions, in section 3.5.

3.2 Picard-Lefschetz approach to oscillatory integrals

Picard-Lefschetz theory deals with oscillatory integrals like

$$I = \int_D \mathrm{d}x\, e^{iS[x]/\hbar}, \tag{3.1}$$

where $\hbar$ is a real parameter, the action $S[x]$ is a real-valued function and the integral is taken over a real domain D, usually defined by the singularities of the integrand or, in higher dimensional or path integral cases, its partial integrals. One is typically interested in the behavior of the integral for small values of the parameter $\hbar$: in quantum mechanical applications, taking $\hbar$ to zero is a nice way to study the classical limit. Picard-Lefschetz theory was originally developed and applied to ordinary integrals, in finite dimension, for example, in the work of Arnol'd et. al. [13]. More recently, Witten [342] and others have discussed its use in quantum mechanical path integrals. For example, it has been used to develop new Monte-Carlo techniques capable of addressing the notorious "sign problem" in some quantum field theories [85].

In the main part of this chapter, we have a far humbler goal. We address simple minisuperspace models of quantum cosmology, which reduce to a single one-dimensional integral. Hence, for the purposes of this brief introduction, we shall review the theory in its most trivial case. It is important to note, however, that in principle Picard-Lefschetz theory may equally be applied in higher dimensions and even, in principle, in the infinite-dimensional context relevant to physically realistic path integrals.

When faced with an integral in the form of (3.1), the idea of Picard-Lefshetz theory is to interpret $S[x]$ as a holomorphic function of $x \in \mathbb{C}$, the complex plane. Cauchy's theorem

allows us to deform the integration contour from the real domain D on the real x-axis into a contour we now call $\mathcal{C}$ in the complex x-plane, while keeping its endpoints fixed. In particular, we seek to deform $\mathcal{C}$ into a "steepest descent" contour passing through one or more critical points of $S[x]$, *i.e.*, points where $\partial_x S = 0$. By the Cauchy-Riemann equations, the real part of the exponent, $\mathrm{Re}[iS[x]]$, which controls the magnitude of the integrand, has a saddle point in the real two-dimensional (Re[x],Im[x])-plane there. The steepest descent contour through the saddle point is defined as the path along which $\mathrm{Re}[iS[x]]$ decreases as rapidly as possible.

A simple example is provided by $S[x] = x^2$, with a critical point at $x = 0$. Writing $x = \mathrm{Re}[x] + i\,\mathrm{Im}[x]$, we have $\mathrm{Re}[iS[x]] = -2\,\mathrm{Re}[x]\mathrm{Im}[x]$. The magnitude of the integrand decreases most rapidly along the contour Im[x]=+Re[x] which is the steepest descent contour. Conversely, it increases most rapidly along the contour $\mathrm{Im}[x] = -\mathrm{Re}[x]$, which is the steepest ascent contour. As we shall discuss, steepest descent contours generically lead to convergent integrals, and in this case they are known as Lefschetz thimbles $\mathcal{J}_\sigma$.

In more detail, we write the exponent $\mathcal{I} = iS/\hbar$ and its argument x in terms of their real and imaginary parts, $\mathcal{I} = h + iH$ and $x = u^1 + iu^2$. Downward flow is then defined by

$$\frac{\mathrm{d}u^i}{\mathrm{d}\lambda} = -g^{ij}\frac{\partial h}{\partial u^j}\,, \tag{3.2}$$

with λ a parameter along the flow and g_{ij} a Riemannian metric introduced on the complex plane. The real part of the exponent h decreases on such a flow away from its critical points, because $\frac{\mathrm{d}h}{\mathrm{d}\lambda} = \sum_i \frac{\partial h}{\partial u^i}\frac{\mathrm{d}u^i}{\mathrm{d}\lambda} = -\sum_i \left(\frac{\partial h}{\partial u^i}\right)^2 < 0$, with the fastest rate of decrease occuring in the direction of "steepest descent", which maximises the magnitude of the gradient. Defining the latter requires that we introduce a metric. Witten points out that the freedom to choose this metric may be exploited in interesting ways [342].

For the simple examples we discuss here, the obvious metric $\mathrm{d}s^2 = |\mathrm{d}x|^2$ is sufficient. Defining complex coordinates, $(u, \bar{u}) = \big((\mathrm{Re}([x] + i\mathrm{Im}[x]), (\mathrm{Re}[x] - i\mathrm{Im}[x])\big)$, the metric is $g_{uu} = g_{\bar{u}\bar{u}} = 0$, $g_{u\bar{u}} = g_{\bar{u}u} = 1/2$. Then $h = (\mathcal{I} + \bar{\mathcal{I}})/2$ and (3.2) becomes

$$\frac{\mathrm{d}u}{\mathrm{d}\lambda} = -\frac{\partial\bar{\mathcal{I}}}{\partial\bar{u}}, \quad \frac{\mathrm{d}\bar{u}}{\mathrm{d}\lambda} = -\frac{\partial\mathcal{I}}{\partial u}\,. \tag{3.3}$$

The imaginary part of the exponent $H = \mathrm{Im}[iS/\hbar]$ is conserved along these flows, since

$$\frac{\mathrm{d}H}{\mathrm{d}\lambda} = \frac{1}{2i}\frac{\mathrm{d}(\mathcal{I} - \bar{\mathcal{I}})}{\mathrm{d}\lambda} = \frac{1}{2i}\left(\frac{\partial\mathcal{I}}{\partial u}\frac{\mathrm{d}u}{\mathrm{d}\lambda} - \frac{\partial\bar{\mathcal{I}}}{\partial\bar{u}}\frac{\mathrm{d}\bar{u}}{\mathrm{d}\lambda}\right) = 0\,. \tag{3.4}$$

Thus the integrand $e^{iS[x]/\hbar}$ – which was a purely oscillatory factor in the original integral – does not oscillate at all when evaluated along a downward flow (see Fig. 3.1). Instead,

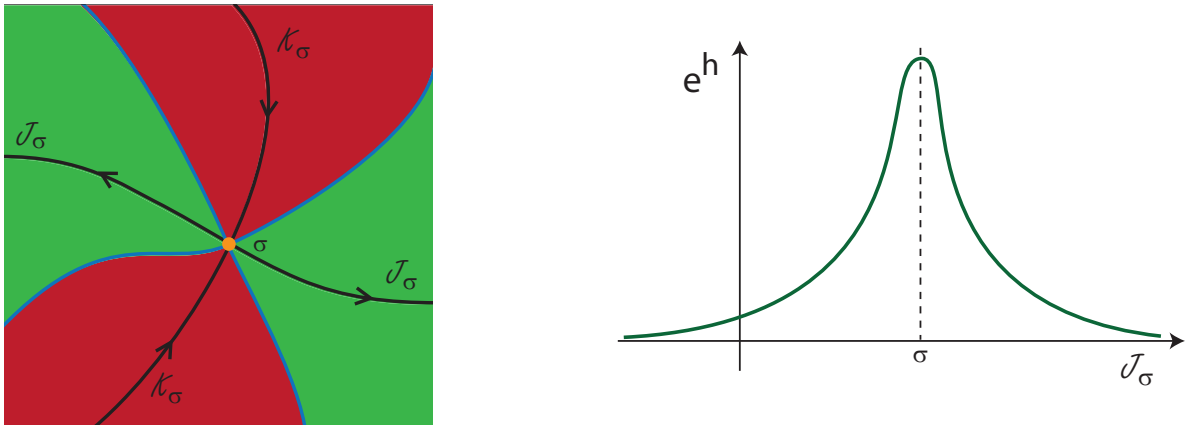


Figure 3.1: *Left panel:* From a saddle point σ emanate upward ($\mathcal{K}_\sigma$) and downward ($\mathcal{J}_\sigma$) flows, which are located in the wedges J_σ (in green) and K_σ (in red) respectively, defined as the regions where the h-function is lower (higher) than its value at the saddle, respectively. The arrows along the flows indicate the direction of descent, and the downward flow $\mathcal{J}_\sigma$ is known as a Lefschetz thimble. The wedges are separated by blue lines along which h is constant and equal to the value at the saddle point $h(p_\sigma.)$ *Right panel:* Along a Lefschetz thimble the real part h of the exponent decreases as fast as possible, ensuring an absolutely convergent integral.

it decreases monotonically so that the integral converges absolutely and "as rapidly as possible." For a downward flow originating at a saddle, λ runs from $-\infty$ at the saddle point to positive values as h decreases. The Lefschetz thimble associated with a given saddle is defined as the set of downward flows leaving the saddle in this way.

Analogously, upward flows are defined via

$$\frac{\mathrm{d}u^i}{\mathrm{d}\lambda} = +g^{ij}\frac{\partial h}{\partial u^j}\,, \tag{3.5}$$

with H likewise being conserved along these flows. Every critical point has an upward flow which, in analogy to the downward flow, is labelled $\mathcal{K}_\sigma$.

There is a complication in this argument which it is convenient to resolve. It is possible, in principle, for a steepest descent contour from one saddle point p_σ to terminate on another saddle point p'_σ, as $\lambda \to \infty$ so that it coincides with a steepest ascent contour from p'_σ. Such a situation is generically unlikely, but it may occur as the result of a symmetry. For example, if $S[x]$ is a real function of x (meaning it is real for all real x), then any complex saddle point necessarily comes with a complex conjugate partner. The

function $h = i(S[x] - S[\bar{x}])/(2\hbar)$ is generally different at these two saddles, whereas the imaginary part of the exponent H is the same so that, indeed, both the steepest descent flow from the higher saddle and the steepest ascent flow from the lower saddle follow the line $\mathrm{Re}[x] = const.$ Such a situation indeed arises with no-boundary initial conditions, with the relevant contours exhibited in Fig. 5 below.

Such a degeneracy between steepest ascent and steepest descent contours may generally be removed by adding an infinitesimal perturbation to $S[x]$, and defining the contour $\mathcal{C}$ in the limit as the perturbation is taken to zero. In this limit the contribution of the perturbation to the integral is negligible. However, a generic perturbation will break the degeneracy between the values of the imaginary part of the exponent H at the two saddle points, making it impossible, according to (3.4), for steepest ascent and descent flows from the two critical points to coincide. Of course, if a symmetry was responsible for the degeneracy, as we have discussed for the case where $S[x]$ is real, then the perturbation must violate the symmetry if it is to remove the degeneracy. So if $S[x]$ being real is responsible for the degeneracy, an imaginary perturbation will be needed to remove it. This is not a problem, however, since as explained, in the limit that the perturbation is taken to zero, its influence on the integral is negligible.

Once all such degeneracies are removed, we are left with a one-to-one correspondence between saddle points p_σ and the associated steepest ascent and descent contours $\mathcal{J}_\sigma$ and $\mathcal{K}_\sigma$. The generic situation is then that any steepest descent contour from a saddle ends on a singularity where $h \to -\infty$, and any steepest ascent contour likewise ends on a singularity where $h \to +\infty$.

Thus, Lefshetz thimbles and upward flows only intersect at a single critical point, the one where both are defined. With a suitable choice of orientation, we can write for the intersection number

$$\mathrm{Int}(\mathcal{J}_\sigma, \mathcal{K}_{\sigma'}) = \delta_{\sigma\sigma'}. \tag{3.6}$$

Our objective is to deform the original integral (3.1) into one evaluated over a sum of Lefschetz thimbles. That is, we would like to write

$$\mathcal{C} = \sum_\sigma n_\sigma \mathcal{J}_\sigma, \tag{3.7}$$

in a homological sense, for some integers n_σ which may take the values 0 or ± 1 when accounting for the orientation of the contour over each thimble. It follows from these equations that $n_\sigma = \mathrm{Int}(\mathcal{C}, \mathcal{K}_\sigma) = \mathrm{Int}(D, \mathcal{K}_\sigma)$, since the intersection number is topological and will not change if we deform the contour $\mathcal{C}$ back to the original, real domain D. Thus a necessary and sufficient condition for a given thimble $\mathcal{J}_\sigma$ to be relevant is that a steepest

ascent contour from the critical point p_σ intersects the original, real integration domain D. In this circumstance, intuitively, there is no obstacle to smoothly "sliding" the intersection point from the real axis along $\mathcal{K}_\sigma$ down to p_σ, and in the process deforming the original integration contour onto the the thimble $\mathcal{J}_\sigma$. This is the argument we alluded to in the introduction, showing that if one starts from a real Lorentzian theory, one *never* obtains semiclassical enhancement factors such as are found in the Euclidean approach. This intersection rule formally follows from relative homology. See appendix 3.B for a sketch of the proof.

In one complex dimension, the way this works is that the original integral along the real x-axis is deformed into a series of thimbles. With the appropriate choice of orientation, adjacent thimbles end and start on singularities of the h-function, so that there is no obstacle to deforming the combined contour back onto the real x-axis. The two "free ends" in the sum over thimbles, corresponding to the first and last steepest descent contours, from the first and last critical points, run to singularities of h in a complex direction determined by the steepest descent flow. In order to show that the original integral I equals the sum of integrals over thimbles, we must show that the original integration contour, which approaches the initial and final singularities of h along the real x-axis, can be deformed into one which ends on initial and final steepest descent contours which approach the same singularities from a different direction. This requires that the integral taken along an "arc" drawn around the singularity vanishes in the limit that the arc is taken closer and closer to the singularity.

We shall now illustrate this behavior in the integral which arises in the simplest models of minisuperspace quantum cosmology. As we shall see in the next section, this takes the form

$$\int_{0^+}^{\infty} \frac{dN}{\sqrt{N}} e^{if(N)/\hbar}, \tag{3.8}$$

where $f(N)$ is holomorphic in N over the relevant domain. The integrand possesses singularities at $N = 0$ and $N = \infty$ and the contour of integration runs from one to the other, over all positive values of N. We wish to show that it is possible to deform this contour to a sum of the relevant steepest descent contours, the first and last of which approach the singularities of the integrand at some finite angle with respect to the real N-axis.

Consider first a singularity of $f(N)$ which occurs at infinite N. We take the original integral up to some large positive value, N_0. It is convenient to change variables to $N = (\ln z)^2$, so that (3.8) becomes $2\int_1^{z_0} \frac{dz}{z} e^{if\left((\ln z)^2\right)/\hbar}$, with $z_0 = e^{\sqrt{N_0}}$. The relevant steepest descent trajectory at large $|z|$ will be determined by the term with the largest power of N in $f(N)$. It will run to infinity at some angle θ with respect to the real z-axis. To show

that the original integral taken up to some large real value z_0 is accurately approximated by the steepest descent integral taken out to $(|z|, \theta) = (z_0, \theta_0)$, we need to show that the integral along an arc at fixed $|z|$ with the angle θ running from 0 to θ_0, becomes negligible as z_0 is taken to infinity. Assume, for example, that $f(N) = aN$ at large $|N|$, with a positive. Now set $z = e^{\sqrt{N_0}+i\theta}$ so that the integral along the arc at fixed $|z|$ becomes

$$2\int_0^{\theta_0} \frac{dz}{z} e^{ia(\ln z)^2/\hbar} = 2i\int_0^{\theta_0} d\theta e^{ia(\sqrt{N_0}+i\,\theta)^2/\hbar} \equiv iI_0 \rightarrow |I_0| < 2\int_0^{\theta_0} d\theta e^{-2a\sqrt{N_0}\theta/\hbar} < \frac{\hbar}{a\sqrt{N_0}} \quad (3.9)$$

where we used a standard Schwarz-type inequality, and the fact that the last integral is bounded by its value when taken over an infinite range. We have thus bounded the magnitude of the integral along the arc at fixed $|z|$, by a quantity which tends to zero as N_0 tends to infinity. Hence in the limit of large N_0, the original contour may indeed be deformed to one ending on the steepest descent contour at the same value of N_0, with negligible change in the value of the integral. The limit $N_0 \rightarrow \infty$ can now be taken, with the conclusion that the two integrals are identical in this limit. It is not hard to generalize this argument to any holomorphic $f(N)$ behaving as a power of N at large N: one just needs to choose N_0 large enough to ensure that all terms in the real part of the exponent in the analog of (3.9) are bounded by some finite multiple of the term involving the highest power of N_0.

Similarly, the steepest descent contour approaches the singularity at $N = 0$ along a complex direction. For example, if $f(N) \sim -a/N$ as $N \rightarrow 0$, with a positive, then $N = 0$ is approached from positive imaginary values. To show that the original integral (3.9) taken along the real N-axis equals the steepest descent integral, we cut the former off at some small real $N = \epsilon_0$. Setting $N = 1/(\ln z)^2$, (3.9) becomes $2\int_1^{z_0} \frac{dz}{z}\frac{1}{(\ln z)^2} e^{-ia\left((\ln z)^2\right)/\hbar}$ with $z_0 = e^{1/\sqrt{\epsilon_0}}$. By Cauchy's theorem, the original integral taken over $N > \epsilon_0$ may be deformed into an integral along an arc $z = e^{1/\sqrt{\epsilon_0}} e^{-i\theta}$, plus the steepest descent integral taken from the arc's intersection with the steepest descent contour. On the arc, $|\ln z|^2 > 1/\epsilon_0$, so the integral along the arc is bounded by $2\epsilon_0 \int d\theta e^{2a\theta/(\hbar\sqrt{\epsilon_0})} < \hbar\epsilon_0^{\frac{3}{2}}/a$ and hence vanishes as $\epsilon_0 \rightarrow 0$. Therefore the steepest descent integral and the original Lorentzian integral give the same result in the limit as the cutoff is removed.

Once we have deformed the contour from the real axis to run through a set of thimbles associated with the contributing critical points, we have:

$$I = \int_D \mathrm{d}x\, e^{iS[x]/\hbar} = \int_{\mathcal{C}} \mathrm{d}x\, e^{iS[x]/\hbar} = \sum_\sigma n_\sigma \int_{\mathcal{J}_\sigma} \mathrm{d}x\, e^{iS[x]/\hbar}. \quad (3.10)$$

As (3.10) indicates, typically more than one Lefschetz thimble contributes to the Lorentzian

path integral, with given boundary conditions, even in mini-superspace quantum cosmology.

The integral taken over a thimble is absolutely convergent if

$$\left|\int_{\mathcal{J}_\sigma} \mathrm{d}x e^{iS[x]/\hbar}\right| \leq \int_{\mathcal{J}_\sigma} |\mathrm{d}x| \left|e^{iS[x]/\hbar}\right| = \int_{\mathcal{J}_\sigma} |dx| e^{h(x)} < \infty \,. \tag{3.11}$$

Defining the length along the curve as $l = \int |dx|$, the integral will converge if $h(x(l)) < -\ln(l) + A$, for some constant A, as $l \to \infty$, which is a rather weak requirement.

We have then expressed the original integral as a sum of absolutely convergent steepest descent integrals. In an expansion in $\hbar$, we have

$$I = \int_D \mathrm{d}x \, e^{iS[x]/\hbar} = \sum_\sigma n_\sigma \, e^{i\,H(p_\sigma)} \int_{\mathcal{J}_\sigma} e^h \mathrm{d}x \approx \sum_\sigma n_\sigma \, e^{iS(p_\sigma)/\hbar} \left[A_\sigma + \mathcal{O}(\hbar)\right], \tag{3.12}$$

where A_σ represents the result of the leading-order Gaussian integral about the critical point p_σ. Sub-leading terms may be evaluated perturbatively in $\hbar$. In the case of degenerate h, a similar expansion applies – we will encounter such an example later in this chapter.

3.3 Minisuperspace Lorentzian path integral

We here consider a universe with a positive cosmological constant Λ, described by the action

$$S = \frac{1}{2}\int_{\mathcal{M}} \mathrm{d}^4x\sqrt{-g}\,(R - 2\Lambda) + \int_{\partial\mathcal{M}} \mathrm{d}^3y\sqrt{g^{(3)}}K \,, \tag{3.13}$$

where we have set $8\pi G = 1$. The second term, involving the 3-metric $g^{(3)}_{ij}$ and the trace of the second fundamental form K of the boundary $\partial\mathcal{M}$, is needed to ensure the variational principle yields the Einstein equations if the boundary geometries are held fixed. For simplicity, we truncate the theory to the simplest cosmologies, represented by the line element

$$\mathrm{d}s^2 = -N(t)^2\mathrm{d}t^2 + a(t)^2\mathrm{d}\Omega_3^2 \,, \tag{3.14}$$

with $\mathrm{d}\Omega_3^2$ the metric of a homogeneous, isotropic 3-dimensional space with curvature k. This is a gross simplification of the original theory – we no longer have propagating gravitational waves – but we retain a dynamical scale factor $a(t)$ as well as diffeomorphism invariance in the timelike coordinate t, and these will be sufficient for us to illustrate many key features of Lorentzian quantum cosmology.

The Feynman path integral for the reduced theory is

$$G[a_1; a_0] = \int \mathcal{D}N\mathcal{D}\pi\mathcal{D}a\mathcal{D}p\mathcal{D}C\mathcal{D}\bar{P}\, e^{\frac{i}{\hbar}\int_0^1 \left[\dot{N}\pi + \dot{a}p + \dot{C}\bar{P} - NH\right]\mathrm{d}t}\,, \tag{3.15}$$

where, in addition to a, N and the fermionic ghost C, we have introduced the conjugate momenta p, π and $\bar{P}$, and the corresponding Liouville measure. Without loss of generality, we can choose the range of the time coordinate to be $0 \leq t \leq 1$. The Hamiltonian constraint $H[a, p; N, \pi; C, \bar{P}] = H_{EH}[a, p] + H_g[N, \pi; C, \bar{P}]$ consists of the Einstein-Hilbert Hamiltonian H_{EH}, in our case a minisuperspace Hamiltonian, and a Batalin, Fradkin and Vilkovisky (BFV) ghost Hamiltonian H_g[1]. The ghost is necessary since the minisuperspace action is diffeomorphism invariant. The ghost term breaks time reparametrization symmetry and fixes the proper-time gauge $\dot{N} = 0$. For a detailed discussion of the BFV ghost in this setting see Teitelboim [307, 309] and Halliwell [184]. For minisuperspace models, most of the path integrals can be performed analytically, yielding

$$G[a_1; a_0] = \int_{0^+}^{\infty} \mathrm{d}N \int_{a=a_0}^{a=a_1} \mathcal{D}a\, e^{iS(N,a)/\hbar}\,, \tag{3.16}$$

which has a very simple interpretation. The path integral $\int \mathcal{D}a e^{iS(N,a)/\hbar}$ represents the quantum mechanical amplitude for the universe to evolve from a_0 to a_1 in a proper time N. The integral over the lapse function indicates that we should consider paths of every proper duration $0 < N < \infty$. Teitelboim [309] showed that this choice of integration domain leads to the causal ordering of the a_0 and a_1, *i.e.*, a_0 precedes a_1. This allows us to describe both an expanding $a_1 > a_0$ and a contracting $a_1 < a_0$ universe, since the direction of the arrow of time is determined by the Feynman propagator and not by the choice of boundary conditions. For an illustration see Fig. 3.2.

The action in (3.16) reduces to

$$S = 2\pi^2 \int_0^1 \mathrm{d}t N \left(-3a\frac{\dot{a}^2}{N^2} + 3ka - a^3\Lambda \right)\,. \tag{3.17}$$

We are faced with a functional integral over $a(t)$, and an ordinary integral over the proper time N. The former may be performed in the semiclassical approximation. Notice first

[1]The Batalin, Fradkin and Vilkovisky ghost is an extension of the Fadeev-Popov ghost [30]. The Fadeev-Popov ghost is based on the BRST symmetry. In particular, the constraint algebra forms a Lie algebra. In general relativity the constraint algebra does not close, which is why the BFV quantization is required. For minisuperspace we have only one constraint, H, for which the constraint algebra trivially closes. Thus the distinction is inessential here, but the BFV quantization is nevertheless preferable.

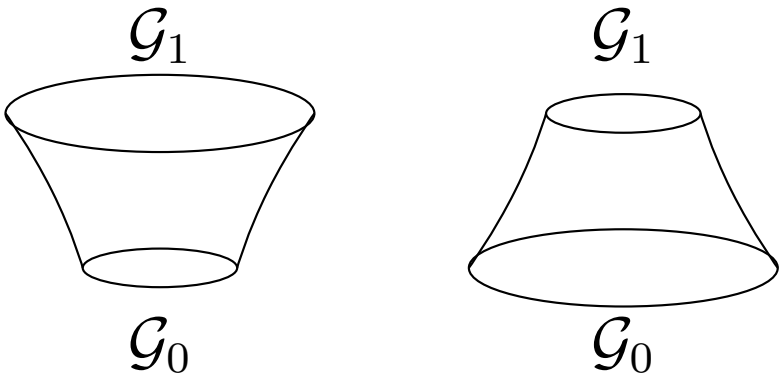


Figure 3.2: A pictorial description of the Feynman propagator, with $\mathcal{G}_0$ and $\mathcal{G}_1$ the initial and final three-geometry. *Left:* an expanding phase. *Right:* a contracting phase.

that the classical equations for $a(t)$ and N are real. In fact, the equations of motion for $a(t)$ yield a unique solution in the form $a = a_c(Nt)$, for arbitrary initial and final a_0 and a_1. However, the constraints which follow from varying N in generally can only be satisfied by complex N. Since the equations of motion are real, it follows that solutions for N come in complex conjugate pairs. Also, reversing the sign of N is classically equivalent to reversing the arrow of time, a symmetry of the classical equations. Hence we may anticipate that, quite generally, there will be four solutions for N, with only two of them being distinct after time reversal symmetry is taken into account.

In fact, we can simplify the calculation by noticing that redefining the lapse function $N(t) \to N(t)/a(t)$ renders the action (3.17) quadratic in $q(t) \equiv a(t)^2$, allowing the path integral over $q(t)$ to be performed exactly [184]. (Actually, there is a subtlety since such a redefinition alters the path integral measure. More fundamentally, one must ensure that the quantum mechanical propagator is properly covariant under such changes of variable. As discussed in [184, 164], the starting point for constructing the propagator is a proper ordering of the Hamiltonian operator. This ordering is determined by covariance under changes of variables including the one just given. With this correction, involving the Ricci curvature on superspace, the quantum Hamiltonians of the theories expressed in terms of a or $q = a^2$ are equivalent. In the case of the redefinition considered here, the correction term is only important at small a. For simplicity, as well as consistency with earlier treatments, we shall ignore it in the leading, semiclassical analysis we perform in this chapter.)

In these new variables, the action (3.17) becomes

$$S = 2\pi^2 \int_0^1 \mathrm{d}t \left(-\frac{3}{4N}\dot{q}^2 + N(3k - \Lambda q) \right) . \tag{3.18}$$

The equation of motion and the constraint following from this action are

$$\ddot{q} = \frac{2\Lambda}{3} N^2; \quad \frac{3}{4N^2}\dot{q}^2 + 3k = \Lambda q \, . \tag{3.19}$$

With boundary conditions $q(0) = q_0$ and $q(1) = q_1$, the general solution to the first equation (before imposing the constraint) is

$$\bar{q} = \frac{\Lambda}{3} N^2 t^2 + \left(-\frac{\Lambda}{3} N^2 + q_1 - q_0 \right) t + q_0 \, . \tag{3.20}$$

Writing the full solution, which does satisfy the constraint as

$$q(t) = \bar{q}(t) + Q(t) \, , \tag{3.21}$$

the path integral becomes

$$G[q_1; q_0] = \int_0^\infty \mathrm{d}N e^{2\pi^2 i S_0/\hbar} \int_{Q[0]=0}^{Q[1]=0} \mathcal{D}Q e^{2\pi^2 i S_2/\hbar} \, , \tag{3.22}$$

with

$$S_0 \;=\; \int_0^1 \mathrm{d}t \left(-\frac{3}{4N}\ddot{\bar{q}}^2 + 3kN - N\Lambda\bar{q} \right) , \quad S_2 = -\frac{3}{4N} \int_0^1 \mathrm{d}t \, \dot{Q}^2 \, . \tag{3.23}$$

The path integral over Q is Gaussian and can be evaluated exactly:

$$\int_{Q[0]=0}^{Q[1]=0} \mathcal{D}Q e^{2\pi^2 i S_2/\hbar} = \sqrt{\frac{3\pi i}{2N\hbar}} \, . \tag{3.24}$$

The propagator thus reduces to an ordinary integral

$$G[q_1; q_0] = \sqrt{\frac{3\pi i}{2\hbar}} \int_0^\infty \frac{\mathrm{d}N}{N^{1/2}} e^{2\pi^2 i S_0/\hbar} . \tag{3.25}$$

Equation (3.25) is an oscillatory integral, to which we apply the methods of the previous section. We lift the lapse N to the complex plane and regard the boundary values 0 and ∞ of the integral as points on the Riemann sphere. The action S_0 can be explicitly evaluated,

$$S_0 = N^3 \frac{\Lambda^2}{36} + N \left(-\frac{\Lambda}{2}(q_0 + q_1) + 3k \right) + \frac{1}{N} \left(-\frac{3}{4}(q_1 - q_0)^2 \right) . \tag{3.26}$$

The action S_0 has four saddle points in the complex plane, which are solutions of

$$\partial S_0/\partial N = \Lambda^2 N_s^4 + \left(-6\Lambda(q_0+q_1)+36k\right) N_s^2 + 9(q_1-q_0)^2 = 0\,, \tag{3.27}$$

given by

$$N_s = c_1 \frac{3}{\Lambda}\left[\left(\frac{\Lambda}{3}q_0 - k\right)^{1/2} + c_2\left(\frac{\Lambda}{3}q_1 - k\right)^{1/2}\right], \tag{3.28}$$

with $c_1, c_2 \in \{-1, 1\}$. The action evaluated at these saddle points is given by

$$\begin{aligned} S_0^{saddle} &= N_s^3\frac{\Lambda^2}{36} + N_s\left(-\frac{\Lambda}{2}(q_0+q_1)+3k\right) + \frac{1}{N_s}\left(-\frac{3}{4}(q_1-q_0)^2\right) \\ &= \frac{1}{N_s}\left[N_s^4\frac{\Lambda^2}{36} + N_s^2\left(-\frac{\Lambda}{2}(q_0+q_1)+3k\right) - \frac{3}{4}(q_1-q_0)^2\right] \\ &= \frac{1}{N_s}\left[-\frac{\Lambda^2}{18}N_s^4 - \frac{3}{2}(q_1-q_0)^2\right] \\ &= -c_1\frac{6}{\Lambda}\left[\left(\frac{\Lambda}{3}q_0 - k\right)^{3/2} + c_2\left(\frac{\Lambda}{3}q_1 - k\right)^{3/2}\right], \end{aligned} \tag{3.29}$$

Each of these four saddle points corresponds to a Lefschetz thimble $\{\mathcal{J}_\sigma\}$, and a steepest ascent contour $\{\mathcal{K}_\sigma\}$. Each is also associated with wedges J_σ, K_σ in which the real part of the exponent $iS/\hbar$ is respectively lower and higher than the saddle point value. Writing the original integration contour in terms of the Lefschetz thimbles

$$(0^+, \infty) = \sum_\sigma n_\sigma \mathcal{J}_\sigma\,, \tag{3.30}$$

we approximate the propagator using the saddle point approximation in the limit $\hbar \to 0$,

$$\begin{aligned} G[q_1;q_0] &= \sum_\sigma n_\sigma \sqrt{\frac{3\pi i}{2\hbar}}\int_{\mathcal{J}_\sigma}\frac{\mathrm{d}N}{N^{1/2}}e^{2\pi^2 i S_0/\hbar} \\ &\approx \sum_\sigma n_\sigma \sqrt{\frac{3\pi i}{2\hbar}}\frac{e^{2\pi^2 i S_0^{saddle}/\hbar}}{N_s^{1/2}}\int_{\mathcal{J}_\sigma}\mathrm{d}N e^{\frac{i\pi^2}{\hbar}S_{0,NN}(N-N_s)^2}\left[1+\mathcal{O}\left(\hbar^{1/2}\right)\right] \\ &\approx \sum_\sigma n_\sigma \sqrt{\frac{3\pi i}{2\hbar}}\frac{e^{2\pi^2 i S_0^{saddle}/\hbar}}{N_s^{1/2}}e^{i\theta_\sigma}\int_{\mathcal{J}_\sigma}\mathrm{d}n e^{-\frac{\pi^2}{\hbar}|S_{0,NN}|n^2}\left[1+\mathcal{O}\left(\hbar^{1/2}\right)\right] \\ &\approx \sum_\sigma n_\sigma \sqrt{\frac{3i}{2N_s|S_{0,NN}|}}e^{i\theta_\sigma}e^{2\pi^2 i S_0^{saddle}/\hbar}\left[1+\mathcal{O}\left(\hbar^{1/2}\right)\right], \end{aligned} \tag{3.31}$$

where we defined $N - N_s \equiv ne^{i\theta}$ with n real and θ being the angle of the Lefschetz thimble with respect to the positive real N axis.

The intersection coefficient n_σ, the angle θ_σ and action at the saddle point S_0^{saddle} all depend on the boundary conditions q_0 and q_1 and the spatial curvature k. In particular, saddle points can become relevant or irrelevant as the boundary conditions are varied. Earlier approaches amount to choosing particular contour in the complex N plane "by hand," on the basis of some preconceived notions. However, as we argued in section 3.2, the virtue of the Lorentzian path integral combined with Picard-Lefschetz theory is that the proper combination of saddle points and relative phases between them is completely fixed.

As can be seen from equation (3.28), for spherical three-geometries, $k = 1$, the saddle points may be located off the real axis and be anywhere in the complex plane, while for the flat and hyperbolic case the saddle points are real. Complex saddle points imply non-classical behaviour since the propagator becomes dominated by non-Lorentzian geometries. In the following sections we concentrate on spherical expanding universes. We study the saddle point approximation (3.31) in four qualitatively different configurations:

- For $q_1 \geq q_0 > \frac{3}{\Lambda}$ the saddle points are real. These boundary conditions represent a classical universe. This case is studied in section 3.3.1.
- For $q_1 > \frac{3}{\Lambda} > q_0$ one of the roots becomes imaginary. This case includes the "no-boundary" proposal and is studied in section 3.3.2.
- The limiting case between the classical and quantum phase, for which $q_1 \geq q_0 = \frac{3}{\lambda}$, is studied in section 3.3.3.
- For $\frac{3}{\Lambda} > q_1 \geq q_0$ both square roots become imaginary and both q_0 and q_1 are in the classically disallowed region. We study this case in section 3.3.4.

3.3.1 Classical boundary conditions

For classical boundary conditions $q_1 \geq q_0 > \frac{3}{\Lambda}$, the four saddle points are real, see Fig. 3.3 for the corresponding lines of steepest descent and ascent. The two positive saddle points

$$N_{s\pm} = \sqrt{\frac{3}{\Lambda}} \left[\left(q_1 - \frac{3}{\Lambda} \right)^{1/2} \pm \left(q_0 - \frac{3}{\Lambda} \right)^{1/2} \right] \tag{3.32}$$

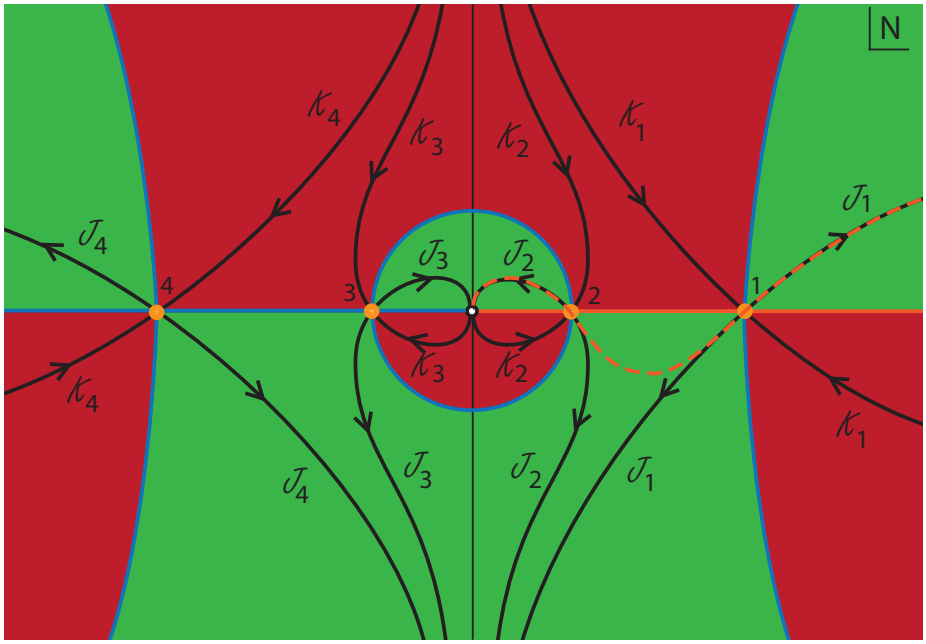


Figure 3.3: A sketch of the wedges and flow lines emanating from the saddle points in the complex N plane, for classical boundary conditions $q_1 > q_0 > \frac{3}{\Lambda}$. The Lefschetz thimbles $\mathcal{J}_\sigma$ reside within the green wedges J_σ (within which the magnitude of the integrand is smaller than at the corresponding saddle point), while the contours of steepest ascent $\mathcal{K}_\sigma$ reside within the red wedges K_σ (within which the magnitude of the integrand is larger than at the corresponding saddle point). The arrows indicate the direction of steepest descent. The original integration contour along the positive real axis is shown in orange, and runs through two saddle points in this case. The deformed contour along which the integral is absolutely convergent comprises the thimbles $\mathcal{J}_1$ and $\mathcal{J}_2$: the dashed orange line indicates how the original contour is deformed onto to these thimbles. Note that neither the flow lines, nor the original integration contour, include the point at $N = 0$.

contribute to the integral since their curves of steepest ascent trivially intersects the original interval $(0, \infty)$. The two negative saddle points do not contribute to the propagator.

The equation of motion is solved by the de Sitter space solution. For classical boundary conditions there exist two solutions: either q_0, q_1 both sit on the same side of the waist of the de Sitter hyperboloid, or they are separated by the waist. Using the classical solution $\bar{q}(t)$ (see equation (3.20)) we can study the two saddle points. At $t = 0$,

$$\frac{\mathrm{d}\bar{q}}{\mathrm{d}t} = -\frac{\Lambda}{3}N_{s\pm}^2 + q_1 - q_0 = 2\left(-q_0 + \frac{3}{\Lambda}\right) \mp 2\sqrt{\left(q_0 - \frac{3}{\Lambda}\right)\left(q_1 - \frac{3}{\Lambda}\right)}\,. \tag{3.33}$$

We observe that N_{s+} corresponds to a decreasing solution. The waist of the de Sitter space sits between the specified boundaries. The other saddle point N_{s-} corresponds to an increasing solution. In this case both boundaries sit on the same side of the waist.

Figure 3.3 illustrates the Lefschetz thimbles corresponding to the saddle points. The first Lefschetz thimble runs from the origin at $N = 0$ up in the positive imaginary N direction[2], curves around, moves through N_{s-}, and asymptotically approaching the negative imaginary axis. The second thimble runs up from the negative imaginary axis, through N_{s+} and asymptotes to positive $Re(N)$ at an angle of $\pi/6$ [3]. Note that the sum of these two thimbles is indeed deformable to the positive real N axis. In Fig. 3.3 the integration contour that runs through the saddle points along the Lefschetz thimbles $\mathcal{J}_{1,2}$ is shown by a dashed orange line – along this contour the integral is manifestly convergent.

Since we have two relevant saddle points, the saddle point approximation of the propagator (3.31) is the sum of two phases,

$$\begin{aligned} G[q_1; q_0] &\approx \left(\frac{3i}{4\Lambda\sqrt{(q_0 - \frac{3}{\Lambda})(q_1 - \frac{3}{\Lambda})}}\right)^{1/2} \left[\, e^{-i\frac{\pi}{4}} e^{iS(N_{s-})/\hbar} + \, e^{i\frac{\pi}{4}} e^{iS(N_{s+})/\hbar}\right] \\ &\approx \frac{e^{i\frac{\pi}{4}} 3^{1/2}}{[(\Lambda q_0 - 3)(\Lambda q_1 - 3)]^{1/4}} \cos\left(\frac{4\pi^2\Lambda^{1/2}}{3^{1/2}\hbar}\left(q_0 - \frac{3}{\Lambda}\right)^{3/2} - \frac{\pi}{4}\right) e^{-i\frac{4\pi^2\Lambda^{1/2}}{3^{1/2}\hbar}\left(q_1 - \frac{3}{\Lambda}\right)^{3/2}} \end{aligned} \tag{3.34}$$

The factors $e^{\pm i\frac{\pi}{4}}$, arise from aligning the fluctuation integrals with the Lefschetz thimbles (cf. the orange dashed line), as explained in more detail in section 3.3.2.

[2]Note that this Lefschetz thimble does not include the point $N = 0$ itself. In the small N limit, the h-function is approximated by $Re(-i/N)$ and this becomes arbitrarily negative as N tends towards $N = 0$ along the positive imaginary axis, without actually reaching $N = 0$. This is just as well, as the original integration contour also does not include $N = 0$ at which point the metric would be singular.

[3]This angle is easy to determine: the flow lines correspond to lines of constant imaginary part of the integrand, and for large N this means constant $Re(N^3)$.

In this simple model, we have interference effects between the two possible classical trajectories linking our initial and final conditions. In more realistic models involving interactions with other fields, one might expect the two terms in the transition amplitude to decohere. This would suppress the interference.

3.3.2 No-boundary conditions

Implications of Picard-Lefschetz theory

The "no-boundary" conditions were proposed by Hartle and Hawking as a theory of initial conditions for the universe [194, 188, 195]. The idea is that in the path integral one should sum only metrics whose only boundary is provided by the final spatial hypersurface (corresponding to the current state of the universe). To implement "no-boundary" conditions, we must take $q_0 = 0$ and find a 4-metric which is regular there. This is possible for positive k. The "no-boundary" condition is supplemented with the constraint equation (3.19) evaluated at $q = 0$,

$$\dot{q}^2 = -4N^2 k \qquad (q = 0) \ . \tag{3.35}$$

We will take the final boundary to correspond to a late time configuration, where the universe has become large, $q_1 > \frac{3}{\Lambda}$. The saddle points of the action are given by

$$N_{s,nb1} = +\frac{3}{\Lambda}\left[i \pm \left(\frac{\Lambda}{3}q_1 - 1\right)^{1/2}\right] , \quad N_{s,nb2} = -\frac{3}{\Lambda}\left[i \pm \left(\frac{\Lambda}{3}q_1 - 1\right)^{1/2}\right] , \tag{3.36}$$

with corresponding actions

$$S_{0,nb1} = -\frac{6}{\Lambda}\left[-i \pm \left(\frac{\Lambda}{3}q_1 - 1\right)^{3/2}\right] , \quad S_{0,nb2} = +\frac{6}{\Lambda}\left[-i \pm \left(\frac{\Lambda}{3}q_1 - 1\right)^{3/2}\right] . \tag{3.37}$$

Note that saddle points in the upper half plane lead to a $e^{i2\pi^2 S_0} \sim e^{-12\pi^2/(\hbar\Lambda)}$, while those in the lower half plane lead to $e^{i2\pi^2 S_0} \sim e^{+12\pi^2/(\hbar\Lambda)}$.

Given the saddle points, we can determine the wedges and the curves of steepest descent and ascent emanating from them. We use the fact that curves with $\mathrm{Re}(iS_0)$ specify the boundaries of the wedges, and that $\mathrm{Im}(iS_0)$ is constant along the flow lines to determine them numerically – see also [186]. For the case of interest to us, the wedge boundaries and flow lines are shown in Fig. 3.4, while the directions of the flows are sketched in Fig. 3.5.

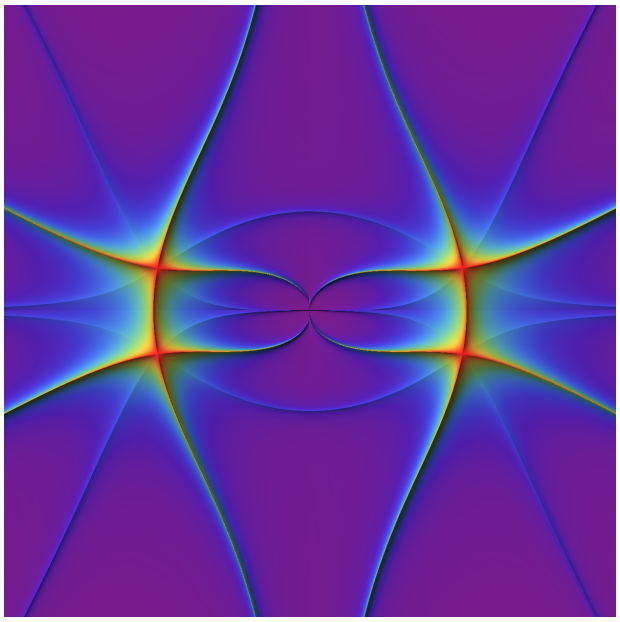


Figure 3.4: For this numerical example we have chosen $k = 1, \Lambda = 3, q_0 = 0, q_1 = 10$. The saddle points then lie at $\pm 3 \pm i$. Shown in the present figure are both the boundaries of wedges (lines of constant real part of the integrand/imaginary part of the action – light blue lines) and the flow lines (lines of constant real part of the action – red/green lines). More specifically, the plot shows both $Abs[\mathrm{Im}(S(N) - S(N_s))]$ and $Abs[\mathrm{Re}(S(N) - S(N_s))]$, where lighter colours correspond to smaller values. The four saddle points are located at the intersections of the flow lines. More details are provided in Fig. 3.5.

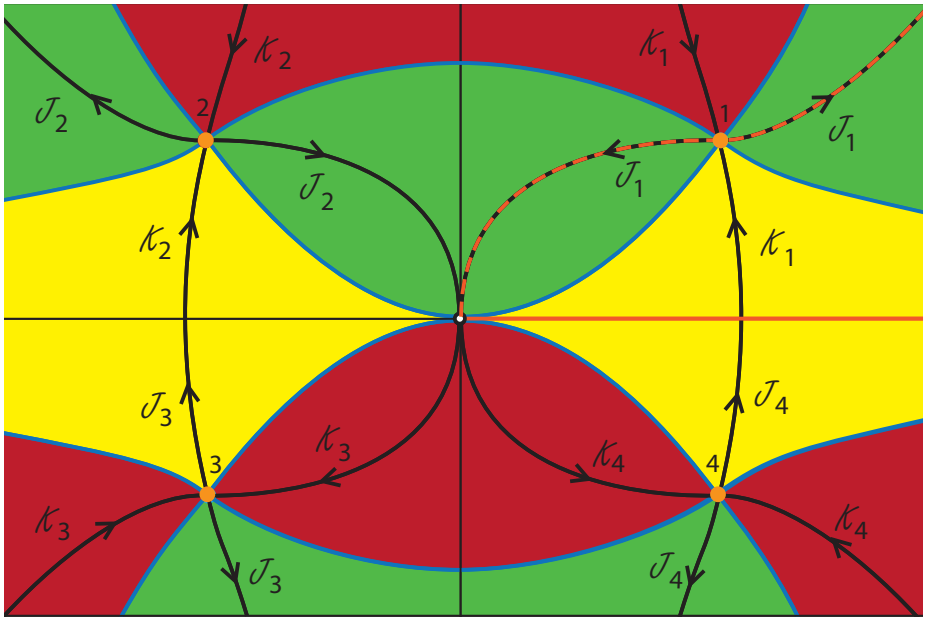


Figure 3.5: A sketch of the wedges and flow lines emanating from the saddle points in the complex N plane, for "no-boundary" conditions $q_0 = 0, q_1 > \frac{3}{\Lambda}$. The loci of the steepest ascent/descent flows (in black) and of the boundaries between wedges (in blue) were determined numerically in Fig. 3.4. Here the arrows indicate the direction of steepest descent. We have coloured the wedges such that regions J_σ with a lower value of the magnitude of the integrand than the corresponding saddle point are green, and regions K_σ with a higher value are red, with the exception of the yellow regions which have a value intermediate between the two saddle point values. Comparing with the adjacent colours then avoids any ambiguity. Notice that, due to the symmetry explained above equation (3.6), there are 'degenerate' ascent and descent flows that link saddle points. This degeneracy is broken by adding an infinitesimal perturbation to the action, as shown in Fig. 3.6. The original integration contour along the positive real axis is shown in orange, and the deformed contour which Picard-Lefschetz theory picks out as the preferred integration cycle is marked in dashed orange. Again neither the flow lines, nor the original or final integration contours, include the point at $N = 0$. Only saddle point 1 in the upper right quadrant can be linked to the original integration contour via an upward flow, and this implies that the (orange-dashed) downward flow from this saddle point is the correct Lefschetz thimble along which the path integral should be performed.

One can identify the direction of the flows analytically by expanding the action around a saddle point,

$$\delta S_0 = S_0^{saddle} + \frac{1}{2} S_{0,NN}^{saddle} (\delta N)^2 + \dots \tag{3.38}$$

The second derivative is given by

$$\begin{aligned} S_{0,NN}^{saddle} &= N_s \frac{\Lambda^2}{6} - \frac{3}{2N_s^3}(q_1 - q_0)^2 & (3.39) \\ &= \frac{1}{6N_s^3}\left[\Lambda^2 N_s^4 - 9(q_1 - q_0)^2\right] . & (3.40) \end{aligned}$$

By evaluating the second derivative, we can determine the direction along which the imaginary part of the action stays constant. Our numerical example serves as an illustration. For the saddle point (number 2) at $N_s = -3 + i$ for instance, the second derivative is given by $S_{,NN} = 9/5 \times (-1 + 3i)$. This means that $\alpha \equiv Arg(S_{,NN}) = \pi - ArcTan(3) \approx 1.89$. We want $Im(iS - iS(N_s)) = 0$. Around the saddle point, the change in iS goes like $\Delta(iS) \propto iS_{,NN}(\delta N)^2 \sim n^2 e^{i(\pi/2 + 2\theta + \alpha)}$, where we have written $\delta N = n e^{i\theta}$. The change in the imaginary part will be proportional to $\sin(\pi/2 + 2\theta + \alpha)$, and this change is zero if

$$\begin{aligned} \theta &= \frac{k\pi}{2} - \frac{\pi}{4} - \frac{\alpha}{2}, \qquad k \in \mathbb{R} & (3.41) \\ &\approx -0.16,\ 1.41,\ 2.98,\ 4.55,\ \dots & (3.42) \end{aligned}$$

This is in good agreement with the flow lines shown in the figure. Note that the change in the real part of the integrand $h = Re(iS)$ is given by $\cos(\pi/2 + 2\theta + \alpha)$. The direction of steepest descent is thus given by $\cos(\pi/2 + 2\theta + \alpha) = -1$, , *i.e.*, for

$$\begin{aligned} \theta &= k\pi + \frac{\pi}{4} - \frac{\alpha}{2}, \quad k \in \mathbb{R}, \quad (\textit{steepest descent}) & (3.43) \\ &\approx -0.16,\ 2.98,\ \dots & (3.44) \end{aligned}$$

while the curves of steepest ascent are at $\theta \approx 1.41,\ 4.55,\ \dots$ Thus the line of steepest descent of h is towards the origin $N = 0$, while the curve of steepest ascent is down towards the real line. This line eventually connects with the saddle point at $-3 - i$. Thus we encounter the degenerate situation described in section 3.2 where the curves of steepest ascent from one saddle point coincides with the curve of steepest descent from another. As discussed there, we can lift this degeneracy by considering a small complex perturbation of the action, and subsequently take the limit where the perturbation vanishes. The effect of such a perturbation is shown in Fig. 3.6, where one can clearly see that the degeneracy is now lifted, and the intersection formula (3.6) can be applied. It is straightforward to repeat

this calculation for the other three saddle points, with the result that the two saddles in the upper half plane have flow lines that are mirror images of each other, while the two saddles in the lower half plane have their upward and downward flow reversed.

Note that the downward flow lines (Lefschetz thimbles) of the upper saddle points can indeed be deformed to the real N line, while the downward flow lines of the lower saddle points cannot. Moreover, only saddle point 1 can be linked to the original integration contour (the positive real half line) via an upward flow, and hence the appropriate integration contour, along which the integral will be manifestly convergent, is given by the Lefschetz thimble $\mathcal{J}_1$ also indicated by the dashed orange line in Fig. 3.5. More precisely, it is implied by the arguments presented around Eq. (3.9) that the integral along the arc at infinity linking the real integration domain to the Lefschetz thimble $\mathcal{J}_1$ vanishes, and thus the path integral manifestly converges. Saddle point 1 lies at

$$N^{+}_{s,nb1} = +\frac{3}{\Lambda}\left[i + (\frac{\Lambda}{3}q_1 - 1)^{1/2}\right] , \tag{3.45}$$

and the action evaluated on the saddle point is

$$S^{+}_{0,nb1} = -\frac{6}{\Lambda}\left[-i + (\frac{\Lambda}{3}q_1 - 1)^{3/2}\right] . \tag{3.46}$$

For saddle points of the form (3.28), we have

$$S_{0,NN} = \frac{2c_2}{N_s}\left(\Lambda q_0 - 3\right)^{1/2}\left(\Lambda q_1 - 3\right)^{1/2} , \tag{3.47}$$

implying that $Arg(N_s) = -\alpha + Arg\left[\left(\Lambda q_0 - 3\right)^{1/2}\left(\Lambda q_1 - 3\right)^{1/2}\right]$. For the "no-boundary" conditions we thus find $Arg(N_s) + \alpha = \frac{\pi}{2}$, and combined with (3.43) this implies $\theta - \frac{1}{2}Arg(N_s) = 0$. In the saddle point approximation, we thus obtain the wavefunction

$$G_{nb}[q_1; 0] \approx e^{i\frac{\pi}{4}}\frac{3^{1/4}}{2(\Lambda q_1 - 3)^{1/4}}e^{-12\pi^2/(\hbar\Lambda) - i4\pi^2\sqrt{\frac{\Lambda}{3}}(q_1 - \frac{3}{\Lambda})^{3/2}/\hbar} . \tag{3.48}$$

Note that the real part of the classical action for the dominant saddle point is negative, as expected from the general arguments presented in section 3.2. This concludes the explicit derivation of our result that the relevant saddle point contributes a weighting $e^{-12\pi^2/(\hbar\Lambda)}$, the inverse of the Hartle-Hawking result.

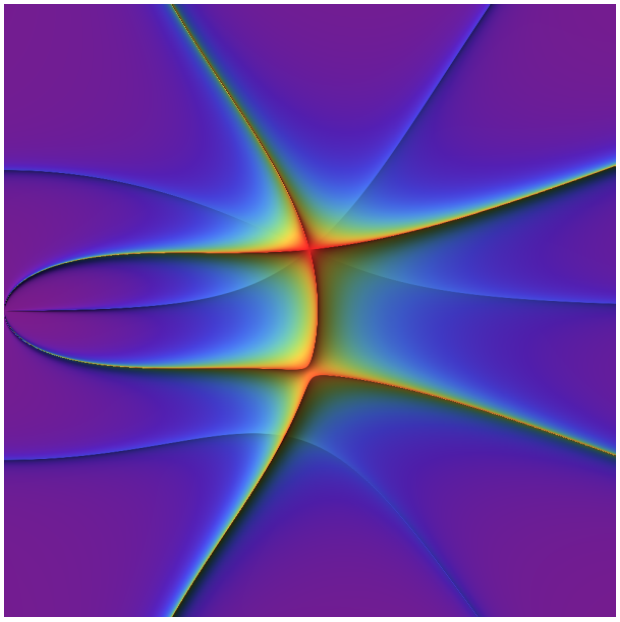


Figure 3.6: Due to the reality of the action, discussed above equation (3.6), and as shown in Fig. 3.5, a curve of steepest ascent from saddle point 1 coincides with a curve of steepest descent from saddle point 4. This degeneracy can be lifted by adding a small complex perturbation to the action, in this example $\Delta S_0 = iN/100$. The new contours of steepest ascent and descent, as well as the level sets of the magnitude of the integrand are shown for saddle point 1. In the presence of the the perturbation, the lower steepest ascent contour from saddle point 1, instead of joining saddle point 4, now runs left to join the origin $N = 0$ from below. Hence the formula for the intersection number (3.6) may now be used to unambiguously determine that saddle point 1 is relevant to the Lorentzian path integral. Note that the two lower curves, one green/red and one blue, do not now pass through a saddle point and should be ignored.

Classicality

The properties of the physical spacetime should be inferred from the quantum mechanical amplitude. In particular, whether or not we are describing a classical spacetime depends on how the amplitude changes as its arguments are varied. Above we have calculated $G_{nb}[q_1]$ as a function of q_1, the scale factor on a spatial hypersurface. The amplitude $G_{nb} = e^{A+iP}$ that we have obtained has a slowly varying amplitude A and a fast-varying phase P as the universe expands, *i.e.*, in the large q_1 limit

$$\frac{\partial A/\partial q_1}{\partial P/\partial q_1} \sim \frac{1}{(q_1 - \frac{3}{\Lambda})^{3/2}} \to 0\,. \tag{3.49}$$

This implies that the amplitude is increasingly classical in a WKB sense as the universe expands. Hence it describes a classical universe. The scaling of the WKB condition for large q_1 is inversely proportion to the volume of space since the spatial volume is proportional to $q_1^{3/2}$. This is what is expected from studies of inflationary "no-boundary" instantons in the limit of an exactly flat potential [236].

Relation to the Euclidean path integral

It is interesting to ask why our results differ from the earlier approaches that took as their starting point the Euclidean path integral. After all, one can simply translate our results into this language by replacing the lapse function N by iN. The graphs we plotted would then simply rotate by 90 degrees. Why would any physical results be changed? The crucial point is that the Euclidean approach assumes the Euclidean time to be fundamental. The path integral should really be performed along the imaginary N axis. In other words, in the Euclidean approach one would take the original integration contour to extend from $N = 0$ (again excluding the point at $N = 0$ itself) to infinity in the positive or negative *imaginary* direction. At this point it is useful to take another look at Fig. 3.5. First note that *none* of the saddle points are related to the imaginary axis by an upward flow line. There are two flow lines that tend towards $N = 0$ asymptotically, but they do not intersect the imaginary axis. This immediately implies that one cannot perform the integral thus defined using the saddle point method. In other words, no combination of saddle points provides a good estimate of the value of the integral. What is more, the integral has no chance of converging. In the positive imaginary direction, the integral diverges at large values of $i|N|$, while along the negative imaginary axis it diverges as it approaches $N = 0$. Another way to say this is to observe that every integration path containing a saddle point passes through a region where the integral is divergent, when one tries to smoothly deform it to

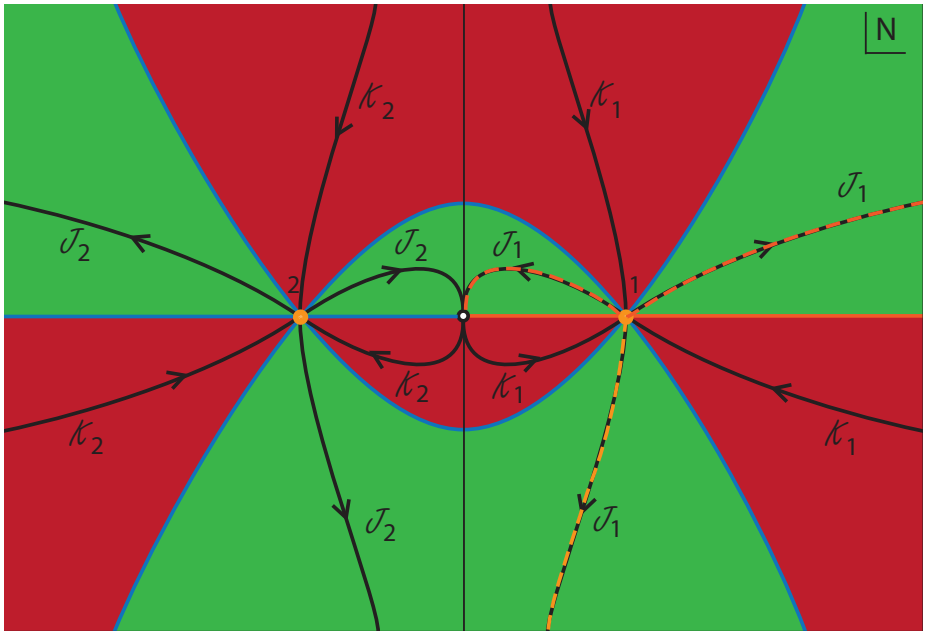


Figure 3.7: A sketch of the wedges and flow lines emanating from the saddle points in the complex N plane, for the boundary conditions $q_1 > q_0 = \frac{3}{\Lambda}$. The colours and arrows are as described in the caption of Fig. 3.3. For these boundary conditions, the saddle points are degenerate, and there are three lines of steepest ascent and descent emanating from them.

the imaginary axis using Cauchy's theorem. Hence we conclude that the Euclidean path integral is simply ill-defined. By contrast, the real time path integral leads to unambiguous and convergent results.

3.3.3 Boundary conditions at the classical limit

The saddle points and flow lines in the case of "no-boundary" conditions look rather different than those obtained with classical boundary conditions. One may wonder how the two descriptions link up as the boundary conditions are continuously varied from classical to non-classical, *i.e.*, from $q_0 > 3/\Lambda$ to $q_0 < 3/\Lambda$, while keeping the final condition classical, $q_1 > 3/\Lambda$. Here we consider limiting case $q_1 > q_0 = \frac{3}{\Lambda}$. The wedges, flow lines and their description are given in Fig. 3.7.

The action iS_0 has two degenerate saddle points at

$$N_{s,limit\pm} = \pm\sqrt{\frac{3}{\Lambda}}\left(q_1 - \frac{3}{\Lambda}\right)^{1/2} . \tag{3.50}$$

The saddle point $N_{s,limit+}$ lies on the original integration contour, and will contribute to the path integral. The other saddle point $N_{s,limit-}$ is irrelevant to the propagator. The saddle points are of order 2, since

$$\frac{\partial S_0}{\partial N} = \frac{\partial^2 S_0}{\partial N^2} = 0 \quad \text{at} \quad N = N_{s,limit\pm} , \tag{3.51}$$

while the third derivative is non-zero. This means that the Taylor expansion around the saddle point is dominated by a cubic term,

$$S_0 = S_0(N_{s,limit+}) + \frac{1}{6}\frac{\partial^3 S_0}{\partial N^3}(N_{s,limit+})(\delta N)^3 + \ldots \tag{3.52}$$

This explains why the flow lines now intersect at angles of $\pi/3$. It is straightforward to evaluate the third derivative,

$$S^{saddle}_{0,NNN} = \frac{\Lambda^2}{6} + \frac{9}{2N_s^4}(q_1 - q_0)^2 \tag{3.53}$$

$$= \frac{6}{N_s^4}\left[\left(q_0 - \frac{3}{\Lambda}\right)^2 + \left(q_1 - \frac{3}{\Lambda}\right)^2\right] . \tag{3.54}$$

In our case $S_{0,NNN}(N_{s,limit+}) = 2\Lambda^2/3$. Given that the third derivative is real-valued and positive, we can determine the directions of the upward and downward flow, as indicated in Fig. 3.7.

This helps us to understand the contour of integration. Starting from $N = 0$, the contour follows the Lefschetz thimble $\mathcal{J}_1$ and moves in the positive imaginary direction, and passes through the saddle point with positive real part. It then runs down, asymptotically towards negative imaginary values, and subsequently comes back along the same path, crossing the degenerate saddle point once more before shooting off at an angle of $\pi/6$ with respect to the real N axis. The middle part of the contour sums to nothing. The Lefschetz thimble transitions to being located entirely in the upper half plane when the boundary condition on q_0 becomes non-classical, cf. Fig. 3.5.

Since for the boundary condition $q_0 = \frac{3}{\Lambda}$ the function h is not a Morse function, *i.e.*, h has two degenerate critical points, the saddle point approximation of equation (3.31) does

not apply. However, using the integral $\int_0^{+\infty} \mathrm{d}n e^{-k\,n^3} = \Gamma(\frac{4}{3})k^{-1/3}$, we can approximate

$$\begin{aligned}\int \frac{\mathrm{d}N}{N^{1/2}} e^{ikN^3} &\approx \frac{1}{N_s^{1/2}} \int \mathrm{d}[\delta N] e^{ik(\delta N)^3} \\ &\approx \frac{1}{N_s^{1/2}} \left[e^{-i\frac{\pi}{6}} \int_{-\infty}^{0} \mathrm{d}n e^{+kn^3} + e^{+i\frac{\pi}{6}} \int_0^{\infty} e^{-kn^3} \right] \\ &\approx \frac{\sqrt{3}\Gamma(\frac{4}{3})}{N_s^{1/2} k^{1/3}}\,, \end{aligned} \tag{3.55}$$

in order to obtain the saddle point approximation of the propagator

$$G[q_1; q_0 = 3/\Lambda] \approx \frac{e^{i\frac{\pi}{4}} 3^{17/12} \Gamma(\frac{4}{3})}{2^{5/6}\pi^{1/6}(\hbar\Lambda)^{1/6}(\Lambda q_1 - 3)^{1/4}}\, e^{-i4\pi^2\sqrt{\frac{\Lambda}{3}}(q_1 - \frac{3}{\Lambda})^{3/2}/\hbar}\,. \tag{3.56}$$

The prefactors $e^{\pm i\frac{\pi}{6}}$ arise from writing $\delta N = n e^{\pm i\frac{\pi}{6}}$ so that the respective integrals are performed along the Lefschetz thimbles.

3.3.4 Non-classical boundary conditions

Finally, we consider boundary conditions that are classically impossible, $\frac{3}{\Lambda} > q_1 \geq q_0$, where both scale factors are smaller then the waist of de Sitter space. Even though such configurations are impossible in Lorentzian signature, they exist in Euclidean signature and correspond to sections of a 4-dimensional sphere (*e.g.* one may picture them as surfaces of constant latitude). Correspondingly, the saddle points are pure imaginary. In the upper half plane they are

$$N_{upper\pm} = i\sqrt{\frac{3}{\Lambda}} \left[\left(\frac{3}{\Lambda} - q_1 \right)^{1/2} \pm \left(\frac{3}{\Lambda} - q_0 \right)^{1/2} \right], \tag{3.57}$$

while in the lower half plane

$$N_{lower\pm} = -i\sqrt{\frac{3}{\Lambda}} \left[\left(\frac{3}{\Lambda} - q_1 \right)^{1/2} \pm \left(\frac{3}{\Lambda} - q_0 \right)^{1/2} \right]. \tag{3.58}$$

There are two possibilities, as for the case of classical boundary conditions. The two spatial hypersurfaces are either on one side of the equator of the 4-sphere, or are separated by

the equator. We can determine which is which by looking at the derivative of the classical solution

$$\frac{\mathrm{d}\bar{q}}{\mathrm{d}t} = \frac{2\Lambda}{3}N_s^2 t - \frac{\Lambda}{3}N_s^2 + q_1 - q_0 = 0, \quad 0 < t < 1\,. \tag{3.59}$$

It is straightforward to see that for N_{upper+} and N_{lower+} the scale factor squared q reaches a maximum for $0 < t < 1$, while for N_{upper-} and N_{lower-} the maximum is only reached for $t > 1$, which is outside of the range for which the solution has been determined. Thus the saddle points at N_{upper+} and N_{lower+} correspond to the configuration where the initial and final hypersurface lie on different sides of the equator of the sphere.

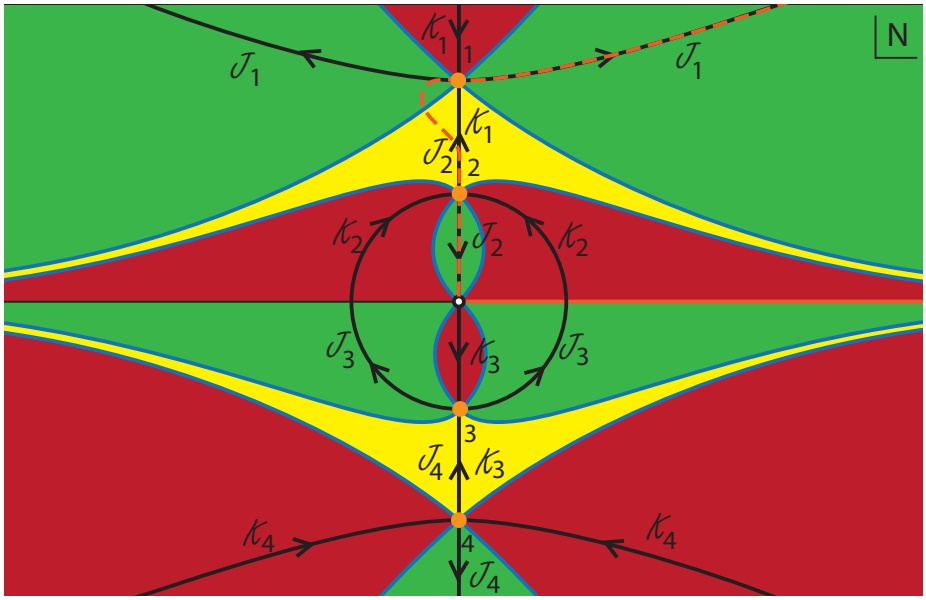


Figure 3.8: A sketch of the wedges and flow lines emanating from the saddle points in the complex N plane, for non-classical boundary conditions $\frac{3}{\Lambda} > q_1 \geq q_0$. The lines, arrows and colours are as described in the caption of Fig. 3.5. Here we again have yellow wedges that are both higher than one saddle point and lower than another one, $J_2 = K_1$ and $J_4 = K_3$. The original integration contour can be deformed into the dashed preferred contour by flowing down. Note that the preferred, dashed contour is chosen to follow the paths of steepest descent near the saddle points, so as to facilitate a saddle point approximation of the propagator.

An analysis of the Lefschetz thimbles indicates that only the saddle points in the upper half plane contribute to the propagator, as these are related to the original integration contour via an upward flow (in the case of N_{upper+} the flow proceeds through the saddle

point N_{upper-} this degeneracy can once again be removed by adding a small complex perturbation to the action, and letting it vanish at the end of the calculation). The proper integration contour runs up from $N = 0$ in the positive imaginary direction, passes through the two upper saddle points (along paths of steepest descent) and then follows the Lefschetz thimble in the upper right quadrant off to infinity at an asymptotic angle of $\pi/6$ with respect to the real N axis – see Fig. 3.8. The propagator can thus be approximated as

$$G[q_1; q_0] \approx c_- e^{-\frac{12\pi^2}{\hbar\Lambda}\left[\left(1-\frac{\Lambda q_0}{3}\right)^{3/2}+\left(1-\frac{\Lambda q_1}{3}\right)^{3/2}\right]} + c_+ e^{-\frac{12\pi^2}{\hbar\Lambda}\left[\left(1-\frac{\Lambda q_0}{3}\right)^{3/2}-\left(1-\frac{\Lambda q_1}{3}\right)^{3/2}\right]}. \qquad (3.60)$$

Keeping in mind that for the saddle point N_{upper-} we have the angle along the Lefschetz thimble being $\theta = \frac{\pi}{2}$ and thus $e^{i\left[\theta-\frac{1}{2}Arg(N)\right]} = e^{i\frac{\pi}{4}}$, while for N_{upper+} we have $\theta = 0$ and $e^{i[\theta-\frac{1}{2}Arg(N)]} = e^{-i\frac{\pi}{4}}$, the normalization constants are

$$c_\pm \;=\; e^{\mp i\frac{\pi}{4}}\left(\frac{3i}{4\Lambda\sqrt{\left(\frac{3}{\Lambda}-q_0\right)\left(\frac{3}{\Lambda}-q_1\right)}}\right)^{1/2}. \qquad (3.61)$$

Note that the exponents in the amplitude are purely real, so it most certainly does not describe classical cosmological evolution. The implied interference between the two terms in the wavefunction may lead to interesting effects, whose investigation we leave to future work. Here we simply note that the saddle point at N_{upper-} dominates, *i.e.*, the path integral gives a higher weighting to the smaller 4-geometry connecting initial and final boundary values, and not the geometry for which the equator of the 4-sphere sits in between. In particular, in the limit where the two boundaries become equal the preference is for having a vanishing 4-geometry, a physically reasonable result.

3.4 Relation to the Wheeler-DeWitt equation

So far we have focussed on path integral quantization. It is natural to wonder how our results are related to canonical methods. In the canonical approach, one obtains a time independent Schrödinger-type equation on superspace, *i.e.*, on the space of 3-dimensional spatial geometries. This is the Wheeler-DeWitt equation. Below we will derive this equation in two separate ways, first from the canonical approach, and afterwards by starting from the path integral. Both approaches lead to the same equation, as they must. The advantage of the path integral approach, however, is that it incorporates the correct boundary conditions and leaves no residual ambiguity in the solutions of the Wheeler-DeWitt equation.

3.4.1 Canonical derivation of the Wheeler-DeWitt equation

The minisuperspace model is described by the action (3.18), with the corresponding Lagrangian

$$L = 2\pi^2 \left[-\frac{3}{4N}\dot{q}^2 + 3kN - N\Lambda q \right]. \tag{3.62}$$

The canonical momentum corresponding to q is given by

$$p = \frac{\partial L}{\partial \dot{q}} = -\frac{3\pi^2}{N}\dot{q}\,. \tag{3.63}$$

The classical Hamiltonian of the system takes the simple form

$$H = \dot{q}p - L = -\frac{N}{6\pi^2}\left[p^2 + 12\pi^4(3k - \Lambda q)\right] = N\mathcal{H}\,. \tag{3.64}$$

The phase-space representation of the action reads

$$S = \int (\dot{q}p - N\mathcal{H})\,\mathrm{d}t = \int \left(\dot{q}p + \frac{N}{6\pi^2}\left[p^2 + 12\pi^4(3k - \Lambda q)\right] \right) \mathrm{d}t\,. \tag{3.65}$$

Observe that the lapse N is a Lagrange multiplier leading to the classical constraint

$$\mathcal{H} = 0\,. \tag{3.66}$$

In the canonical quantization scheme we obtain the Hamiltonian operator $\hat{H}$ in the q-representation by the substitution $p \mapsto \hat{p} = -i\frac{\partial}{\partial q}$,

$$\hat{H} = \frac{1}{6\pi^2}\left[\frac{\partial^2}{\partial q^2} + 12\pi^4(\Lambda q - 3k) \right]. \tag{3.67}$$

The corresponding Wheeler-DeWitt equation is given by

$$\hat{H}\psi = 0 \to \hbar^2 \frac{\partial^2 \psi}{\partial q^2} + 12\pi^4(\Lambda q - 3k)\psi = 0\,, \tag{3.68}$$

with ψ the wave-function of the universe. The corresponding Feynman propagator G satisfies [184]

$$\hat{H}G = -i\delta(q_0 - q_1)\,, \tag{3.69}$$

where the Hamiltonian operator acts either on q_0 or on q_1. We call this equation the inhomogeneous Wheeler-DeWitt equation. For a the De Sitter universe,

$$\hbar^2 \frac{\partial^2 G}{\partial q_1^2} + 12\pi^4(\Lambda q_1 - 3k)G = -6\pi^2 i\delta(q_0 - q_1)\,. \tag{3.70}$$

In the next section we prove that the path integral is propagator satisfying this equation.

3.4.2 Integral representation of the Feynman propagator

Our discussion in section 3.3 has shown that the propagator/wave function of the universe is given by

$$G[q_1; q_0] = \int_0^\infty \mathrm{d}N G[q_1; q_0; N]\,. \tag{3.71}$$

The integrand

$$G[q_1; q_0; N] = \int_{q=q_0}^{q=q_1} \mathcal{D}q e^{iS(N,q)/\hbar} \tag{3.72}$$

is the propagator to propagate from q_0 to q_1 in parameter time N. In analogy with non-relativistic quantum mechanics and by construction of the path integral, the propagator will satisfy a Schrödinger-like equation

$$i\frac{\partial G[q_1; q_0; N]}{\partial N} = \hat{H} G[q_1; q_0; N]\,, \tag{3.73}$$

with the Hamiltonian operator acting on either q_0 or q_1 with the boundary condition

$$\lim_{N\to 0} G[q_1; q_0; N] = \delta(q_0 - q_1)\,. \tag{3.74}$$

It follows that the total propagator of minisuperspace models is a Greens function of the Wheeler-DeWitt equation

$$\begin{aligned}
\hat{H} G[q_1; q_0] =& \int_0^\infty \mathrm{d}N \hat{H} G[q_1; q_0; N] \\
=& i\int_0^\infty \mathrm{d}N \frac{\partial G[q_1; q_0; N]}{\partial N} \\
=& iG[q_1; q_0; N]\big|_{N=0}^{N=\infty} \\
=& -i\delta(q_0 - q_1)\,.
\end{aligned} \tag{3.75}$$

3.4.3 Solution of the Feynman propagator

The Wheeler-DeWitt equation is solved by bilinear expressions in Airy functions,

$$Ai\left[\frac{\sqrt[3]{-12\pi^4}(q\Lambda - 3k)}{\Lambda^{2/3}}\right], \quad Bi\left[\frac{\sqrt[3]{-12\pi^4}(q\Lambda - 3k)}{\Lambda^{2/3}}\right], \tag{3.76}$$

where q stands for either q_0 or q_1 here. Note that in the arguments above, $\sqrt[3]{-1}$ can stand for $e^{i\frac{\pi}{3}}, -1, e^{-i\frac{\pi}{3}}$ – either choice, combined with a suitable linear combination of Airy functions, solve the Wheeler-DeWitt equation [4]. For definiteness we will choose $\sqrt[3]{-1} = e^{i\frac{\pi}{3}}$, and write

$$z \equiv \frac{e^{i\frac{\pi}{3}}(12\pi^4)^{1/3}(\Lambda q - 3)}{(\hbar\Lambda)^{2/3}}\,. \tag{3.77}$$

We also use the notation $z_0 = z(q \to q_0)$ and $z_1 = z(q \to q_1)$.

In the previous discussion, we approximated the Feynman propagator using the WKB approximation. Now, using the homogeneous solutions of the Wheeler-DeWitt equation we can construct the space of all possible Green's functions and subsequently determine the Feynman propagator. The propagators of the Wheeler-DeWitt equations must be of the form

$$G[q_1; q_0] = -6\pi^2 i \frac{\psi_1(q_0)\psi_2(q_1)\theta(q_0 - q_1) + \psi_1(q_1)\psi_2(q_0)\theta(q_1 - q_0)}{\psi_1(q_0)\psi_2'(q_1) - \psi_1(q_1)\psi_2'(q_0)}, \tag{3.78}$$

with ψ_1 and ψ_2 two linearly independent homogeneous solutions. Explicitly, if we take $\psi_1 = Ai[z] + aBi[z]$ and $\psi_2 = Ai[z] + bBi[z]$ then we obtain the general form of the propagator

$$G[q_1; q_0] = \frac{i}{(b-a)} \frac{2^{1/3}(-3)^{2/3}\pi^{5/3}}{(\Lambda\hbar)^{1/3}} \left[\psi_1(q_0)\psi_2(q_1)\theta(q_0 - q_1) + \psi_1(q_1)\psi_2(q_0)\theta(q_1 - q_0)\right]. \tag{3.79}$$

We can try to find the appropriate linear combination of solutions by solving for the path integral directly near $z_0 = z_1 = 0$, , *i.e.*, near $q_0 = q_1 = \frac{\Lambda}{3}$ [186]. Re-scaling the lapse function by $\Lambda^{2/3}$ and writing $z_0 = z_1 = Z$, the path integral is given by

$$G[Z, Z] = \sqrt{\frac{3\pi i}{2}} \frac{1}{\Lambda^{1/3}} \int \frac{\mathrm{d}N}{N^{1/2}} e^{\frac{i\pi^2}{18}N^3 - \gamma N Z} \tag{3.80}$$

where $\gamma = \frac{2\pi^2 i}{(-12\pi^4)^{1/3}}$. In order to find the Taylor expansion near $Z = 0$, we must evaluate the path integral and its first few derivatives at $Z = 0$. Writing $N = ne^{i\frac{\pi}{6}}$ in order to

[4] This is due to the fact that Airy functions contain three regions of convergence in the complex plane of the argument, where these regions are invariant under rotations by $2\pi/3$ radians. The Airy functions Ai and Bi are defined as two linearly independent combinations of the possible convergent contours. Rotating the arguments by $2\pi/3$ then translates into taking different linear combinations of the convergent integration contours, *i.e.*, simply corresponds to taking different linear combinations of Airy functions [342]. An example is provided by the relation $Ai(ze^{\pm i\frac{2\pi}{3}}) = \frac{1}{2}e^{\pm i\frac{\pi}{3}}[Ai(z) \mp iBi(z)]$.

match to the appropriate Lefschetz thimble, we find

$$\begin{aligned}
G\left[\frac{\Lambda}{3},\frac{\Lambda}{3}\right] &= \sqrt{\frac{3\pi i}{2}}\frac{1}{\Lambda^{1/3}}e^{i\frac{\pi}{12}}\int_0^\infty \frac{\mathrm{d}n}{n^{1/2}}e^{\frac{-\pi^2}{18}n^3} = \sqrt{\frac{3\pi i}{2}}\frac{1}{\Lambda^{1/3}}\frac{(2i)^{1/6}\Gamma(\frac{1}{6})}{(9\pi)^{1/3}} \qquad (3.81)\\
\frac{\partial G}{\partial Z}\left[\frac{\Lambda}{3},\frac{\Lambda}{3}\right] &= \sqrt{\frac{3\pi i}{2}}\frac{(-\gamma)}{\Lambda^{1/3}}e^{i\frac{\pi}{4}}\int_0^\infty \mathrm{d}n\, n^{1/2}e^{\frac{-\pi^2}{18}n^3} = \sqrt{\frac{3\pi i}{2}}\frac{(-\gamma)}{\Lambda^{1/3}}e^{i\frac{\pi}{4}}\sqrt{\frac{2}{\pi}} \qquad (3.82)\\
\frac{\partial^2 G}{\partial Z^2}\left[\frac{\Lambda}{3},\frac{\Lambda}{3}\right] &= \sqrt{\frac{3\pi i}{2}}\frac{\gamma^2}{\Lambda^{1/3}}e^{i\frac{5\pi}{12}}\int_0^\infty \mathrm{d}n\, n^{3/2}e^{\frac{-\pi^2}{18}n^3} = \sqrt{\frac{3\pi i}{2}}\frac{\gamma^2}{\Lambda^{1/3}}e^{i\frac{5\pi}{12}}\frac{2^{5/6}3^{2/3}\Gamma(\frac{5}{6})}{\pi^{5/3}}
\end{aligned}$$

The Taylor series $G[Z,Z] = G\left[\frac{\Lambda}{3},\frac{\Lambda}{3}\right] + G_{,Z}\left[\frac{\Lambda}{3},\frac{\Lambda}{3}\right]Z + \frac{1}{2}G_{,ZZ}\left[\frac{\Lambda}{3},\frac{\Lambda}{3}\right]Z^2$ precisely matches that of

$$G[Z,Z] = \frac{2^{1/3}(-3)^{2/3}\pi^{5/3}}{(\hbar\Lambda)^{1/3}}Ai[Z]\left(Ai[Z] - iBi[Z]\right)\,. \qquad (3.83)$$

The asymptotic limits at large and small q, to which we will turn shortly, then imply that we should take

$$\begin{aligned}
G[q_1;q_0] = \frac{2^{1/3}(-3)^{2/3}\pi^{5/3}}{(\hbar\Lambda)^{1/3}}\Big[&\left(Ai[z_0] - iBi[z_0]\right)Ai[z_1]\theta(q_1-q_0)\\
&+ \left(Ai[z_1] - iBi[z_1]\right)Ai[z_0]\theta(q_0-q_1)\Big]\,. \qquad (3.84)
\end{aligned}$$

Note that $a = -i$ and $b = 0$, by which we see that (3.84) is a propagator. This is equal the evaluation of a non-relativistic particle with a linear potential [174].

For the "no-boundary" case, we are interested in small (or zero) q_0 and large q_1. For $\Lambda q_0 \ll 3$ we have the approximate formulae

$$\begin{aligned}
Ai\left[e^{i\frac{\pi}{3}}\frac{(12\pi^4)^{1/3}}{\Lambda^{2/3}}(\Lambda q_0 - 3)\right] &\approx \frac{(\hbar\Lambda)^{1/6}e^{-i\frac{\pi}{12}}}{\sqrt{\pi}(12\pi^4)^{1/12}(3-\Lambda q_0)^{1/4}}\sin\left(i\frac{4\pi^2}{\sqrt{3}\hbar\Lambda}(3-\Lambda q_0)^{\frac{3}{2}} + \frac{\pi}{4}\right)\\
Bi\left[e^{i\frac{\pi}{3}}\frac{(12\pi^4)^{1/3}}{\Lambda^{2/3}}(\Lambda q_0 - 3)\right] &\approx \frac{(\hbar\Lambda)^{1/6}e^{-i\frac{\pi}{12}}}{\sqrt{\pi}(12\pi^4)^{1/12}(3-\Lambda q_0)^{1/4}}\cos\left(i\frac{4\pi^2}{\sqrt{3}\hbar\Lambda}(3-\Lambda q_0)^{\frac{3}{2}} + \frac{\pi}{4}\right)
\end{aligned}$$

while for $\Lambda q_1 \gg 3$ we have

$$Ai\left[e^{i\frac{\pi}{3}}\frac{(12\pi^4)^{1/3}}{\Lambda^{2/3}}(\Lambda q_1 - 3)\right] \approx \frac{(\hbar\Lambda)^{1/6}e^{-i\frac{\pi}{12}}}{2\pi^{1/2}(12\pi^4)^{1/12}(\Lambda q_1 - 3)^{1/4}}e^{-i\frac{\sqrt{12}\pi^2}{\hbar\Lambda}(\Lambda q_1 - 3)^{3/2}}\,.$$

For the total propagator we obtain

$$\begin{aligned}
G[q_1;q_0=0] &= \frac{2^{1/3}(-3)^{2/3}\pi^{5/3}}{(\hbar\Lambda)^{1/3}}\left(Ai\left[-3\frac{e^{i\frac{\pi}{3}}(12\pi^4)^{1/3}}{(\hbar\Lambda)^{2/3}}\right] - iBi\left[-3\frac{e^{i\frac{\pi}{3}}(12\pi^4)^{1/3}}{(\hbar\Lambda)^{2/3}}\right]\right)Ai[z_1]\\
&\approx \frac{e^{i\frac{\pi}{4}}3^{1/4}}{2(\Lambda q_1 - 3)^{1/4}}e^{-\frac{12\pi^2}{\hbar\Lambda} - i\frac{\sqrt{12}\pi^2}{\hbar\Lambda}(\Lambda q_1 - 3)^{3/2}}\,. \qquad (3.85)
\end{aligned}$$

This agrees exactly with (3.48), including the subleading terms.

A similar agreement can be found for classical and non-classical boundary conditions, by evaluating the appropriate limits of the Airy functions. In both cases the expressions agree with Eq. (3.34) respectively Eq. (3.60), including the first sub-leading term.

With boundary conditions at the classical limit, $q_0 = \frac{3}{\Lambda}$ and $\Lambda q_1 \gg 3$, the propagator is given by

$$G\left[q_1; q_0 = \frac{3}{\Lambda}\right] = \frac{2^{1/3}(-3)^{2/3}\pi^{5/3}}{(\hbar\Lambda)^{1/3}} \left(Ai[0] - iBi[0]\right) Ai[z_1]\,. \tag{3.86}$$

Using the exact expression $Ai[0] - iBi[0] = e^{i\frac{2\pi}{3}} \frac{2\ 3^{1/3}}{\Gamma(-\frac{1}{3})} = e^{-i\frac{\pi}{3}} \frac{3^{5/6}\Gamma(\frac{4}{3})}{\pi}$, one may verify that the asymptotic form of this propagator also agrees with the path integral expression (3.56).

3.5 Discussion

We hope this calculation has brought a new element of rigor into quantum cosmology. We have argued that the Lorentzian path integral, combined with Picard-Lefschetz theory, is to be preferred over the Euclidean version. In particular, in the simplest cosmology – a closed FRW universe, with a positive cosmological constant – we explained how it eliminates the ambiguities associated with the Euclidean path integral, including the conformal factor problem and the question of which saddle points are relevant. We have shown that Picard-Lefschetz theory identifies precisely which saddle point solutions contribute to the Feynman propagator, with which factors, and hence how a consistent semiclassical expansion may be developed. We have also shown how the path integral formulation of the causal propagator eliminates the problem of defining boundary conditions on superspace, an ambiguity which plagues attempts to obtain the "wavefunction of the universe" by solving the homogeneous Wheeler-DeWitt equation.

As we have seen, for "no-boundary" conditions the Feynman propagator includes a semiclassical factor $e^{-12\pi^2/(\hbar\Lambda)}$, arising from the classical action of the relevant saddle point solution. We gave a general argument in the introduction, detailed in section (3.2), that relevant complex classical solutions only give suppression factors, and never enhancement factors such as are obtained from Hartle and Hawking's Euclidean approach. Furthermore, we explained in detail why the Euclidean path integral is divergent and hence cannot be taken to be a fundamental starting point of the theory.

We showed in simple minisuperspace examples how the Lorentzian path integral reduces to a perfectly convergent (albeit conditionally convergent) integral over the space

of fields. It is very plausible that this result extends to include all dynamical modes. In particular, the fluctuations about our homogeneous, isotropic but sometimes complex saddle point solutions, will themselves possess quadratic, complex actions. It is clear that Picard-Lefschetz theory applies rather trivially to this case and *always* yields a convergent measure on the space of field fluctuations. Furthermore, field interactions lead to higher powers in the action and hence better and better convergence when the Picard-Lefschetz approach is employed. It is also clear how unitarity is recovered for cosmological backgrounds corresponding to real (or nearly real) saddle point solutions, because the starting point of the whole theory is a path integral over real fields. We believe these arguments, as well as the examples we have investigated in detail, provide compelling evidence that the Lorentzian formulation of quantum cosmology is to be preferred.

As we have already mentioned, there is significant overlap between this work and that of [163, 164] which describes a quantum cosmological bounce for conformal-invariant matter and free scalar fields, also in a Lorentzian formulation. In that context, two of us have recently shown how cosmological time can emerge [137].

An obvious extension will be the inclusion of other types of matter, such as pressure-free matter and scalar fields with nontrivial potentials. In particular it will be interesting to revisit previous uses of instantons, in both inflationary [189] and ekpyrotic [32, 31] cosmologies, as well as attempts to describe quantum transitions between contraction and expansion in such models [56], in the light of Picard-Lefschetz theory. In this context, exactly solvable models such as those of [156, 28] may provide useful insight. As we have seen, for the simplest cosmology with only a cosmological constant we obtained the same result as Vilenkin, in his "tunneling" proposal for the wavefunction of the universe. In chapter 4, we apply the same technique to a more realistic model, where we consider not only the background spacetime but include fluctuations.

More generally, it will be interesting to see how tunneling, and other nonperturbative quantum gravity processes, can be treated in this Lorentzian-Picard-Lefschetz (LPL) framework.

3.A The Feynman propagator and the Wheeler-DeWitt equation in a de Sitter universe

In section 3.3 we showed that the Feynman propagator for a De Sitter minisuperspace model is given by

$$G[q_1;q_0] = \sqrt{\frac{3\pi i}{2}} \int_0^\infty \frac{\mathrm{d}N}{N^{1/2}} e^{2\pi^2 i S_0}\,, \tag{3.87}$$

with

$$S_0 = N^3\,\frac{\Lambda^2}{36} + N\left(-\frac{\Lambda}{2}(q_0+q_1)+3k\right) + \frac{1}{N}\left(-\frac{3}{4}(q_1-q_0)^2\right)\,, \tag{3.88}$$

in our minisuperspace model of gravity with a positive cosmological constant Λ. Starting with the propagator, we can derive the Wheeler-deWitt equation by taking derivatives. The partial derivative of G with respect to q_1 is

$$\begin{aligned}\frac{\partial G}{\partial q_1} =& \sqrt{\frac{3\pi i}{2}} \int \frac{\mathrm{d}N}{N^{1/2}} 2\pi^2 i S_{0,q_1} e^{2\pi^2 i S_0} \\ =& \sqrt{\frac{3\pi i}{2}} \int \frac{\mathrm{d}N}{N^{1/2}} 2\pi^2 i \left[-\frac{N}{2}\Lambda - \frac{3}{2N}(q_1-q_0)\right] e^{2\pi^2 i S_0}.\end{aligned} \tag{3.89}$$

The second order partial derivative of G with respect to q_1 is

$$\begin{aligned}\frac{\partial^2 G}{\partial q_1^2} =& \sqrt{\frac{3\pi i}{2}} \int \frac{\mathrm{d}N}{N^{1/2}} \left[2\pi^2 i S_{0,q_1q_1} - 4\pi^4 S_{0,q_1}^2\right] e^{2\pi^2 i S_0} \\ =& \sqrt{\frac{3\pi i}{2}} \int \frac{\mathrm{d}N}{N^{1/2}} \left[-4\pi^4\left(\frac{N}{2}\Lambda + \frac{3}{2N}(q_1-q_0)\right)^2 - \frac{3\pi^2 i}{N}\right] e^{2\pi^2 i S_0}\,.\end{aligned} \tag{3.90}$$

The argument of the integral depends on N. We would like to remove this dependence by using the properties of the Lefschetz thimbles. From the fundamental theorem of calculus and partial integration we have that

$$\left[N^{-\frac{1}{2}} e^{2\pi^2 i S_0}\right]_0^\infty = \int \mathrm{d}N \frac{\mathrm{d}}{\mathrm{d}N}\left[N^{-\frac{1}{2}} e^{2\pi^2 i S_0}\right] = -\frac{1}{2}\int \frac{\mathrm{d}N}{N^{\frac{3}{2}}} e^{2\pi^2 i S_0} + 2\pi^2 i \int \frac{\mathrm{d}N}{N^{\frac{1}{2}}} S_{0,N} e^{2\pi^2 i S_0}\,. \tag{3.91}$$

Substituting this relation in the second order partial derivative of G with respect to q_1 gives

$$\begin{aligned}\frac{\partial^2 G}{\partial q_1^2} =&\sqrt{\frac{3\pi i}{2}}\left[\int \frac{\mathrm{d}N}{N^{1/2}}\left[-4\pi^4\left(\frac{N}{2}\Lambda+\frac{3}{2N}(q_1-q_0)\right)^2\right]e^{2\pi^2 iS_0}-3\pi^2 i\int\frac{\mathrm{d}N}{N^{\frac{3}{2}}}e^{2\pi^2 iS_0}\right]\\ =&\sqrt{\frac{3\pi i}{2}}\left[\int\frac{\mathrm{d}N}{N^{1/2}}\left[-12\pi^4\left(\Lambda q_1-3k\right)\right]e^{2\pi^2 iS_0}+6\pi^2 i\left[N^{-\frac{1}{2}}e^{2\pi^2 iS_0}\right]_0^\infty\right]\\ =&-12\pi^4\left(\Lambda q_1-3k\right)G+6\pi^2 i\sqrt{\frac{3\pi i}{2}}\left[N^{-\frac{1}{2}}e^{2\pi^2 iS_0}\right]_0^\infty. \end{aligned}\tag{3.92}$$

Hence

$$\frac{\partial^2 G}{\partial q_1^2}+12\pi^4\left(\Lambda q_1-3k\right)G=6\pi^2 i\sqrt{\frac{3\pi i}{2}}\left[N^{-\frac{1}{2}}e^{2\pi^2 iS_0}\right]_0^\infty.\tag{3.93}$$

The contribution corresponding to the limit $N\to\infty$ vanishes, since the Lefschetz thimble is constructed such that $e^{2\pi^2 iS_0}\to 0$ and $\frac{1}{\sqrt{N}}$ certainly becomes small in this limit. The propagator G thus satisfies

$$\frac{\partial^2 G}{\partial q_1^2}+12\pi^4\left(\Lambda q_1-3k\right)G=-6\pi^2 i\sqrt{\frac{3\pi i}{2}}\lim_{N\to 0}\frac{e^{2\pi^2 iS_0}}{\sqrt{N}}.\tag{3.94}$$

In the limit $N\to 0$ the action diverges as

$$S_0\to\frac{1}{N}\left(-\frac{3}{4}(q_1-q_0)^2\right).\tag{3.95}$$

Writing $N=in$, since the Lefschetz thimbles approach the origin along the imaginary axis,

$$\lim_{N\to 0}\frac{e^{2\pi^2 iS_0}}{\sqrt{N}}=\sqrt{\frac{2\pi}{i}}\lim_{n\to 0}\frac{e^{-3\pi^2\frac{(q_1-q_0)^2}{2n}}}{\sqrt{2\pi n}}=\sqrt{\frac{2}{3\pi i}}\delta(q_0-q_1).\tag{3.96}$$

So, reinstating $\hbar$, the Wheeler-deWitt (propagator) equation is given by

$$\hbar^2\frac{\partial^2 G}{\partial q_1^2}+12\pi^4\left(\Lambda q_1-3k\right)G=-6\pi^2 i\delta(q_0-q_1).\tag{3.97}$$

Note that, had we integrated over a contour from $N=-\infty$ to $N=+\infty$ (ignoring the singularity at $N=0$) the Dirac delta function term on the right hand side would have been absent.

3.B Homology

Picard-Lefschetz theory of the h-function can be formalized using relative homology. Homology can generally be used to study topological structure of a topological spaces X. We here shortly describe smooth singular homology $H_n(X;\mathbb{Z})$ and subsequently the corresponding relative homology $H_n(X,Y;\mathbb{Z})$. For a more detailed exposition see [286, 246, 190].

3.B.1 Smooth singular homology

Smooth singular homology is based on simplexes defined by

$$\Delta^n = \left\{ x = (x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \middle| \sum_{i=0}^{n} x_i = 1 \text{ and } x_i > 0 \text{ for all } i \right\}. \tag{3.98}$$

In particular Δ^n is known as a n-symplex. Simplexes are a generalization of triangles: Δ^0 is a point, Δ^1 is an line element, Δ^2 is a triangle, Δ^3 is an tetrahedron. The p-symplex is trivially embedded in a corresponding hyperplane

$$\Delta^p \subset V^n \left\{ x = (x_0, \ldots, x_n) \in \mathbb{R}^{n+1} \middle| \sum_{i=0}^{n} x_i = 1 \right\}. \tag{3.99}$$

A smooth singular p-symplex is a continuous map $\sigma : \Delta^n \to X$ that smoothly extends to an open neighborhood in the hyperplane V^n. One should picture this as a smooth embedding of the n-symplex in the topological space.

A p-symplex has $p+1$ vertices, labeled by $e_0, e_1, \ldots, e_n$. We can represent a smooth singular n-symplex by its vertices as

$$[p_0, \ldots, p_n]\,. \tag{3.100}$$

We define a boundary operator ∂_n on the smooth singular n-symplex, mapping the symplex its boundary,

$$\partial_n \sigma = \sum_{k=0}^{n} (-1)^k [p_0, \ldots, p_{k-1}, p_{k+1}, \ldots, p_n]\,. \tag{3.101}$$

The minus sign denotes the orientation of the boundary elements. In particular, the boundary operator maps the line to its extremal points, and the triangle to its three edges.

Given the smooth singular n-simplexes, we can construct p-chains as finite formal sums

$$c = \sum_{\sigma} c_\sigma \sigma \,, \tag{3.102}$$

with $c_\sigma \in \mathbb{Z}$. Chains add according to addition in $\mathbb{Z}$. The group of all n-chains is the n-chain group $C_n(X)$.

The boundary map acts linearly on chains and can be extended to $\partial_n : C_n \to C_{n-1}$, forming a sequence

$$\cdots \to C_n(X) \to C_{n-1}(X) \to C_{n-2} \to \ldots \tag{3.103}$$

It can be proven that the boundary map is nilpotent, i.e. $\partial_n \circ \partial_{n-1} = 0$. That is to say, the boundary of a boundary vanishes for all chains in $C_n(X)$. This is equivalent to saying that the above sequence is an exact sequence. This allows us to define the homology group. First consider the n-cycle group $Z_n(X)$ consisting of all chains with vanishing boundaries,

$$Z_n(X) = \text{Ker } \partial_n \,, \tag{3.104}$$

and the n-boundary group $B_n(X)$ consisting of all chains which are a boundary of another $(n+1)$-chain,

$$B_n(X) = \text{Im } \partial_{n+1} \,. \tag{3.105}$$

The n-homology group is given by the quotient group

$$H_n(X) = H_n(C_n(X)) = Z_n(X)/B_n(X) \,. \tag{3.106}$$

This is well defined since $B_n(X)$ is a subgroup of $Z_n(X)$ since $\partial_n \circ \partial_{n-1} = 0$. An element of the homology group is an equivalence class of all cycles which are the boundary of another cycle. Informally, two chains are equivalent if they are deformable into each other.

3.B.2 Relative homology

We now consider the relative homology of a topological space X with respect to a subspace $A \subset X$. Consider the n-chain groups $C_n(A)$ and $C_n(X)$ of the topological spaces A and X. Now consider the quotient group

$$C_n(X)/C_n(A) \,. \tag{3.107}$$

Two elements in $C_n(X)$ correspond to the same equivalence class in the quotient group if they differ by a chain in A. Since the boundary operator ∂ operates trivially on $C_n(A)$ we can define the relative homology group

$$H_n(X, A; \mathbb{Z}) = H_n(C_n(X)/C_n(A)) \,. \tag{3.108}$$

The homology groups again consists of equivalence classes of chains on X. However this time a chain has to begin and end in the subspace A and is up to equivalence of chains in A.

3.B.3 Relative homology and Lefschetz thimbles

Picard-Lefschetz theory is developed using relative homology. Consider the region, A, of points in X below an upper bound $U \in \mathbb{R}$, *i.e.*,

$$A = \{x \in X | h(x) \leq U\} \subset X \,. \tag{3.109}$$

The set A normally consists of a number of disconnected regions A_i corresponding to the poles of the exponent. In these regions, the contribution to the integral is exponentially suppressed, the integrand is dominated by e^U. When the upper bound U is taken negative, *i.e.*, $U \ll 0$, the contribution of the integral in the region A is negligible. Now consider the relative homology of paths starting and ending in the subspace A,

$$H_n(X, A; \mathbb{Z}) = H_n(C_n(X)/C_n(A)) \,. \tag{3.110}$$

The relative homology consists of the equivalence classes of paths between the regions A_i, each of which can be identified with a steepest descent contour corresponding to a saddle point.

Consider the downwards flow $\gamma : I \subset \mathbb{R} \to X$, defined by

$$\frac{\mathrm{d}\gamma}{\mathrm{d}\lambda} = -g^{ij} \frac{\partial h}{\partial x^j} \,, \tag{3.111}$$

with g the metric. When the flow does not terminate on a critical point, the function h diverges to $-\infty$ in the limit $\lambda \to \infty$. The region of X for which the downward flow terminates in the component A_i as $\lambda \to \infty$, is the descending manifold $\mathcal{A}_i$.

We can also consider the region, $B \subset X$, consisting of the points above the lower bound $L \in \mathbb{R}$, *i.e.*,

$$B = \{x \in X | h(x) \geq L\} \,. \tag{3.112}$$

The set B again generally consists of the disjoint components B_i corresponding to the poles of the h-function. We define the ascending manifold $\mathcal{B}_i$ of the h-function as the subset of X terminating at B_i as $\lambda \to -\infty$. By intersecting the ascending and descending manifolds we obtain the Morse-Smale cells $\mathcal{C}_i$ which partition X [256, 303]. The boundaries of these cells, $\mathcal{C}_i$, are the steepest descent

$$\mathcal{J}_\sigma = \{x \in X | c(0) = x \,, \lim_{\lambda \to -\infty} \gamma(\lambda) = p_\sigma\} \,. \tag{3.113}$$

and the steepest ascent contours

$$\mathcal{K}_\sigma = \{x \in X | c(0) = x \,, \lim_{\lambda \to \infty} \gamma(\lambda) = p_\sigma\} \,. \tag{3.114}$$

corresponding to the saddle point p_σ.

Since the Morse-Smale cells, $\mathcal{C}_i$, partition the space X, it follows that equivalence classes of relative homology group $H_n(X, A; \mathbb{Z})$ can be uniquely identified with the steepest descent contours. Now, since the integration contour, $\mathcal{C}$, can be deformed to the steepest descent contour $\mathcal{J}_\sigma$ when the original contour is in the ascending manifold corresponding to the saddle point p_σ, it follows that the saddle points p_σ is relevant if and only if the steepest ascent contour $\mathcal{K}_\sigma$ intersects $\mathcal{C}$, *i.e.*,

$$n_\sigma = \mathrm{Int}(\mathcal{C}, \mathcal{K}_\sigma) \,. \tag{3.115}$$

Chapter 4

No smooth beginning for spacetime

> We can compare spacetime to an open, conic cub.... The bottom of the cup is the origin of atomic disintegration, it is the first instant at the bottom of space-time, the now which has no yesterday because, yesterday, there was no space.
>
> Georges Lemaître

Abstract

We identify a fundamental obstruction to any theory of the beginning of the universe, formulated as a semiclassical path integral. Hartle and Hawking's no-boundary proposal and Vilenkin's tunneling proposal are examples of such theories. Each may be formulated as the quantum amplitude for obtaining a final 3-geometry by integrating over 4-geometries. We introduce a new mathematical tool – Picard-Lefschetz theory – for defining the semiclassical path integral for gravity. The Lorentzian path integral for quantum cosmology with a positive cosmological constant is meaningful in this approach, but the Euclidean version is not. Framed in this way, the resulting framework and predictions are unique. Unfortunately, the outcome is that primordial tensor (gravitational wave) fluctuations are unsuppressed. We prove a general theorem to this effect, in a wide class of theories.

4.1 Introduction

In chapter 3, we discussed both the Hartle and Hawking [188, 182] and Vilenkin [323, 329], respectively HH and V, proposed for the quantum creation of universes using the gravitational path integral. In these proposals, one is supposed to integrate over 4-geometries g bounded by a final 3-geometry h. Formally, one writes

$$\text{HH}: \int^{h} [dg] e^{-S_E[g]/\hbar} \qquad \text{V}: \int_{\emptyset}^{h} [dg] e^{iS[g]/\hbar}. \tag{4.1}$$

Hartle and Hawking advocate integrating over compact Euclidean 4-geometries bounded by h, as a means to evaluate the highly oscillatory Lorentzian integral. Vilenkin advocates integrating over Lorentzian 4-geometries interpolating between a vanishing initial 3-geometry, labelled $\emptyset$, and h.

In chapter 3, we evaluated the integral over homogeneous and isotropic spacetimes with a positive cosmological constant Λ using Picard-Lefschetz theory [13, 342]. This led to an interesting saddle point solutions. We shall, in this chapter, carefully analyze both the Hartle-Hawking and Vilenkin proposals by treating the 4-geometry as a homogeneous, isotropic cosmological background with gravitational waves described by general relativistic perturbation theory. The path integral is taken over all contributing 4-geometries, modulo diffeomorphism equivalence. In a suitable time-slicing, illustrated in Fig. 4.1, any topologically trivial 4-metric may be expressed as $-N(x^k)^2 dt^2 + h_{ij}(t, x^k) dx^i dx^i$, where x^k are the space coordinates. One may choose t to run from 0 to 1, with the final 3-metric $h_{ij}(1, x^k)$.

When the lapse N is real, the four-geometry is Lorentzian; if imaginary, the four-geometry is Euclidean. Formally, one passes from the Lorentzian to the Euclidean theory with the replacement $N \to -iN \equiv N_E$, the sign being chosen to conform to the usual Wick rotation. Very generally, one cannot integrate N_E over the infinite real range $-\infty < N_E < \infty$. Any real Euclidean action obtained from a real Lorentzian action is necessarily odd in N_E. Furthermore, if its equations of motion are time-reversal invariant, they are even in N_E. Hence integrating out the dynamical variables always leaves one with an effective Euclidean action for N_E which is odd in N_E. If it diverges to $+\infty$ as $N_E \to +\infty$, then it diverges to $-\infty$ as $N_E \to -\infty$, and vice versa. Therefore, in any meaningful semiclassical Euclidean path integral, one cannot integrate N_E over all real values. There are three available options: i) integrate N_E over a half-range, should that integral converge; ii) leave the lapse real and Lorentzian, or iii) deform the lapse integral onto some other complex contour. We consider (and rule out) all three options.

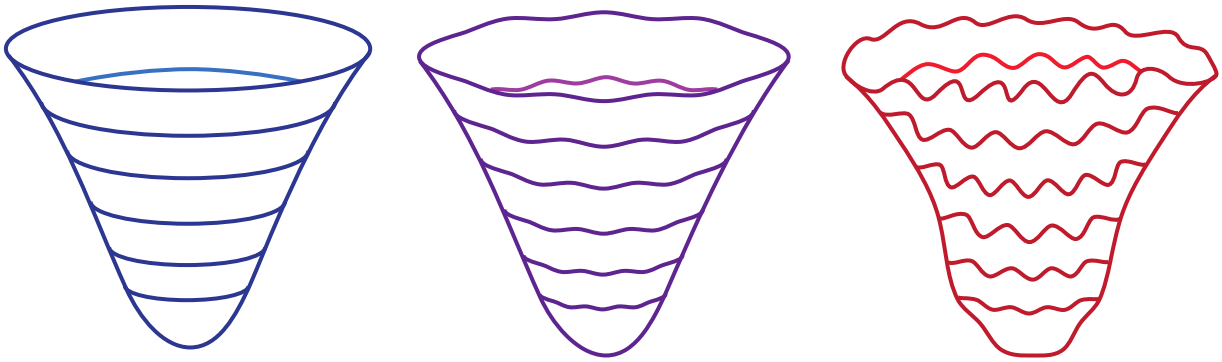


Figure 4.1: *Left:* the smooth, regular picture of the no-boundary background. *Middle:* the no-boundary picture with hoped-for small fluctuations, in agreement with observations. *Right:* the fluctuations implied by the more rigorous, Lorentzian-Picard-Lefschetz approach developed here. Our analysis shows that, to leading semiclassical order, large fluctuations are preferred, leading to a breakdown of the theory.

We perform the path integrals (4.1) in the saddle point (semiclassical) approximation. First, we integrate over the background scale factor. Then, neglecting backreaction, we integrate over tensor (gravitational wave) perturbations. Both path integrals are Gaussian and present no difficulty. Finally, we carefully integrate over the lapse, using Picard-Lefschetz theory to identify the relevant saddle points and steepest descent contours, as discussed in chapter 3.

Our key findings are as follows. First, for $\Lambda > 0$ the Euclidean path integral diverges, unlike the Lorentzian version. Taken over $0^+ < N_E < +\infty$, it diverges at $N_E = 0$ due to the "wrong sign" kinetic term for the scale factor. Integrating from $-\infty < N_E < 0^-$ it diverges due to the cosmological constant. Thus we revert to the Lorentzian path integral which, taken either over real $0^+ < N < \infty$, or $-\infty < N < \infty$, yields a conditionally convergent, meaningful result. Picard-Lefschetz theory allows us to render the integral absolutely convergent by distorting the N contour into the complex N-plane. We then obtain unambiguous predictions which are, unfortunately, unacceptable since they include unsuppressed perturbations on the final 3-geometry.

4.2 The conformal factor problem

There is a basic conundrum at the heart of quantum cosmology, whose resolution underlies our main claims. The problem is that the scale factor of the universe has a *negative* kinetic

term, unlike all other degrees of freedom. This simple, but fundamental fact prevents one from Wick rotating time so that the phase factor $e^{iS/\hbar}$ appearing in Lorentzian path integrals becomes a real suppression factor $e^{-S_E/\hbar}$ for all degrees of freedom. Our approach, discussed in chapter 3, is to perform no Wick rotation at all, but instead use Picard-Lefschetz theory to make sense of the original, Lorentzian path integral. In doing so, we uncover an important subtlety. For the simplest case of a closed, Λ cosmology, the relevant saddle is a round Euclidean four-sphere, just as Hartle and Hawking claimed, but obtained via the conjugate continuation from de Sitter. This inverts the semiclassical weighting, from $e^{+12\pi^2/(\hbar\Lambda)}$ to $e^{-12\pi^2/(\hbar\Lambda)}$, agreeing with Vilenkin and representing a more physically intuitive $\hbar \to 0$ limit.

However, the perturbations present new difficulties. The semiclassical amplitude is fixed by the complex, classical solution to the linearized Einstein equations giving the perturbation of the final 3-geometry. The Picard-Lefschetz construction ensures the convergence of the path integral and determines the prefactors uniquely. However, we will see that as a result of the abovementioned complex-conjugate nature of the background, the path integral yields an *inverse* Gaussian weighting for the final perturbation. Hence, large perturbations are favored and the theory is out of control.

4.3 The Bunch-Davies vacuum: quantum fields on a classical background

To set the stage, we briefly review the path integral computation of perturbations in the flat slicing of a *classical* de Sitter background. The line element is $a^2(\eta)(-d\eta^2 + d\vec{x}^2)$ with $a(\eta) = -1/(H\eta)$, (constant) Hubble parameter H and conformal time $-\infty < \eta < 0$. The Fourier modes of the perturbations decouple and can be treated independently. The quadratic action for a perturbation mode ϕ – for example, a gravitational wave – of wavenumber k takes the form $S^{(2)}_{0,1} = \frac{1}{2}\int_{\eta_0}^{\eta_1} \mathrm{d}\eta\, a^2(\eta)\left[(\phi_{,\eta})^2 - k^2\phi^2\right]$, with η_0 the initial and η_1 the final conformal time. We assume $|k\,\eta_0| \gg 1$ so that the perturbations start out in the local adiabatic vacuum at some early time η_0. For simplicity, we take $\eta_1 \to 0^-$, so the mode ends up frozen, with its physical wavelength far outside the Hubble radius. The amplitude for a final perturbation ϕ_1 is then given by

$$G^{(2)}_{\phi}[\phi_1] = \int \mathcal{D}\phi\, e^{iS^{(2)}_{0,1}[\phi]/\hbar - \frac{1}{2}ka_0^2\phi_0^2/\hbar}\,, \tag{4.2}$$

where the action $S^{(2)}_{0,1}$ incorporates the boundary conditions $\phi(\eta_{0,1}) = \phi_{0,1}$, and the functional measure includes an integral over ϕ_0. The second factor represents the initial (as-

sumed) adiabatic ground state wavefunction.

The functional integral is Gaussian so the saddle point method is exact. Stationarizing with respect to ϕ_0 and using the Hamilton-Jacobi equation $\partial S^{(2)}_{0,1}/\partial\phi_0 = -\pi_\phi(\eta_0) = -a^2\phi_{,\eta}(\eta_0)$, we find the saddle point solution to be "negative frequency" at early times. Solving the perturbation equation $\phi_{,\eta\eta} - (2/\eta)\phi_{,\eta} + k^2\phi = 0$, with the given boundary conditions, the classical solution is $\phi \approx \phi_1 e^{ik\eta}(1 - ik\eta)$. Evaluating the semiclassical exponent and carefully taking the limit $\eta_1 \to 0^-$, we find

$$G^{(2)}_\phi[\phi_1] \propto e^{-\frac{k^3}{2H^2}\phi_1^2 + i\frac{k^2}{2H^2\eta_1}\phi_1^2}\,. \tag{4.3}$$

In this thesis, we interpret the Feynman propagator as the amplitude to evolve from an initial to a final state. We interpreted the modulus squared as the relative probability density for this propagation to take place. For a more detailed discussion see Chapter 7 of this thesis. By squaring the propagator we observe that the divergent phase (which physically represents the final momentum of the mode) disappears and that we recover the familiar result of a scale-invariant power spectrum for ϕ_1.

The same result can be obtained by analytic continuation from the Euclidean theory. First, we Weyl transform the line element to flat space, and ϕ to $\chi = a\phi$. After an integration by parts, the Lorentzian action becomes $S^{(2)}_{0,1} = \frac{1}{2}\int_{\eta_0}^{\eta_1} \mathrm{d}\eta\, \left[(\chi_{,\eta})^2 - (k^2 - 2/\eta^2)\chi^2\right]$. Now we pass to Euclidean time $X \equiv i\eta$ and $S_E \equiv -iS$, obtaining

$$S_E = \frac{1}{2}\int_{X_0}^{X_1} \mathrm{d}X\, \left[(\chi')^2 + (k^2 + 2/X^2)\chi^2\right]\,, \tag{4.4}$$

with $' \equiv d/dX$, , *i.e.*,, a positive Euclidean action. We compute $G_\chi[\chi[X_1]]$ from the Euclidean path integral over χ. Again, we seek a classical saddle point solution. Finiteness of S_E imposes regularity at $X \to -\infty$, automatically selecting the ground state wavefunction. The desired classical solution is $\chi(X) = \chi_1 f(X)/f(X_1)$, with $f(X) = e^{kX}(1/X - k)$. The on-shell action is $S_E(X_1) = \frac{1}{2}\chi\chi'(X_1) = \frac{1}{2}\chi_1^2 f'(X_1)/f(X_1)$. We continue back to Lorentzian time by setting $X_1 = i\eta_1$. Taking the limit $\eta_1 \to 0^-$ again yields (4.3), with an additional phase generated from the change of variables from ϕ to χ.

4.4 Quantum fields on a quantum background

Let us now turn to a consistent semiclassical path integral treatment of *both* the background and the perturbations, in order to understand why this fails to yield the above-mentioned standard results.

4.4.1 The background

We assume a homogeneous and isotropic background cosmology: $ds^2 = -N_p(t_p)^2 \mathrm{d}t_p^2 + a(t_p)^2 \mathrm{d}\Omega_3^2$, with lapse function N_p, scale factor $a(t_p)$ and unit 3-sphere metric $\mathrm{d}\Omega_3^2$. The time t_p is the physical time if N_p is set to unity. The Einstein-Λ action for the background is

$$S_{0,1}^{(0)} = 2\pi^2 \int_0^1 \left[-3a \frac{a_{,t_p}^2}{N_p} + N_p(3a - a^3\Lambda) \right] \mathrm{d}t_p \,, \tag{4.5}$$

(in units where $8\pi G = 1$). As we saw in chapter 3, the path integral to evolve from $a(0) = 0$ to $a(1) = a_1$ is [307, 184]

$$G^{(0)}[a_1; 0] = \int_{0^+}^{\infty} \mathrm{d}N \int_0^{a_1} \mathcal{D}a \, e^{iS^{(0)}[a,N]/\hbar} \,. \tag{4.6}$$

Re-defining the lapse and the time coordinate via $N_p \, \mathrm{d}t_p \equiv (N \, \mathrm{d}t)/a$ renders the action quadratic in $q \equiv a^2$,

$$S^{(0)} = 2\pi^2 \int_0^1 \left[-\frac{3}{4N} \dot{q}^2 + N(3 - \Lambda q) \right] \mathrm{d}t \,. \tag{4.7}$$

The path integral over q can now be performed exactly[1]. The classical solution satisfying $q(0) = 0$, $q(1) = q_1$ is

$$q(t) = \frac{\Lambda}{3} N^2 t^2 + \left[q_1 - \frac{\Lambda}{3} N^2 \right] t \,. \tag{4.8}$$

The propagator reduces to

$$G^{(0)}[q_1; 0] = \sqrt{\frac{3\pi i}{2\hbar}} \int_{0^+}^{\infty} \frac{\mathrm{d}N}{N^{1/2}} e^{iS^{(0)}[q_1;0;N]/\hbar} \tag{4.9}$$

with

$$S^{(0)}[q_1; 0, N] = 2\pi^2 \left(N^3 \, \frac{\Lambda^2}{36} + N(3 - \frac{1}{2}\Lambda \, q_1) - \frac{3q_1^2}{4N} \right) \,. \tag{4.10}$$

This integral is then evaluated by deforming the integration contour into the complex N-plane, using Picard-Lefschetz theory [13, 342] to identify the relevant saddle points and steepest descent contours.

[1] Modulo issues regarding operator ordering and the path integral measure, and the restriction $q \geq 0$, further discussed in chapter 4 and [163].

The on-shell background action $S^{(0)}[q_1; 0, N]$ has four saddle points, each located in a different quadrant of the complex N-plane. The relevant saddle is located at

$$N_s = \frac{3}{\Lambda}\left(i + \sqrt{\frac{\Lambda}{3}q_1 - 1}\right), \tag{4.11}$$

yielding for the no-boundary propagator

$$G[q_1; 0] \propto e^{-\frac{12\pi^2}{\Lambda} - i4\pi^2\sqrt{\frac{\Lambda}{3}}\left(q_1 - \frac{3}{\Lambda}\right)^{3/2}}. \tag{4.12}$$

As discussed in chapter 3, Picard-Lefschetz theory implies semiclassical suppression, in agreement with Vilenkin but not with Hartle and Hawking.

We have performed the analogous calculation with a slow-roll inflaton field φ whose potential is well-approximated by $V(\varphi) \approx \Lambda - \frac{1}{2}m^2\varphi^2$ near $\varphi = 0$. We find that, as one would naively expect, for small φ_1,

$$G[q_1, \varphi_1; 0, 0] \propto e^{-\frac{12\pi^2}{V(\varphi_1)}} \times phase \tag{4.13}$$

so there is a higher weighting for a larger initial potential energy $V(\varphi)$. Given that the radius of the universe is approximately $\sqrt{3/V(\varphi)}$ when space and time become classical, this supports the intuition that it is easier to nucleate a small rather than a large universe.

The same results can be obtained in physical time t_p using the correspondence

$$\sinh(Ht_p) = H^2\, N\, t - i\,, \tag{4.14}$$

where we define $H = \sqrt{\Lambda/3}$ and $a(t_p) = \frac{1}{H}\cosh(Ht_p)$. The no-boundary point $t = 0$ corresponds to $Ht_p = -\frac{\pi}{2}i$.

4.4.2 The fluctuations

Let us now extend our analysis to include perturbations – for example, gravitational waves – treated at leading (quadratic) order. The full propagator is

$$G[q_1, \phi_1; 0] = \int_{0^+}^{\infty} \mathrm{d}N \int^{q_1} \mathcal{D}q \int^{\phi_1} \mathcal{D}\phi\, e^{iS/\hbar}\,, \tag{4.15}$$

where $S = S^{(0)}[q; 0, N] + S^{(2)}[q, \phi, N]$, with

$$S^{(2)}[q, \phi, N] = \frac{1}{2} \int N \mathrm{d}t \left[q^2 \left(\frac{\dot{\phi}}{N} \right)^2 - l(l+2)\phi^2 \right] , \tag{4.16}$$

and l the principal quantum number on the 3-sphere. For notational economy we explicitly include just one, orthonormalized mode ϕ; all modes occur in similar fashion. For tensor perturbations, $l \geq 2$ [158]. (In general, one may also have scalar or vector perturbations, with $l \geq 0$ and $l \geq 1$ respectively: see, *e.g.*, Ref. [170]). The lapse perturbation is nondynamical in the absence of matter and may be set to zero. The no-boundary condition is then implemented by specifying $q(0) = 0$ and requiring the action to be finite and stationary under all variations which vanish on the final boundary.

The path integral over the perturbations is again quadratic, so the saddle point method gives the ϕ_1 dependence exactly. The equation of motion for ϕ is $\ddot{\phi} + 2\frac{\dot{q}}{q}\dot{\phi} + \frac{N_s^2}{q^2} l(l+2)\phi = 0$, where we use the saddle point N_s of the background, neglecting backreaction. The finite action solution is $\phi(t) = \phi_1 F(t)/F(1)$, with

$$F(t) = \left(1 + \frac{i}{H^2 N_s t - i} \right)^{\frac{l}{2}} \left(1 - \frac{i}{H^2 N_s t - i} \right)^{-\frac{l+2}{2}} \left(1 - \frac{i(l+1)}{H^2 N_s t - i} \right) . \tag{4.17}$$

Note $\phi(t) \propto t^{\frac{l}{2}}$ as $t \to 0$, implying ϕ is regular there.

The classical action for the perturbations reduces to a surface term on the final boundary,

$$S^{(2)}[q_1, \phi_1, N_s] = \frac{1}{2} \int_0^1 \mathrm{d}t \frac{\mathrm{d}}{\mathrm{d}t} \left[\frac{q^2}{N_s} \phi \dot{\phi} \right] \tag{4.18}$$

$$= \frac{q_1^2}{2N_s} \phi_1^2 \frac{\dot{F}(1)}{F(1)} \tag{4.19}$$

$$= \frac{\phi_1^2}{2} \left[-\frac{l(l+2)}{H} \sqrt{q_1} - i \frac{l(l+1)(l+2))}{H^2} + \mathcal{O}\left(\frac{1}{\sqrt{q_1}} \right) \right] . \tag{4.20}$$

The full propagator for the perturbed background factorizes at this order $G[q_1, \phi_1; 0] = G[q_1; 0] G_\phi[\phi_1; 0]$, with

$$G_\phi[\phi_1; 0] \propto e^{\frac{l(l+1)(l+2)}{2\hbar H^2} \phi_1^2} \times phase \tag{4.21}$$

corresponding to an *inverse* Gaussian distribution.

In order to compare our results with the Bunch-Davies vacuum, we convert (4.17) to conformal time $\mathrm{d}\eta = \mathrm{d}t_p/a$. The physical time and the conformal time are related by $\tan\left(\frac{\pi}{4}+\frac{\eta}{2}\right) = \tanh\left(\frac{Ht_p}{2}\right)$, where $-\infty < t_p < \infty$ corresponds to $-\pi < \eta < 0$. Thus, as $\eta \to 0$,

$$\sinh(Ht_p) = 2\frac{\tan(\frac{\pi}{4}+\frac{\eta}{2})}{1-\tan^2(\frac{\pi}{4}+\frac{\eta}{2})} \to -\frac{1}{\eta} + \frac{\eta}{3} + \frac{\eta^3}{45} + \dots$$

which, using (4.14), leads to the late time approximation

$$\phi = \phi_1 \left[1 + \frac{1}{2}l(l+2)\eta^2 - \frac{i}{3}l(l+1)(l+2)\eta^3 + \dots\right].$$

This is the late time expansion of the "positive frequency" mode function, confirming that the no-boundary condition selects the "wrong" mode function as compared to the adiabatic ground state.

4.5 More general smooth beginnings

Having demonstrated our claim that the perturbations are out of control in the no-boundary description of quantum de Sitter spacetime, we would like to establish how general the result is. To begin with, we shall consider a fluid more general than a cosmological constant, but which is still "adiabatic", namely, the background pressure P is a function of the energy density ρ so that there is a unique cosmological history parameterized by the scale factor a. Furthermore, we assume this classical evolution results in a smooth "bounce" of the scale factor such as occurs in the closed slicing of de Sitter spacetime.

From our discussion above, it is clear that the on-shell classical action is all that is needed to determine the semiclassical exponent in the quantum propagator both for the background and for the perturbations. In the no-boundary solutions, $q = a^2$ runs from $q_0 = 0$ to q_1, a positive value. Thus q itself may be used as a time coordinate. The Friedmann constraint allows us to express the background line element as

$$ds^2 = -\frac{\mathrm{d}q^2}{4q\,(\frac{1}{3}\rho(q)q - 1)} + q\,\mathrm{d}\Omega_3^2, \tag{4.22}$$

where we allow the energy density $\rho(q)$ to vary with q.

Cauchy's theorem enables us to deform the time (or q) contour upon which we evaluate the classical action as long as it does not cross any singularity. In particular, we

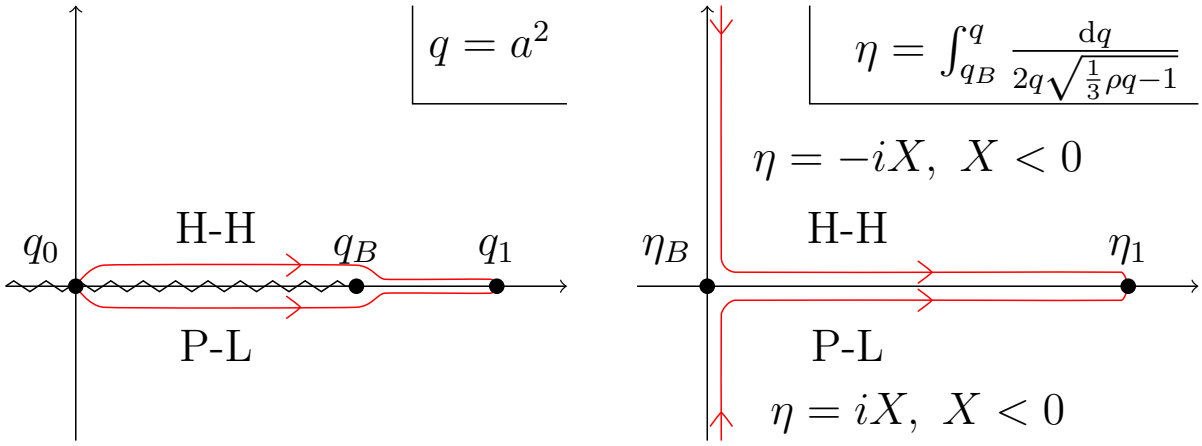


Figure 4.2: *Left:* Analytic continuation contours (red) in the Hartle-Hawking (H-H) and Picard-Lefschetz (P-L) descriptions, above and below the branch cut in the complex q-plane. *Right:* Corresponding contours for the conformal time.

can deform it to one in which q is real everywhere. The line element (4.22) is Lorentzian for $q > 3/\rho(q)$ but Euclidean for $0 \leq q < 3/\rho(q)$, and is easily checked to be regular at $q = 0$. At $q = 3/\rho(q)$, where $q = q_B$, the real, Lorentzian solution "bounces," and q therefore ceases to be a single-valued time coordinate. Our complex saddle point solution (4.8) passes *below* this point in the complex q-plane: it is precisely this topological fact which results in the suppression of the semiclassical amplitude, required by Picard-Lefschetz theory (see chapter 3). Using the Friedmann constraint, the classical action (4.7) gives $iS^{(0)} = -6\pi^2 i \int \mathrm{d}q \sqrt{\rho\, q/3 - 1}$. Since we start in the Lorentzian region we take the branch cut to run leftwards from the point q_B, the classical "bounce." Continuing the q integral below the branch cut to $q = 0$, we obtain for the real part of the semiclassical exponent $-6\pi^2 \int_0^{q_B} \sqrt{1 - \rho\, q/3}$. For a cosmological constant $\rho(q) = \Lambda$, we obtain $-12\pi^2/\Lambda$. Continuing above the branch cut yields $+12\pi^2/\Lambda$, Hartle and Hawking's result, which is inconsistent with Picard-Lefschetz theory.

To analyze the perturbations, we pass to coordinates in which the metric is conformally static: for $q > q_B$, we set $d\eta = dq/(2q\sqrt{\rho\, q/3 - 1})$ to obtain the line element $q(\eta)(-d\eta^2 + \mathrm{d}\Omega_3^2)$. We take $\eta = 0$ to correspond to the "bounce," so η is positive in the Lorentzian region. Now, when q passes below the branch cut commencing at q_B, the square root in the definition of $d\eta$ means that η continues from the positive real η-axis onto the negative imaginary η-axis, $\eta = iX$ with $X < 0$ in the Euclidean region. Conversely, following X forward from the Euclidean region, it "turns right" into the Lorentzian region, whereas in the usual Wick rotation, assumed by Hartle and Hawking, it "turns left" (see Fig. 4.2). Taking the continuation implied by Picard-Lefschetz theory for the background, the

Euclidean action for the perturbations has the "wrong" sign. We can still impose regularity of the modes in the Euclidean region, but the resulting semiclassical weighting factor will inherit the wrong sign.

As in our earlier discussion, it is convenient to go to a Weyl frame in which the kinetic terms are canonical. So we set $\phi = \chi/a$, obtaining for the Lorentzian action

$$iS^{(2)} = i\pi^2 \int d\eta \left[(\chi_{,\eta})^2 + \frac{a_{,\eta\eta}}{a}\chi^2 - l(l+2)\chi^2 \right] . \tag{4.23}$$

The background equations imply that $a_{,\eta\eta}/a = \frac{1}{2}(\frac{1}{3} - w)\rho\, a^2 - 1$, where $w = P/\rho$ is the equation of state. Analytically continuing η back into the Euclidean region and then on to $q = 0$ (corresponding to $X = -\infty$), as explained above, we obtain the Euclidean action

$$-S_E^{(2)} = \pi^2 \int_{-\infty}^{0} dX \left[\chi'^2 + U(X)\chi^2 \right] , \tag{4.24}$$

where $\chi' \equiv d\chi/dX$ and $U(X) \equiv l(l+2)+1+\frac{1}{2}q\,(w_E - \frac{1}{3})\rho_E$. Here, w_E and ρ_E are the analytic continuations of their Lorentzian counterparts into the Euclidean region. Whatever the equation of state of the matter, $U(X)$ is positive at large l, since regularity demands that ρ_E remains finite, and correspondingly $w_E \to -1$, as $q \to 0$. In fact, $U(X)$ is positive for all tensor modes as long as $\rho_E > 0$ and $w_E > -17/3$. As before, the propagator's dependence on the final perturbation χ_1 is given by the classical action. Finiteness of the action selects the mode $\chi = f(X)$ which is regular at $q = 0$, *i.e.*, which vanishes at $X = -\infty$ (in the large l limit, $f(X) \sim e^{\sqrt{l(l+2)}X}$). Using an integration by parts, from (4.24) we obtain the on-shell Euclidean action $-S_E^{(2)} = \pi^2\chi_1^2 f'(X_1)/f(X_1)$. The quantity $f'(X)/f(X)$ is positive at $X = -\infty$: as long as $U(X)$ is real and positive, the classical equation of motion for f implies $f'(X)/f(X)$ remains positive throughout the Euclidean region.

Continuing the conformal time into the Lorentzian region, we can show that the real part of the semiclassical exponent remains positive. Expressing the mode function in terms of its real and imaginary parts, $f(X) = R(X) + iI(X)$, we have shown that $\mathrm{Re}[f'/f] = (RR' + II')/(R^2 + I^2) > 0$ at $X = 0$. When X turns in the negative imaginary direction, $X = -i\eta$, with η positive, the Cauchy-Riemann equations yield $R' + iI' = i(R_{,\eta} + iI_{,\eta})$. Therefore, at $X = \eta = 0$, we have $R_{,\eta} = I'$ and $I_{,\eta} = -R'$ and follows that the Wronskian $IR_{,\eta} - RI_{,\eta}$, which is independent of η, equals $(R^2+I^2)\mathrm{Re}[f'/f]$ at $X = 0$, which is positive. Now, the real part of the semiclassical exponent, at a final Lorentzian time η_1 is similarly given, after an integration by parts, by $\pi^2\chi_1^2\mathrm{Re}[i\, f_{,\eta}(\eta_1)/f(\eta_1)] = \pi^2\chi_1^2(IR_{,\eta} - RI_{,\eta})/(R^2 + I^2)$ (in fact, I vanishes there by assumption). Since the Wronskian is positive, it follows that the semiclassical exponent for the perturbation χ_1 is positive, for all positive η.

In more general situations, the background pressure may not be expressible in terms of the density. In this case, it may not be possible to describe both the Euclidean and Lorentzian regions in terms of a real potential U. Nevertheless, even in this more general situation, where the "bounce" point q_B satisfying $q_B = 3/\rho_B$ is complex, we still need to pass below it in the complex q-plane to be consistent with Picard-Lefschetz theory. This topological result again implies that the conformal time η runs from $-i\infty$ in the region around $q = 0$ to positive, nearly real values in an approximately "Lorentzian" region. For modes of large l, the (in general complex) potential $U(X)$ is dominated by the l^2 term, and the no-boundary solution is accurately described by the WKB Euclidean growing mode, so that $\mathrm{Re}[f'/f] \sim \sqrt{l(l+2)} + O(l^{-1})$ at large l. The arguments above again demonstrate that the final semiclassical exponent has a positive real part. We conclude that the problem of unbounded perturbations, at small wavelengths, is unavoidable.

4.6 Discussion

The no-boundary and tunneling proposals had as their objective to provide theories of initial conditions for the universe, and in particular to explain the initial smoothness of the universe. As we have demonstrated here, when analyzing these proposals in a well-defined mathematical setting, the picture that emerges is rather the opposite. Large perturbations are preferred, to such an extent that the propagator becomes non-normalizable and the entire framework fails. In the next chapter, we show that the situation cannot be improved by considering steeper inflationary potentials, even in cases where the lapse function becomes real at the saddle points (and even for negative ekpyrotic potentials the same problems arise). We furthermore show that back-reaction does not significantly change the situation. A smooth semi-classical beginning to the universe, where the big bang singularity is avoided, is thus not an option.

Chapter 5

No rescue for the no-boundary proposal: pointers to the future of quantum cosmology

> The trouble with changing the rules of a game (in this case the standard rules of quantum gravity) is that there are no rules about changing the rules.
>
> Arlen Anderson and Bryce DeWitt

Abstract

In chapters 3 and 4, we introduced Picard-Lefschetz theory as a tool for defining the Lorentzian path integral for quantum gravity in a systematic semiclassical expansion. This formulation avoids several pitfalls occurring in the Euclidean approach. Our method provides, in particular, a more precise formulation of the Hartle-Hawking no-boundary proposal, as a sum over real Lorentzian four-geometries interpolating between an initial three-geometry of zero size, *i.e*, a point, and a final three-geometry. With this definition, we calculated the no-boundary amplitude for a closed universe with a cosmological constant, assuming cosmological symmetry for the background and including linear perturbations. We found the opposite semiclassical exponent to that obtained by Hartle and Hawking for the creation of a de Sitter spacetime "from nothing". Furthermore, we found the linearized perturbations to be governed by an *inverse* Gaussian distribution, meaning they

are unsuppressed and out of control. Recently, Diaz Dorronsoro *et al.* [104] followed our methods but attempted to rescue the no-boundary proposal by integrating the lapse over a different, intrinsically complex contour. Here, we show that, in addition to the desired Hartle-Hawking saddle point contribution, their contour yields extra, non-perturbative corrections which again render the perturbations unsuppressed. We prove there is *no* choice of complex contour for the lapse which avoids this problem. We extend our discussion to include backreaction in the leading semiclassical approximation, fully nonlinearly for the lowest tensor harmonic and to second order for all higher modes. Implications for quantum de Sitter spacetime and for cosmic inflation are briefly discussed.

5.1 Introduction

The no-boundary proposal of Hartle and Hawking represents an attempt to explain the quantum origin of spacetime and provide an initial condition for cosmic inflation [194, 188]. All it apparently requires is:

(i) domination of the energy density by a positive cosmological constant or gently sloping scalar field potential, just as is assumed for inflation,

(ii) a closed (positively curved, compact) universe, and

(iii) that the quantum mechanical amplitude for a given three-geometry Σ be given by the Feynman path integral over all compact four-geometries bounded only by Σ.

The latter geometrical picture, in particular, offers to realize a hope dating back to the very beginnings of modern cosmology [239, 314, 58, 325], that the unification of quantum mechanics and general relativity might resolve the big bang singularity and explain the beginning of the universe. The no-boundary proposal has furthermore been influential well beyond cosmology, particularly in areas of mathematical physics including holography as well as conformal and topological field theory, where it has been used to motivate and define interesting quantum states.

However, since its beginnings, the proposal has suffered from the lack of a precise mathematical formulation. In chapters 3 and 4, we attempted to rectify this shortcoming. Our starting point is the *Lorentzian* path integral for quantum gravity, treated as a low energy, effective field theory within a semiclassical expansion. We argued that in the presence of a positive cosmological constant, the Lorentzian path integral propagator to evolve from a geometry of zero initial size to a given final three-geometry provides a mathematically meaningful definition of the no-boundary amplitude. This is precisely the

point of view adopted by Vilenkin in his early papers, although he never performed any path integral calculations. Instead, he imposed "outgoing" boundary conditions on solutions of the Wheeler-DeWitt equation, a prescription which is, however, incomplete when perturbations are considered. Our Lorentzian path integral formulation in contrast allows us to simultaneously handle both the background and perturbations with no ambiguities.

5.1.1 No-boundary de Sitter: Picard-Lefschetz theory

We chose to focus on the simplest example of a quantum cosmology, namely the no-boundary or "tunneling from nothing" version of quantum de Sitter spacetime in the closed slicing, performing a semiclassical quantization of both the background and the perturbations. On the positive side, we found unique, well-defined results, free of the diseases such as the "conformal factor" problem, which plague the Euclidean approach. However, we also found unexpectedly negative results concerning the semiclassical Hartle-Hawking state for quantum fields and fluctuations.

We claim that the Lorentzian path integral amplitude for a closed universe with a positive cosmological constant Λ to emerge "from nothing" into a period of de Sitter expansion, acquiring a frozen, dimensionless tensor perturbation ϕ_1 on the final three-geometry, is given by

$$\int \mathcal{D}g e^{iS[g]/\hbar} \propto e^{-\frac{12\pi^2}{\hbar\Lambda}+\frac{3}{2\hbar\Lambda}l(l+1)(l+2)\phi_1^2}, \tag{5.1}$$

with $S[g]$ being the usual Einstein-Hilbert-Λ action taken in units where $8\pi G = 1$. The path integral is taken over all compact four geometries bounded only by the final three-geometry. Here, we omit the functional determinant and a phase representing the late time classical evolution since we wish to focus on the semiclassical weighting factor, given by the real part of the semiclassical exponent. To avoid notational clutter, we also include just one tensor mode, with principal quantum number l and amplitude ϕ_1 on the final three-geometry, assuming that the physical wavelength of that mode is larger than the de Sitter radius at the final time, so that the mode has dynamically "frozen out". To quadratic order in the perturbations, any number of modes may be included by simply replacing $l(l+1)(l+2)\phi_1^2$ with $\sum_{lmn} l(l+1)(l+2)\phi_{1,lmn}^2$, where $\phi_{1,lmn}$ are the coefficients in the expansion of the final tensor perturbation in real, orthonormalized spherical harmonics on the three-sphere, with quantum numbers l, m, n (see, *e.g.*, [158]).

The sign of the exponent in (5.1) is the *opposite* of that usually associated with the Hartle-Hawking no-boundary proposal. Hartle and Hawking obtained their result by considering the Euclidean action for quantum gravity, obtained by performing the usual Wick

rotation on quantum fields (including tensor modes) and finding a saddle point representing a Euclidean four-sphere. This procedure recovers the usual Euclidean vacuum for quantum fields at short distances.

Unfortunately, as we showed in chapter 3, the Euclidean path integral for the relevant cosmological background is, in the case at hand, a meaningless divergent integral. We avoided that problem by *not* Wick rotating: instead, we evaluate the Lorentzian path integral directly. Our main tool is Picard-Lefschetz theory, a powerful and rigorous means of converting an oscillatory, conditionally convergent multidimensional integral into a sum of absolutely convergent, steepest descent integrals. In this case, we find perfectly unambiguous results for both the background and the perturbations. However, because the background scale factor has a kinetic term of the opposite sign to that of the perturbations (and other quantum fields), when we integrate out the background, this in effect imposes a Wick rotation of the opposite sign to the usual one, yielding an inverse Gaussian distribution for quantum fields and implying that the perturbations are out of control. On this basis, we claimed that the no-boundary proposal (or its "tunneling" equivalent) cannot in any way describe the emergence of a realistic cosmology.

Let us now discuss the two terms in the exponent of (5.1), since there is an interesting story behind each of them. The first term comes from integrating out the cosmological background. It is convenient to rewrite the usual FLRW cosmological line element $-dt^2\bar{N}^2(t) + a(t)^2 d\Omega_3^2$ as $-dt^2 N^2/q(t) + q(t)d\Omega_3^2$, with $q(t) = a(t)^2$ and $d\Omega_3^2$ the standard metric on the unit three-sphere. Fixing a gauge in which the background lapse N is a constant, the path integral over $q(t)$, being Gaussian, may be performed without difficulty. One is left with an ordinary, one-dimensional integral over N, given in equation (5.13) below, with the exponent given in terms of the appropriate classical action (5.10).

Figure 5.1 exhibits the structure of the real part of the exponent (the h-function) in the complex N-plane. The orange points indicate N_σ, the saddle points of the h-function, which are also (by the Cauchy-Riemann equations) stationary points of the exponent, which is a holomorphic function of N. Since the effective action is real for real N, these saddles come in complex conjugate pairs. At these saddle points the four-geometry is a completely regular solution of the complexified Einstein equations. Hartle and Hawking took the two lower saddles, labelled 3 and 4. However, when we define the contour $\mathcal{C}$ to be that appropriate to the causal Lorentzian propagator, or its complex conjugate, Picard-Lefschetz theory identifies the two conjugate saddles, 1 and 2 respectively, as the relevant ones. The classical action for the upper two saddles is the complex conjugate of that of the lower two saddles, $S_{cl}(N_\sigma^*) = S_{cl}^*(N_\sigma)$. Hence their semiclassical weighting factor $|e^{iS_{cl}/\hbar}| = e^{i(S_{cl}-S_{cl}^*)/(2\hbar)}$ is the inverse of the weighting for the two lower ones.

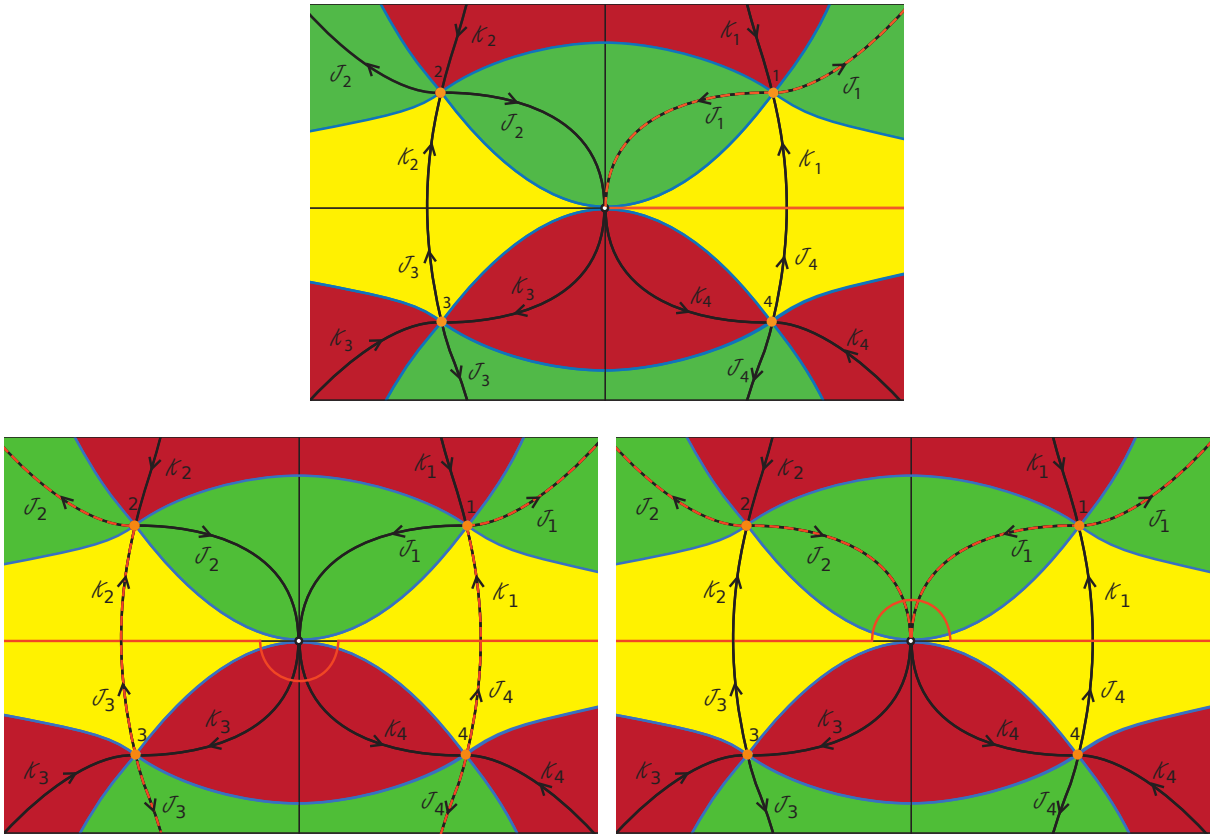


Figure 5.1: The h-function for the background is plotted in the complex N-plane, for a closed, homogeneous and isotropic Λ-dominated cosmology. The solid orange line is the defining integration contour $\mathcal{C}$, and the dashed orange line is the corresponding deformed contour, passing along Lefschetz thimbles. As N tends to infinity in the complex N-plane, the real part of the exponent in the integrand tends to $+\infty$ in the red regions or $-\infty$ in the green regions. It is constant along the blue contours. *Upper panel:* the real Lorentzian contour $0^+ < N < \infty$ used for the causal Lorentzian propagator. *Lower left panel:* the contour for N running from $-\infty$ to $+\infty$ *below* the origin, as proposed by Diaz Dorronsoro *et al.* [104]. *Lower right panel:* the real part of the causal propagator, equivalent to a continuous contour for N running from $-\infty$ to $+\infty$ *above* the origin.

Once the saddle points are identified, Picard-Lefschetz theory allows us to transform conditionally convergent integrals into absolutely convergent integrals as follows. Each saddle point σ is generically the intersection of two contours – one of steepest descent, labelled $\mathcal{J}_\sigma$, and one of steepest ascent, labelled $\mathcal{K}_\sigma$. The real part of the exponent (the h-

function) decreases monotonically on the former and increases monotonically on the latter. The steepest descent and steepest ascent contours are shown in black. The solid orange lines in the Figure show three possible choices we shall consider for the contour $\mathcal{C}$ over which the integral over N may be taken. In all cases, we take $\mathcal{C}$ to run from one singularity where the h-function diverges to $-\infty$ to another. It is a general result of Picard-Lefschetz theory that in order for a saddle point to be relevant to an integral over $\mathcal{C}$, the steepest ascent contour from that saddle must intersect $\mathcal{C}$ (see, *e.g.* chapter 3). This being the case, provided the exponent is holomorphic in the relevant region of N one can deform $\mathcal{C}$ into the complex N-plane so that it passes over $\mathcal{J}_\sigma$, with Cauchy's theorem ensuring that the value of the integral is preserved. One must also be careful to check that any additional arcs introduced near the two limits of the integral (in our case, near $N = 0$ and $N = \infty$) give a vanishing contribution.

The upper panel in the Figure shows the defining contour for the causal Lorentzian propagator, $0^+ < N < \infty$. One can deform this contour by "sliding it down" the steepest ascent contour $\mathcal{K}_1$ onto the steepest descent contour $\mathcal{J}_1$, known as a "Lefschetz thimble" (the dashed orange line). In this way one obtains an equal, absolutely convergent integral over N. Since a saddle point is relevant if and only if its steepest ascent contour intersects the original integration contour, and since the classical action is real-valued on the real line (so the h-function is zero there), it follows that the real part of the semiclassical exponent at any relevant saddle must always be negative. As we argued in chapter 4, this argument is already sufficient to rule out the Hartle-Hawking result.

How, then, did Hartle and Hawking reach the opposite conclusion? They took the saddle points in the lower-half N-plane to be the relevant ones, on the basis that one should reproduce the usual Wick rotation for quantum fields, but they never explicitly performed the path integral over the lapse. First, consider integrating $N_E = iN$ over the infinite real range $-\infty < N_E < \infty$. Any real Euclidean action obtained from a real Lorentzian action is necessarily odd in N_E. Furthermore, if its equations of motion are time-reversal invariant, they are even in N_E. Hence, in the absence of any singular behavior, integrating out the dynamical variables always leaves one with an effective Euclidean action for N_E which is odd in N_E. If it diverges to $-\infty$ as $N_E \to +\infty$, then it diverges to $+\infty$ as $N_E \to -\infty$, and vice versa. So the semiclassical path integral over all N_E always diverges. Therefore, in any meaningful semiclassical Euclidean path integral, one simply cannot integrate N_E over all real values. Note that this means that a semiclassical Euclidean path integral *cannot* be used to obtain a solution of the homogeneous Wheeler-DeWitt equation, or a "wave function of the universe," as Hartle and Hawking hoped. There are three available options: i) integrate N_E over a half-range, should that integral converge; ii) leave the lapse real and Lorentzian, or iii) deform the lapse integral onto some other complex contour. We consider

all three options in this chapter, and show none is viable.

Exploring the first option, if one integrates N_E over positive values, the integral diverges at the origin (in the red region below it, shown in Fig. 5.1). This divergence is due to the essential singularity at $N = 0$, which is nothing but the usual one for a quantum mechanical propagator in the limit of short times. For the Einstein action, with the condition that the initial three-geometry has zero size, at small N the propagator behaves as $e^{-i3\pi^2 q_1^2/(2\hbar N)}$, where q_1 is the value of the scale factor squared on the final three-geometry. The minus sign is unusual and due to the negative kinetic term for the scale factor. Conversely, if one integrates N_E over negative values, it diverges at $-\infty$ in the uppermost red region in the Figure. Hence, there is no Euclidean contour for the lapse which gives a meaningful result. Hence, in our work we reverted to option ii) and integrated over real $0^+ < N < +\infty$. As we shall explain, integrating over all real N yields the real part of our answer, so one obtains (5.1) once again.

5.1.2 Hartle-Hawking rescued?

In their recent paper, Diaz Dorronsoro *et al.* attempted to recover the predictions of the original Euclidean formulation of the no-boundary proposal path integral by following our Lorentzian-Picard-Lefschetz approach. They claim to identify a new contour for the lapse, shown as the solid orange contour in the lower left panel of Figure 5.1, which recovers both the original Hartle-Hawking weighting for the background and a Gaussian distribution for the fluctuations. Their contour runs from $N = -\infty$ to $N = +\infty$, passing *below* the essential singularity at $N = 0$. It is immediately apparent that their contour cannot be deformed onto the real N-axis, to make the spacetime four-metric real and Lorentzian, because the integrand diverges as one approaches the origin from below. Hence their contour cannot be legitimately termed Lorentzian.

Diaz Dorronsoro *et al.* emphasize that the path integral along their contour is real, despite the contour being complex. This is indeed the case because the Lorentzian action is odd in N and their contour is even under $N \to -N^*$. Second, they stress that it solves the homogeneous Wheeler-DeWitt equation, whereas our causal Lorentzian propagator does not. Combining these points in a rhetorical flourish, they emphasize that their construction provides a "real" wavefunction. In section 5.2 of this chapter, we discuss the basic physical principles underlying the causal Lorentzian propagator in quantum gravity, explaining why it is complex, like the Feynman propagator for a relativistic particle or a string, and why, when the Hamiltonian is applied, it yields $-i$ times a delta function, rather than zero. Should we wish to, we may trivially obtain a "real" wavefunction (in both senses) from our

causal Lorentzian propagator merely by taking its real part. This is equivalent to using the orange contour illustrated in the lower right panel of Figure 5.1. Since the integrand becomes exponentially small as the origin is approached from the upper half N-plane, one may equivalently describe the contour in terms of two disconnected pieces, $-\infty < N < 0^-$ and $0^+ < N < \infty$, or over a single complex contour running from $-\infty$ to ∞ passing above the origin. The contour proposed by Diaz Dorronsoro *et al.* has no "real" advantage over the causal, Lorentzian propagator (or its real part) in these terms. As we discuss in section 5.2, solutions of the homogenous Wheeler-DeWitt equation are arbitrary without further information about the quantum state. In contrast, the causal "no-boundary" propagator as we have described it is in principle unique.

Nevertheless, let us further investigate their proposed wavefunction. It is not hard to see that the Hartle-Hawking saddles 3 and 4 are indeed relevant to Diaz Dorronsoro *et al.*'s proposed contour, as they claim. All we need to do is follow the steepest ascent contours, $\mathcal{K}_3$ and $\mathcal{K}_4$ from the saddle points towards the essential singularity at $N = 0$. Since they intersect the small orange semicircle below the origin (where the integrand is diverging), Picard-Lefschetz theory tells us they are relevant. The real part of the semiclassical exponent is indeed allowed to be positive, precisely because their defining contour is not Lorentzian. However, if we follow the steepest ascent contours $\mathcal{K}_1$ and $\mathcal{K}_2$ from the upper two Lorentzian-Picard-Lefschetz saddles, we see that these *also* intersect their contour and thus all four saddles are relevant to their contour. As a consequence, their wavefunction includes contributions of the form of (5.1), bringing along with them unsuppressed perturbations.

The idea of using more general contours for the lapse goes back many years. Halliwell and Louko, in particular, investigated steepest descent contours in de Sitter minisuperspace models, producing contour diagrams very similar to ours [186]. Halliwell and Hartle further developed the idea, realizing also that certain saddle point solutions would lead to unacceptable quantum field theory distributions [185]. In fact, they used this very argument against the tunneling proposal. (For a related discussion, from the Wheeler-DeWitt point of view, see also [287]). They seemingly did not, however, appreciate our point here, which is that any contour for N, running from one singularity of the h-function to another and yielding a convergent integral, *inevitably* includes contributions from the unacceptable saddle points. In section 5.6 of this chapter, we shall prove by simple enumeration that *no* contour for N avoids the problem of unsuppressed fluctuations.

5.1.3 Perturbation conundrum

The observant reader will have noticed a logical conundrum raised by the above arguments. If Picard-Lefschetz theory tells us that the real part of the semiclassical exponent is always negative, how do we explain the dependence of (5.1) on the perturbation amplitude? Clearly, by increasing ϕ_1, one can make the real part of the exponent arbitrarily positive. Of course, linear perturbation theory breaks down at large ϕ_1, so one might hope that nonlinearities somehow prevent the second term from ever overcoming the first. This, we shall show in Section 5.5, is not the correct explanation. Instead, something more subtle and interesting is going on. The point is that general relativity is not a regular theory. In particular, time evolution generically allows for the development of singularities. Around these singularities, the perturbations develop unusual, non-analytic behavior. We shall show that this introduces branch cuts into the effective action for the lapse N and breaks the analyticity assumptions underlying the use of Picard-Lefschetz theory. For the de Sitter model with perturbations, the singularities do not occur on the Picard-Lefschetz thimble for the higher-dimensional theory. Hence they do not introduce any ambiguity into our results. Rather they occur on the original, defining contour for the Lorentzian path integral, if one integrates out the background and the perturbations, at real, off-shell values for the lapse N, *i.e.*, for real, Lorentzian off-shell and singular four-geometries. For the perturbations (and only for the perturbations!), this breakdown of analyticity allows the effective exponent for N to gain a positive real part as one approaches the real N-axis from above.

Before delving into further detail, let us outline the basic steps in our approach. We fix a convenient gauge, in which the lapse N is a constant in time and the perturbations are taken to be transverse-traceless to eliminate unphysical degrees of freedom. Then we perform the Lorentzian path integral for Einstein-Λ gravity in three steps. First, we integrate out the radius squared $q(t)$ of the spherical background universe. As we have mentioned, this is a Gaussian path integral presenting no difficulties. Next, we integrate out the perturbations, treated to quadratic order in general relativistic linear perturbation theory. Finally, we integrate over the lapse N.

Picard-Lefschetz theory plays an important role in ensuring these calculations make sense. Let us start by assuming that we can represent all variables appearing in the path integral as finite sums over Fourier modes in the cosmological time t and, furthermore, that the answer is independent of any UV cutoff appearing in this sum. In this way, the path integrals become ordinary integrals, albeit in high dimension. Picard-Lefschetz theory (and Cauchy's theorem) may now be rigorously used to deform the original, highly oscillatory integral into an equivalent, absolutely convergent integral over a many-dimensional Picard-

Lefschetz thimble. Once this has been done, Fubini's theorem (see, *e.g.*, [3]) assures us we may evaluate the high-dimensional integral iteratively as a series of one dimensional integrals, and that the final result will be independent of the order in which those integrals are performed. If, on the contrary, we do not distort all contours to the Picard-Lefschetz thimble before integrating out some variables, we can easily generate singularities on the original, real contour for the remaining variables. An example is provided in Appendix 5.A, showing how such singularities are generated and, equally, how they are avoided in the higher-dimensional Picard-Lefschetz procedure.

We followed this Picard-Lefschetz procedure in calculating the causal, Lorentzian propagator in the approximation where we neglected backreaction of the perturbations on the background. It led unambiguously to the result (5.1). How, then, did it generate a positive real, semiclassical exponent where Picard-Lefschetz flow arguments would appear to forbid one? The explanation lies in the fact that, for a range of real but off-shell values of N, the background metric develops singularities which lead to non analytic behavior of the perturbations. The consequence is that integrating out the perturbations generates a pair of finite branch cuts on the real N-axis. These give a positive real part to the semiclassical exponent on the upper side of the real N-axis, including the points where the steepest ascent contours from saddles 1 and 3 meet the real axis and therefore allowing for those saddle points to contribute positively to the semiclassical exponent.

5.1.4 Resolution: real strong singularities

It is important to emphasize that the singularities we are discussing only occur at real values of N and *do not* occur on the Picard-Lefschetz thimble $\mathcal{J}_1$ relevant to the causal Lorentzian propagator. Therefore, they introduce no ambiguity in our result (5.1). However, they *do* occur on the defining contour for the Lorentzian path integral, after the perturbations are integrated out. In that sense, they are similar to the singularities in the prefactor described in Appendix 5.A, which again occur after a partial integration is performed. However, and this is a key point, in any finite-dimensional Gaussian integral of the form $\int d\vec{x} e^{i(\vec{x}^T O \vec{x} + \vec{u}^T \vec{x})/\hbar}$, where the matrix O is real, symmetric and nonsingular and the vector $\vec{u}$ is real, any number of partial integrations will not alter the imaginary nature of the exponent. Each such integration may be performed as a saddle point integral, and the integrated out variable always assumes a real value at its unique saddle point value. Hence, after any number of partial integrals, the real part of the remaining semiclassical exponent is always equal to zero. Furthermore, if the matrix elements of O are merophorphic in some variable – in our case, the lapse N – the exponent will remain meromorphic in that variable. One can never generate branch cuts in the exponent by performing partial integrations.

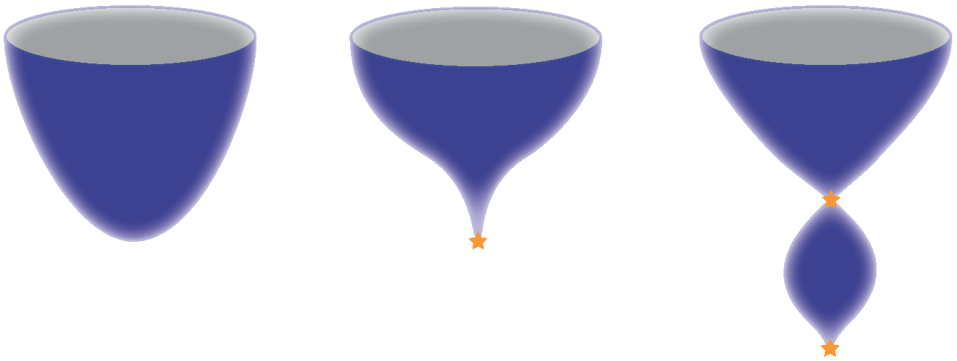


Figure 5.2: The classical background geometries appearing in the no-boundary path integral. *Left:* The regular, complex saddle point geometry. *Middle:* A real Lorentzian off-shell background geometry appearing at $N_-^2 \leq N^2 \leq N_\star^2$, possessing one strong singularity. *Right:* The real Lorentzian geometry appearing at $N^2 > N_\star^2$, possessing two strong singularities.

In our case, something different and inherently infinite dimensional takes place. At off-shell, real values of the lapse N, the background develops strong singularities. By this we mean that once we integrate out the background variable q, for a range of real N the quadratic operator appearing in the action of the perturbations becomes singular. The left panel of Figure 5.2 illustrates the complex but regular geometry which appears as a full saddle point of the path integral. However, at real N (and only at real N) the background geometries – stationary in q but off-shell in N – may exhibit either one strong singularity, if $|N|$ exceeds a critical value N_- or two strong singularities, if it exceeds an even larger value $N_\star$. These cases are illustrated in the middle and right panels of Figure 5.2, respectively.

To discuss what happens near the singularity at $t = 0$, consider momentarily setting the cosmological constant Λ to zero. Then the solution to the second order equation of motion for q is simply $q = q_1 t$, where $0 \leq t \leq 1$ and q_1 denotes the final value. The background line element is $-dt^2 N^2/(q_1 t) + (q_1 t) d\Omega_3^2$, and the background action is given in (5.8) below. For a tensor perturbation mode ϕ, with principal quantum number l on the 3-sphere, it is convenient to rewrite the action (given in general form in (5.19) below) and the associated equation of motion in terms of the canonically normalized variable $\chi(t) \equiv q(t)\phi(t)$ as follows:

$$S^{(2)} = \int_0^1 dt \frac{1}{2N}\left(\dot{\chi}(t)^2 + \frac{\gamma^2 - 1}{4\,t^2}\chi(t)^2\right) - \frac{1}{2N}\left[\frac{\dot{q}}{q}\chi^2\right]_0^1, \qquad -\ddot{\chi}(t) + \frac{\gamma^2 - 1}{4\,t^2}\chi = 0, \quad (5.2)$$

where (for $\Lambda = 0$) $\gamma = \sqrt{1 - 4l(l+2)N^2/q_1^2}$. Notice the perturbation action and equation of motion are both meromorphic in N. However, the two solutions to the equation of motion, $\chi_\pm = t^{\frac{1}{2}(1\pm\gamma)}$, are not, because γ has a branch cut in N. First consider real N satisfying $N^2 < q_1^2/(4l(l+2)$, so γ is real and smaller than one. While both solutions for χ vanish at $t = 0$, only χ_+ has finite action. Therefore, we take this to be the relevant saddle point solution. Normalizing it to obtain $\phi = \phi_1$ at $t = 1$, the classical action for the perturbation is $(\gamma - 1)q_1^2\phi_1^2/(4N)$. We now consider analytically continuing in N to other values in the complex N-plane. Evidently, the action has branch points at $N_\pm = \pm q_1/(2\sqrt{l(l+2)})$. It is convenient to draw the branch cuts to run to $\pm\infty$ respectively. As we increase N along the real axis, we must either pass above or below the branch cut. Passing above, the real part of γ remains positive but the imaginary part becomes negative. Thus, the real part of the semiclassical exponent, $iS_{cl}(N)/\hbar$ is positive on the real N-axis, above the branch cut. Conversely, it is negative below the branch cut. Examining the perturbation solutions for imaginary values of γ one sees they undergo an infinite number of oscillations as t tends to zero. This means they cannot be approximated with any finite sum of Fourier modes in t, and there is no contradiction with the argument given in the opening paragraph of this subsection. Notice also that the non-analyticity in the partially integrated exponent arises precisely at values of N on the branch cuts, where the perturbation action fails to select a particular perturbation mode. It is plausible that this is precisely the edge of the wedge of convergence associated with the higher-dimensional Picard-Lefschetz contour, at the limit where the original path integral ceases to be absolutely convergent and hence cannot be performed iteratively.

The above simplified case exemplifies the mechanism operating in our path integral for de Sitter. In the full situation, with $\Lambda > 0$, in the vicinity of $t = 0$ the perturbations are still described by (5.2) but the power γ appearing in the asymptotic behavior of modes near $t = 0$ has a more intricate structure in N, given in equation (5.25) below. It possesses four square root branch points instead of two, at values $N = \pm N_\pm$ with $N_+ > N_- > 0$, with two finite branch cuts connecting them on the real N-axis (red lines in Figure 5.3). Furthermore, γ develops simple poles at $\pm N_\star$, where $N_\star \equiv \sqrt{N_+ N_-}$, requiring separate analysis (see Appendix 5.D). In Appendix 5.B we prove that in the full problem, just as in our simplified case, the real part of γ is positive for all complex N away from the branch cuts. This means that the mode χ_- has infinite action and must be excluded. Furthermore, as pictured in Figure 5.2, for real N satisfying $|N| > N_\star$, a second Lorentzian singularity forms, at a value $t = t_s$, where $0 < t_s < 1$. The behavior of the perturbations near this second singularity is similar to that near $t = 0$, but this time it turns out that that χ_+ has divergent action, and hence it must be eliminated. Thus, *no* perturbation mode has finite action for real N with $|N| > N_+$. This finding shall be important in our analysis of how

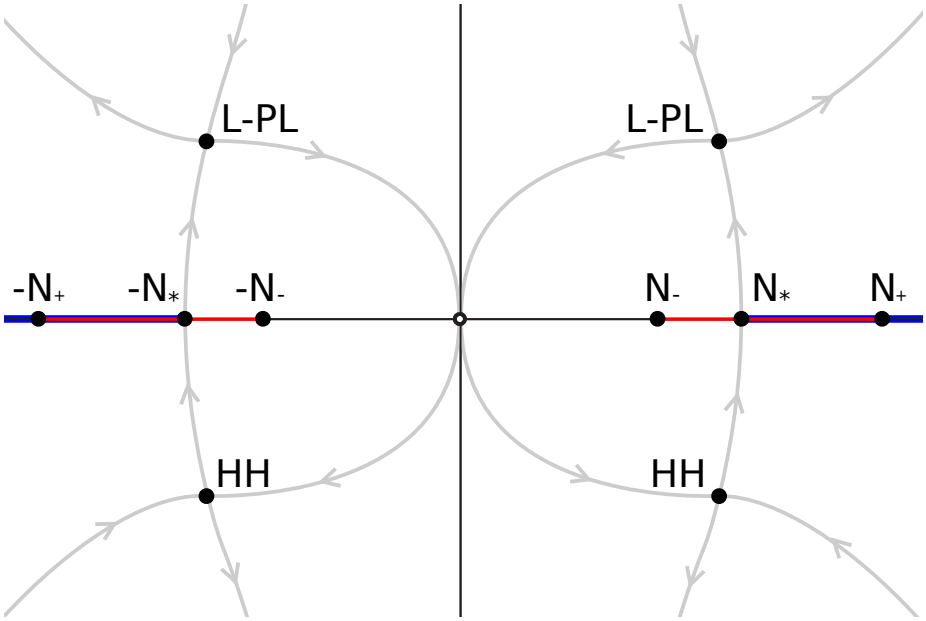


Figure 5.3: The branch cuts (in red) on the real N-axis, for $-N_+ < N < -N_-$ and $N_- < N < N_+$, form impenetrable barriers for Picard-Lefschetz theory. The classical scale factor squared q crosses zero for a second time (as in the right panel of Fig. 5.2) on the blue lines. The Hartle-Hawking and Lorentzian-Picard-Lefschetz saddles are indicated HH and L-PL respectively. The gray lines are the lines of steepest ascent and descent emanating from the four saddle points, with the arrows indicating directions of descent.

to deform Diaz Dorronsoro *et al.*'s proposed contour into one over which the N-integral becomes absolutely convergent.

The subtle, and inherently infinite-dimensional phenomenon just described turns out to explain why it is possible to obtain a positive real term in the semiclassical exponent for the fluctuations, and still remain consistent with Picard-Lefschetz flow away from the branch cuts on the real N-axis, where the effective action for N is still analytic. For example, in our treatment of the background, the original steepest ascent contour from saddle 1 intersects the real N-axis at precisely $N_\star$, the value at which the geometry becomes doubly singular, as indicated in Figure 5.2. As we have described, integrating out the perturbations generates a positive real part of the exponent proportional to ϕ_1^2 on the upper side of the branch cut. Therefore, although our saddle point contribution (5.1) grows exponentially with increasing ϕ_1^2, so does the real part of $iS_{cl}(N)/\hbar$ on the upper side of the N-axis, where the steepest ascent contour from our saddle meets it. There is therefore no inconsistency

with Picard-Lefschetz flow: even as we increase ϕ_1, saddle point 1 remains relevant to the Lorentzian causal propagator.

In Appendix 5.C, we analyse the no-boundary path integral for $\Lambda = 0$ in detail, showing that, in that case, both the Euclidean and the Lorentzian contours make sense as defining contours for N. In the former case, one must take N to run from just above the origin to $+i\infty$. In the latter, it runs over $0^+ < N < \infty$. The result of this analysis, however, is that the introduction of gravity inverts the Euclidean vacuum distribution of the quantum fields, because the background "chooses the wrong Wick rotation," as was explained in chapter 4. This result holds equally for the Euclidean or Lorentzian definitions of the no-boundary path integral.

The situation is more subtle with the contour proposed by Diaz Dorronsoro *et al*, because in this case the background Picard-Lefschetz thimble descending from a Hartle-Hawking saddle intersects the branch cut at $N_\star$. Strictly speaking, Picard-Lefschetz flow fails at this point when we integrate out the perturbations since the effective action for N is no longer analytic in N. However, Cauchy's theorem still holds. By distorting the background steepest descent contour in N to run around the branch cut, and continue along the original contour in the upper half complex N-plane, *before* we even integrate out the perturbations, we can maintain the absolute convergence of the integral, as well as the validity of Cauchy's theorem. At first sight, it appears that there might be two ways to circumnavigate the cut in Figure 5.3, namely on the side nearest to the origin or farthest from it. But here, the second singularity in the background geometry imposes an additional constraint. As mentioned above, since no perturbation mode has finite action for real N satisfying $|N| > N_+$, we cannot go around the branch cut on that side of it. So, as it turns out, when perturbations are included there is only one way to go round the branch cut – on the side nearest the origin. This is the unique choice for distorting the Picard-Lefschetz thimbles associated with the Hartle-Hawking saddles, and it is illustrated in Figure 5.6 below.

The fact that we obtain a unique result even for (what we regard as) an unphysical choice of the defining contour for the lapse and in a situation which is inherently infinite-dimensional is a sign in favor of the mathematical validity of our approach. Nevertheless, one should emphasize that Lorentzian Einstein gravity becomes singular at these off-shell values $|N| \geq N_\star$, with the background geometry developing a second singularity as illustrated in Figure 5.2. Furthermore, the infinite oscillations developed in the interval $|N| > N_-$ might also lead one to doubt the validity of the Einstein action as a correct low energy effective theory, this far off-shell. While the analytic continuation we perform to avoid the branch cuts is, we believe, an entirely natural definition of the off-shell, low energy theory, we cannot rule out the possibility that new degrees of freedom enter and

significantly alter the result. Were this true, however, it would presumably invalidate the no-boundary proposal.

If we distort the Picard-Lefschetz background thimbles as described above, the real part of the exponent $iS_{cl}(N)/\hbar$ becomes more and more positive on the upper side of the branch cuts and, by symmetry, more and more negative on the lower side (see Fig. 5.5). Thus, even as we increase the perturbation amplitude ϕ_1 and the height of the Hartle-Hawking saddle point falls, its steepest descent contour $\mathcal{J}_4$ still runs down to hit the branch cut. The integral around the branch cut yields an additional contribution to the path integral which, again, gives an inverse Gaussian in the perturbations, $e^{+\frac{3}{2\hbar\Lambda}l(l+1)(l+2)\phi_1^2}$ but this time without the $e^{-\frac{12\pi^2}{\hbar\Lambda}}$ suppression factor associated with the Lorentzian-Picard-Lefschetz saddle (5.1).

The conclusion of this analysis is quite striking. Namely, if one wishes to include the Hartle-Hawking saddles in the Lorentzian path integral, then Cauchy's theorem and the choices we are forced to make to obtain an absolutely convergent integral imply there are additional non-perturbative contributions giving unsuppressed fluctuations. In section 5.6 we prove that *no* contour in N can avoid such contributions. As we discuss in the conclusions, this has potentially profound implications for quantum de Sitter spacetime and for inflation.

As a final remark, note that throughout this section we have only treated the perturbations to quadratic order in the action. This is at best a partial representation of the complete theory, and one might wonder whether higher order effects might significantly alter the analyticity properties of the effective action for the lapse near singularities such as those we encountered on the real N-axis. Fortunately, at the semiclassical level we are working at, it is not difficult to study nonlinear backreaction using numerical methods. We do so in Section 5.5, with the conclusion that the basic picture we have obtained using general relativistic linear perturbation theory remains unchanged.

5.1.5 Wider implications

We believe that our results have implications well beyond the no-boundary proposal. For instance, the fact that the no-boundary amplitude is out of control has a bearing on the question of topology change. Smooth, topology changing transitions can be thought of as combinations of no-boundary amplitudes – see the left panel in Fig. 5.4. Our analysis suggests that such transitions must be disallowed. This does not mean that topology change cannot occur, but it indicates that any topology changing transition would have to proceed via a singular, more quantum transition. That such a transition might be feasible

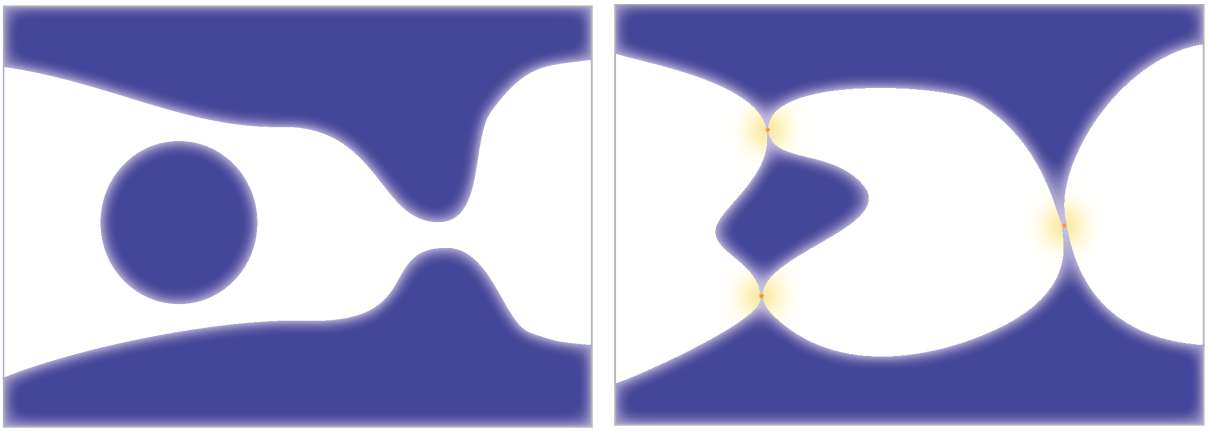


Figure 5.4: Our results imply that a smooth, Λ-mediated topology changing transition is ill-defined. Thus topology change should not be thought of as illustrated in the left panel (where the physical regions of spacetime are blue, space is horizontal and time vertical). Rather, topology change most likely requires passage through a singularity, where massless degrees of freedom play a crucial role in enabling the transition and extensions of the semi-classical methods employed in the present chapter are needed.

is supported by earlier studies indicating that a singular, radiation-dominated bounce appears to be possible, and appears not to suffer from unsuppressed fluctuations [163]. This finding actually resonates, to some extent, with the description of singularity resolution in string theory, where it is typically found that new, massless degrees of freedom appear which are crucial in regularizing topology change. Likewise, it is in accordance with what we know from observations about the standard big bang cosmology, that the early universe was dominated by radiation. Even if one is interested in inflationary scenarios, our findings suggest that a "beginning" with only inflationary potential energy is not allowed. An earlier phase such as a radiation-dominated phase may have been required prior to inflation.

This chapter is structured as follows. In section 5.2, we briefly review the broad physical and mathematical principles of Lorentzian path integral quantum cosmology. We emphasize that the basic definition of the theory involves an integration over a real, nonzero lapse functions N, although it is mathematically convenient to deform that integration contour to an equivalent complex one using Cauchy's theorem, in order to improve the integral's convergence. In section 5.3, we review the path integral for the background de Sitter cosmology, comparing the contour for the Lorentzian propagator with that proposed by Diaz Dorronsoro *et al.*. We show that the latter contour cannot be deformed to the

real N-axis. In section 5.4, we include perturbations, treated in linear perturbation theory. In section 5.4, we include nonlinear backreaction, to all orders for the lowest tensor mode and up to second order for higher modes, showing that it has little effect on our conclusions. In Section 5.5 we prove by simple enumeration a new theorem, that no possible choice of the integration contour for the lapse – whether physically motivated or not – rescues the no-boundary proposal from the problem of unsuppressed perturbations. Finally, we summarize our main findings and briefly comment on implications for quantum de Sitter spacetime and for inflation. In Appendix 5.A, we illustrate some features of higher-dimensional Picard-Lefschetz theory in a simple two-dimensional oscillatory integral. In Appendix 5.B, we prove that the quantity γ governing the perturbations near background singularities of the types discussed above, obeys $\Re[\gamma] > 0$ for all complex N except on the special intervals noted. In Appendix 5.C, we discuss the no-boundary path integral for $\Lambda = 0$, showing that both the Euclidean and the Lorentzian contours make sense as defining contours for N. However, unsuppressed perturbations are obtained in both cases. Finally, in Appendix 5.D, we examine the point $N = N_\star$, showing the precise behavior of the modes and the classical action at that special value.

5.2 Basic physical and mathematical principles

The novelty of our work is to combine, for the first time, two theoretical strands each over three decades old. The first is the work of C. Teitelboim (now C. Bunster) in formally developing a Feynman path integral for quantum gravity [307, 309, 310] based on ideas tracing back to Bryce DeWitt, John Wheeler and Richard Feynman. The second is an area of pure mathematics known as Picard-Lefschetz theory, aimed at the evaluation of oscillatory integrals, in any finite number of dimensions, via contour deformation exploiting Cauchy's theorem and saddle point/steepest descent approximations. Although there was an upsurge of interest in semiclassical quantum gravity effects in the early 1980's, with the computation of scalar and tensor quantum fluctuations in inflation as well as Hartle and Hawking and Vilenkin's proposals for the beginning of the universe, it is surprising to us that Teitelboim's foundational work seems to have attracted only casual reference. It is likewise remarkable that Picard-Lefschetz theory, as a rigorous and highly appropriate mathematical tool, seems to have been altogether overlooked. (For a brief review, with applications to Chern-Simons theory, see [342]. For recent and more general applications to quantum field theory, see [74, 36])

Teitelboim's goal was to develop the theory of quantum geometrodynamics, a program initiated by Wheeler, DeWitt and others. The basic quantity of interest in this program is

the quantum mechanical propagator: the amplitude for obtaining a final three-geometry Σ_1 from a given initial, three-geometry Σ_0, represented by a Hamiltonian path integral over all possible intervening four-geometries. As we explained in chapters 3 and 4, the no-boundary proposal is most naturally formulated in this framework as the amplitude for obtaining Σ_1 when Σ_0 is taken to have zero size. When framed in these terms, we showed that the no-boundary proposal becomes equivalent to Vilenkin's "tunneling" proposal [325, 326], and that the relevant Lorentzian no-boundary propagator, for general relativity with a positive cosmological constant, is a relatively well-defined mathematical object, whereas the Euclidean propagator is not.

5.2.1 Why Lorentzian?

It is worth spelling out why we base our approach on the Lorentzian rather than the Euclidean path integral. Obviously, to do so is more conservative: we take the real-time classical theory as fundamental and try to ensure our quantum theory recovers it's successes in the relevant physical regimes. In fact, the motivation for performing a Wick rotation in gravity appears to have been a misguided belief that Lorentzian path integrals are too oscillatory to be well defined. Hartle and Hawking state in the second paragraph of their paper *"The oscillatory integral in (the usual nonrelativistic path integral propagator) is not well defined but can be made so by rotating the time to imaginary values"* [188]. This statement is incorrect: the integrals appearing in real time (Lorentzian) path integrals are typically conditionally, although not absolutely, convergent. If suitable saddle points exist, as they do very generally, then the path integral can be made absolutely convergent by deforming the integration contour to run along the appropriate steepest descent contour. This is precisely what Picard-Lefschetz theory accomplishes.

In general relativity, rotating the time coordinate to imaginary values is problematic in several ways. The kinetic term for the conformal factor has the wrong sign - the well-known "conformal factor problem", making the Euclidean action unbounded below. Gibbons, Hawking and Perry proposed to remove that divergence by rotating the conformal factor to imaginary values [160]. Unfortunately, this rotation does not respect the boundary conditions in geometrodynamics, which involve the initial and final three-geometry. Instead, in the examples which follow, we will first perform the path integral over the scale factor and perturbations, leaving us with an ordinary integral over the lapse N which we perform using steepest descent methods. When we consider the saddle point solution for the spacetime metric, it is complex and includes a "Euclidean" region. But any complex deformations of the contours are, in our method, chosen by the theory and not the theorist. It is worth mentioning that some of our arguments regarding the improved behavior

of the Lorentzian, as opposed to the Euclidean, path integral were anticipated by earlier discussions, for example by Giddings [161, 162] and particularly by Sorkin [300], although with less general methods.

5.2.2 Wavefunction or propagator?

In contrast, Hartle and Hawking took the Euclidean path integral to be fundamental. This seems to be the basis for their belief that the wavefunction has to be real. Second, they gave a formal argument that the Euclidean path integral satisfies the homogeneous Wheeler-DeWitt equation, and in follow-up papers, *e.g.*, [182] claimed that the Euclidean path integral provides boundary conditions for the wavefunction on the boundary of superspace. Diaz Dorronsoro *et al.* [104] emphasize that their proposed wavefunction is both real and solves the homogeneous Wheeler-DeWitt equation, and they implicitly criticize our Lorentzian propagator because it is not real, and yields $-i$ times a delta functional on the right hand side of the Wheeler-DeWitt equation. However, in chapter 3, we explicitly demonstrated that for Einstein gravity with a positive cosmological constant, the Euclidean path integral is divergent. Our arguments above show it cannot provide boundary values for solutions of the Wheeler-DeWitt equation, as was hoped. Therefore there seems little motivation for insisting that the wavefunction should be real. In fact, as we shall discuss momentarily, a real wavefunction presents problems with recovering local quantum field theory unitarity. In contrast, the Lorentzian formulation provides a natural and mathematically meaningful way to formulate the no-boundary amplitude, as the path integral propagator for obtaining a given final three-geometry starting from an initial three-geometry of zero size, a viewpoint emphasized by Vilenkin [326]. In chapter 3 we showed that the Lorentzian no-boundary propagator is well defined, and furthermore that the dominant saddle point contribution for the background is a regular complex four-geometry with the final three-geometry as its only boundary, exactly the semiclassical picture Vilenkin, and Hartle and Hawking, had anticipated.

If, on the contrary, the no-boundary proposal is reduced to choosing some particular solution of the Wheeler-DeWitt equation, all geometrical justification or uniqueness disappears. As a simple example of this ambiguity, our Lorentzian propagator trivially provides a real (in both senses) solution of the homogeneous Wheeler-DeWitt equation, just by taking its real part. Diaz Dorronsoro *et al.*'s wavefunction, based on a complex contour for the lapse, with an appropriate symmetry, provides another. By taking linear combinations of the two, one obtains an infinite number of "real" wavefunctions with no obvious means to choose between them.

5.2.3 Causality or gauge invariance?

Before proceeding any further, it may be helpful to undertake a short excursion in order to explain why the Lorentzian path integral propagator necessarily *does not* satisfy the homogeneous Wheeler-DeWitt equation, and why this in no sense undermines its utility as a fundamental amplitude in the theory. This was actually understood a long time ago in a beautiful series of papers by C. Teitelboim, emphasizing the tension between gauge invariance and causality. We particularly recommend the brief summary article, Ref. [309].

Schematically, the Lorentzian path integral over all four-geometries bounded by an initial three-geometry Σ_0 and a final three-geometry Σ_1, is given by

$$\langle 1|0\rangle = \int \mathcal{D}N \int \mathcal{D}N^i \int_{\Sigma_0}^{\Sigma_1} \mathcal{D}h^{(3)}_{ij} \mathcal{D}\pi^{(3)ij} e^{\frac{i}{\hbar} S[h^{(3)}_{ij};\pi^{(3)ij};N]} \tag{5.3}$$

where the Lorentzian four-geometry is studied in a $3+1$ split with N being the lapse function, N^i the shift, $h^{(3)}_{ij}$ the 3-metric, $\pi^{(3)ij}$ is its conjugate momentum, and $S = \int_0^1 \mathrm{d}t \int \mathrm{d}x^3 [\pi^{(3)ij} \dot{h}^{(3)}_{ij} - N^i H_i - NH]$ the action for general relativity expressed in first order Hamiltonian form. The path integral is taken over all four-geometry bounded by Σ_0 and Σ_1. Here for simplicity we have neglected the ghosts and BVF formalism needed to ensure general covariance, which were worked out by Teitelboim, Henneaux and others, and generalized to supergravity, in the 1980's [199].

Although the expression (5.3) for the propagator is still formal, the ranges of integration for all but one of the variables to be path-integrated over are fairly clear. At each t and at every spatial point one integrates over all possible real three-metrics and momenta. Likewise one integrates over all real values of the shift in order to enforce the Einstein three-momentum constraint (the G_{ti} Einstein equation) at every spacetime point.

The integration over the lapse N is more subtle. As Teitelboim argued, it is generally possible to choose a gauge in which N depends only on the spatial coordinates. The value of N at some point then controls the total proper time between the initial and final three-geometries, and the path integral measure over N becomes an infinite number of ordinary integrals. The question arises whether one should integrate over all real values of N or only over positive values. Classically, N and $-N$ represent the same spacetime geometry, suggesting that it would be overcounting to include both. Teitelboim argued that integrating N only over one of these choices – positive values, for example – is to be preferred, since it allows one to introduce a primitive notion of causality into the theory, independently of the existence of any classical spacetime. His remarks echo Feynman's earlier discussion, in his first papers on quantum electrodynamics, where he obtained his famous propagator as

the quantized amplitude for a relativistic particle, rather than from any consideration of quantum fields (see Appendix 5.A of Ref. [147]). Because Feynman employed the same, primitive, "world-line" notion of causality, his propagator is still referred to as the "causal" propagator.

In quantum geometrodynamics, it is the causality constraint of integrating only over positive N which enables one to globally distinguish an "in" from an "out" state, and to meaningfully define quantum mechanical transition amplitudes. It also prevents one from considering histories (four-geometries) where the final three-geometry crosses the initial three-geometry creating a region where the two reverse roles. However, there is a tension between diffeomorphism invariance and the primitive causality constraint. Through the Lie derivative (and the corresponding Poisson bracket algebra), timelike diffeomorphisms may be used to push the initial three-surface backward or forward. If the final three-surface is held fixed (as it is, in the propagator), as the initial three-surface approaches it one must exclude diffeomorphisms which would push the initial three-surface ahead of the final one. That is, diffeomorphism invariance becomes retricted to half of the usual space of diffeomorphisms. As Teitelboim puts it [309], the causality constraint $N > 0$ "disrupts the group structure of the four-dimensional diffeomorphisms". Hence, one should not be surprised that the causal propagator is *not* annihilated by the Hamiltonian and, in this sense, is no longer completely invariant under the generator of time-like diffeomorphisms.

One can see this very well in lower-dimensional examples of quantum geometrodynamics, such as the quantized relativistic particle, or the quantized free relativistic string. In these cases, as is well known (see *e.g.* [171]), integrating over positive N is precisely what is required to construct the Feynman propagator, used in perturbative calculations of unitary scattering amplitudes (or, for the string, for defining vacuum states). In these examples, the propagator is formally given by

$$\langle 1|0\rangle = \int_{0^+}^{\infty} dN \langle 1|e^{-iNH/\hbar}|0\rangle = -i\hbar \langle 1|H^{-1}|0\rangle \tag{5.4}$$

where H is the Hamiltonian: $H = p^2 + m^2$ for a free particle or $H = L_0 - 1$ for a free open string. In the Picard-Lefschetz approach, we do not actually need to include the usual $i\epsilon$ to ensure convergence of the integration over N [164]. Note that we define the integral over N to run only over positive real values. This is because in the examples of interest, the integrand possesses singular behavior at small N, so that the integral over N is only defined as its lower, real limit is taken to zero. This singular behavior is no accident. It is generated in passing from the Hamiltonian to the Lagrangian formalism: at $N = 0$ the momenta cannot be expressed in terms of the velocities. In our work, we shall take the Hamiltonian formulation, in which all fields including N are real, to be the fundamental

definition of the theory. The restriction to $N > 0$ (or $N < 0$) is then necessary for a well-defined passage to the Lagrangian formulation.

It follows from (5.4) that the propagator is *not* annihilated by the Hamiltonian constraint, even though the latter is required to vanish on all physical states. Indeed, it follows from (5.4) that $H\langle 1|0\rangle$ equals $-i\hbar$ times a matrix element of the identity operator. For the free relativistic particle in d spacetime dimensions, in the coordinate representation one obtains

$$H_{x_1}\langle x_1|x_0\rangle = (-\hbar^2\Box_{x_1} + m^2)\langle x_1|x_0\rangle = -i\hbar\delta^d(x_1 - x_0), \tag{5.5}$$

the usual equation satisfied by the Feynman propagator $\Delta_F(x_1 - x_0) \equiv \langle x_1|x_0\rangle$.

Within four-dimensional quantum geometrodynamics, one expects something similar: the Hamiltonian applied to the causal propagator yields a delta functional which is zero unless the initial and final three-geometries (the analogs of the initial and final spacetime coordinates of the relativistic particle) are identical. Exactly solvable minisuperspace examples are worked out in detail in [163, 164]. For the no-boundary Lorentzian propagator, the delta functional occurring on the right hand side of the Wheeler-DeWitt equation is nonzero only when the final three-geometry degenerates to a point.

5.2.4 Recovering unitarity

Teitelboim ends his short paper [309] as follows: *"Therefore, it appears that in both gravity and supergravity one is faced with the alternative of preserving either gauge invariance or causality. It is the opinion of this author that one should preserve causality. In the case of positron theory, this turns out to be the correct choice ultimately because only by using the Feynman propagator does one obtain a unitary amplitude.(...) Whether or not a similar situation will arise for the quantized gravitational field remains to be seen."* [309]. We believe the same issue indeed arises, as follows. It is presumably a fundamental constraint on any theory of quantum cosmology that for scales and times much shorter than the Hubble length and time, and much longer than the Planck length and time, we should recover local quantum field theory, along with unitarity of scattering amplitudes in the quantum field theory sense. Consider the Lorentzian path integral propagator between two large three-universes, the final one slightly larger than the initial one, with a local quantum field such as a gravitational wave in a stationary state such as the vacuum, or some fixed number of freely propagating quanta. The path integral will have a classical saddle point solution at positive real N representing an expanding universe with the corresponding quantum field state. Because of the symmetry of the classical theory under $N \to -N$ there will inevitably also be a saddle point representing a contracting universe. If we integrate both

positive and negative values of N, we cannot avoid picking up both saddle points. We thus obtain a superposition of amplitudes for the same quantum field state, within an expanding universe and its time reverse respectively. The inferred Schrödinger wavefunctional for the quantum field will combine field wavefunctionals in which the stationary state is evolved both forward and backward in the "time" as represented by the size of the universe. Such evolution is not unitary. Therefore, integrating over both signs of N seems to be inconsistent, at a basic level, with recovering perturbative quantum field theory unitarity in a description of local processes.

One may say this even more strongly as follows. A real wavefunction, as advocated by Hartle and Hawking and Diaz Dorronsoro *et al.* has no chance of directly recovering unitarity which, at a fundamental level, rests upon quantum mechanical amplitudes being complex. This is particularly obvious for stationary states: the norm of $e^{-iEt/\hbar}$ is preserved but the norm of $\cos(Et/\hbar)$ is not. In the quantum cosmology literature, this problem is sometimes side-stepped by regarding the expanding and contracting parts of the Hartle-Hawking wavefunction as describing two "decoherent histories," which should be studied separately. In effect, to describe an expanding universe, one throws half of the Hartle-Hawking wavefunction away. This seems, at best, uneconomical: if one integrates only over positive N in the first place, and takes the causal propagator to the basic amplitude in the theory, there is no such redundancy and no projection is required.

5.3 The Background

In order to be self-contained we briefly summarize the calculation of the path integral for the background. More details, and references to older literature, see chapter 3.

5.3.1 The propagator for a de Sitter cosmology

For a homogeneous, isotropic background four-geometry, the gauge fixed Feynman propagator for the scale factor of the universe a reduces to

$$G[a_1; a_0] = \int_{0^+}^{\infty} \mathrm{d}N \int_{a(0)=a_0}^{a(1)=a_1} \mathcal{D}a \, e^{iS[a;N]/\hbar} \,, \tag{5.6}$$

where S is the Einstein-Hilbert-Λ action. Throughout this chapter, our focus will be on carefully calculating semiclassical exponents, *i.e.*, contributions to the propagator proportional to $e^{iS_{cl}/\hbar}$ with S_{cl} some classical action. We shall ignore terms in the exponent of

higher order in $\hbar$ associated, for example, with operator ordering ambiguities in the quantum Hamiltonian on superspace (see the discussion above equation (18) in chapter 3). Nor shall we keep track of Jacobian factors associated with redefinitions of variables in the path integral measure. We shall proceed by transforming the action S into a convenient form and then simply adopting the canonical phase space measure for these variables. A more careful treatment would include Jacobian and ordering corrections as well as Fadeev-Popov factors associated with the constraints and gauge fixing conditions.

As outlined in the introduction, we consider a positively curved Friedman-Lemaître-Robertson-Walker (FLRW) universe containing only a positive cosmological constant. It is convenient to write the background metric as follows:

$$\mathrm{d}s^2 = -\bar{N}^2\mathrm{d}t^2 + a^2\mathrm{d}\Omega_3^2 \equiv -\frac{N^2}{q}\mathrm{d}t^2 + q\mathrm{d}\Omega_3^2\,, \tag{5.7}$$

where the first expression is the usual FLRW metric, with $\mathrm{d}\Omega_3^2$ the metric on the unit 3-sphere. The second expression is a convenient rewriting, with $q = a^2$ representing the size modulus for the three-geometry and $N = a\bar{N}$ the redefined lapse. This form has the advantage that the Einstein-Hilbert-Λ action (with Λ the cosmological constant) is quadratic in q [184],

$$S^{(0)} = 2\pi^2 \int_0^1 \left[-\frac{3}{4N}\dot{q}^2 + N(3 - \Lambda q)\right] \mathrm{d}t\,. \tag{5.8}$$

It is convenient to pick a gauge in which N is constant. Since the path integral over q is now Gaussian, it may be performed exactly, with the exponent being given by the classical action. The equation of motion, $\ddot{q} = \frac{2\Lambda}{3}N^2$, is solved by

$$\bar{q}(t) = \frac{\Lambda}{3}N^2t^2 + \left(-\frac{\Lambda}{3}N^2 + q_1\right)t\,, \tag{5.9}$$

with the boundary conditions $q_0 = 0$ and $q_1 = a_1^2$, The corresponding classical action,

$$\bar{S}^{(0)}[q_1; 0; N] = 2\pi^2\left[N^3\frac{\Lambda^2}{36} + N\left(3 - \frac{\Lambda}{2}q_1\right) - \frac{3q_1^2}{4N}\right], \tag{5.10}$$

results in the propagator [1]

$$G[q_1; 0] = \sqrt{\frac{3\pi i}{2\hbar}}\int_{0^+}^{\infty}\frac{\mathrm{d}N}{\sqrt{N}}e^{i\bar{S}^{(0)}[q_1;0;N]/\hbar}\,, \tag{5.11}$$

[1] In evaluating the path integral over q, we include all paths from $q[0] = 0$ to $q[1] = q_1$, including those for which q goes negative. Our methods rely on analyticity, hence we do not impose any barrier forcing q to

where the integration measure $1/\sqrt{N}$ arises from the Gaussian integral over q. As mentioned above, our propagator satisfies

$$\hat{H}G[q_1;0] = -i\hbar\delta(q_1)\,, \tag{5.12}$$

with $\hat{H}$ the Hamiltonian operator (see chapter 3).

In order to discuss more general contours $\mathcal{C}$ for the integral over the lapse N, such as that advocated in [104], we will write the propagator as follows:

$$G_{\mathcal{C}}[q_1;0] = \sqrt{\frac{3\pi i}{2\hbar}}\int_{\mathcal{C}}\frac{\mathrm{d}N}{\sqrt{N}}e^{i\bar{S}^{(0)}[q_1;0;N]/\hbar}\,. \tag{5.13}$$

5.3.2 Picard-Lefschetz theory

The generalized propagator (5.13) is a highly oscillatory integral. We rely on Picard-Lefschetz theory to evaluate it, in a semiclassical approximation – for more details see [342] and chapter 3. One starts by analytically continuing the classical action $\bar{S}^{(0)}$ to the complex N-plane. The exponent is expressed in terms of its real and imaginary parts h and H (which are dimensionless) as

$$e^{iS/\hbar} = e^{h+iH}\,. \tag{5.14}$$

The idea then is to deform the integration contour $\mathcal{C}$ into the complex plane, while keeping its end points fixed, in order to turn the oscillating integral into an absolutely convergent one which, moreover, can then be approximated as a saddle point integral. This is achieved by deforming the integration contour onto a set of steepest descent paths $\mathcal{J}_\sigma$ (also known as Lefschetz thimbles) associated to the saddle points of h (each labeled by σ). In principle, one has to also prove that the "contours at infinity" created by this deformation are negligible. This is usually not difficult (see chapter 3 for examples). Along a steepest descent contour the phase H is constant so that the integral is no longer oscillatory. Paths of steepest descent follow the h-function in a downwards flow until h diverges to minus infinity. Thus an integral over a full thimble always runs between singularities of the

remain positive. Should the details of the theory near $q = 0$ strongly affect the relevant semiclassical saddle point solutions, it seems to us this would necessarily imply sensitivity to the UV completion. In contrast, the saddle point solutions we study here all take the form of locally regular (albeit complex) solutions of the classical Einstein-Λ equations, with modest curvature everywhere. In this case, geometrical higher derivative corrections to the low energy Einstein-Λ effective action are consistently small, at least on shell, and the results are therefore more likely to be reliable.

h-function, of this character. As long as h diverges fast enough, which is typically a modest requirement, the integral along the corresponding Lefschetz thimbles is absolutely convergent since

$$|G_{\mathcal{C}}[q_1;0] \leq \sum_{\sigma} \sqrt{\frac{3\pi}{2\hbar}} \int_{\mathcal{J}_\sigma} \left| \frac{\mathrm{d}N}{\sqrt{N}} \right| e^{h(N)} \,. \tag{5.15}$$

Not all saddle points and steepest descent paths contribute to the contour integral along any particular contour $\mathcal{C}$. A Lefschetz thimble $\mathcal{J}_\sigma$ is relevant if and only if the corresponding steepest *ascent* contour $\mathcal{K}_\sigma$ through the same saddle point σ intersects $\mathcal{C}$. The reason for this is quite intuitive: the original integral is highly oscillatory and thus involves many cancellations. If it is to be replaced by a non-oscillatory integral the integrand must be smaller in magnitude than it is along the original contour. Hence, starting from the original contour we flow down to the Lefschetz thimble. In Fig. 5.1 we illustrate the application of Picard-Lefschetz theory to the Lorentzian contour $\mathcal{C}_1 = (0^+, \infty)$ and two alternate contours, $\mathcal{C}_2^-$ which runs from $N = -\infty$ to $N = +\infty$ just *below* the essential singularity at $N = 0$ and $\mathcal{C}_2^+$ which runs from $N = -\infty$ to $N = +\infty$ just *above* the essential singularity at $N = 0$.

An important subtlety in Picard-Lefschetz theory is what to do when a thimble centred on one saddle point runs down to another saddle point. For example, thimble $\mathcal{J}_4$ – the steepest descent contour from saddle point 4 – coincides with the steepest ascent contour from saddle point 1, $\mathcal{K}_1$. The resolution of this dilemma is to add a small (complex) perturbation to the action which removes the degeneracy – for example one can imagine giving Planck's constant a small complex phase. Such a perturbation breaks the degeneracy between $\mathcal{J}_4$ and $\mathcal{K}_1$ and causes $\mathcal{J}_4$ to just avoid saddle point 1. The perturbation can be taken to zero and in this limit does not affect the result. One sign of the phase causes $\mathcal{J}_4$ to narrowly miss saddle 1 and run off to infinity along the right "side" of thimble $\mathcal{J}_1$. The other sign causes $\mathcal{J}_4$ to continue down the left "side" of thimble $\mathcal{J}_1$. Either are perfectly valid definitions of the completion of thimble 4, and their application will yield exactly the same results. For simplicity, in what follows we shall adopt the second definition for $\mathcal{J}_4$, pictured in Figure 5.13 below, and similarly define the completion of $\mathcal{J}_3$ to run to the origin along the right "side" of $\mathcal{J}_2$.

The causal propagator: integrating over positive lapse

The integration domain $\mathcal{C}_1$ only intersects the line of steepest ascent from one of the four saddle points: $\mathcal{K}_1$ corresponding to saddle point 1 (see the left panel of Fig. 5.1). Observe that $\mathcal{C}_1$ (orange line) can be deformed into the Lefshetz thimble $\mathcal{J}_1$ (orange dashed line) without passing any singularity. Moreover, one can easily show that the additional arcs

around the origin and at complex infinity required to complete the deformed contour, have a vanishing contribution as shown in chapter 3. In the saddle point approximation, the propagator is then given by

$$G_{\mathcal{C}_1}[q_1;0] = c_1 e^{-\frac{12\pi^2}{\hbar\Lambda} - i4\pi^2\sqrt{\frac{\Lambda}{3\hbar^2}}(q_1 - 3/\Lambda)^{3/2}}\,, \tag{5.16}$$

where the constant c_1 includes the functional determinants and prefactors. In principle, for small $\hbar$ it can be expressed as a series in $\hbar$. The weighting $e^{-\frac{12\pi^2}{\hbar\Lambda}}$ is the inverse of the famous Hartle-Hawking result $e^{\frac{12\pi^2}{\hbar\Lambda}}$ and agrees with Vilenkin's tunneling proposal [325], as well as with arguments by Sorkin [300]. Evaluation of the integration domain $(-\infty, 0)$ leads to the Lefschetz thimble $\mathcal{J}_2$ giving an equivalent (but complex conjugate) result.

Solutions of the homogeneous Wheeler-DeWitt equation

In an attempt to recover the Hartle-Hawking result, Diaz Dorronsoro *et al.* [104] have instead proposed the integration domain $\mathcal{C}_2^-$. We have already explained why this contour cannot be claimed to be Lorentzian. Nevertheless, let us continue to analyze it. From Figure 5.1 one sees that $\mathcal{C}_2^-$ is intersected by steepest ascent lines from all four saddles. From left to right, the contour intersects $\mathcal{K}_2, \mathcal{K}_3, \mathcal{K}_4$, and $\mathcal{K}_1$. We thus conclude that all four saddles contribute to the path integral. The corresponding deformed contour is indicated by the dashed orange line in the figure. Thus the path integral can be rewritten as a sum over all four thimbles,

$$\begin{aligned} G_{\mathcal{C}_2^-}[q_1;0] \approx & |c_1|\, e^{\frac{12\pi^2}{\hbar\Lambda}} \cos\left(\frac{4\pi^2}{\hbar}\sqrt{\frac{\Lambda}{3}}(q_1 - 3/\Lambda)^{3/2} + \varphi_1\right) \\ & + |c_2|\, e^{-\frac{12\pi^2}{\hbar\Lambda}} \cos\left(\frac{4\pi^2}{\hbar}\sqrt{\frac{\Lambda}{3}}(q_1 - 3/\Lambda)^{3/2} + \varphi_2\right)\,, \end{aligned} \tag{5.17}$$

where $c_1 = |c_1|e^{i\varphi_1}$, $c_2 = |c_2|e^{i\varphi_2}$ are coefficients to be expanded in powers of $\hbar$. $G_{\mathcal{C}_2^-}[q_1;0]$ is real because the contributions from $\mathcal{J}_2, \mathcal{J}_3$ are complex conjugates of those from $\mathcal{J}_1, \mathcal{J}_4$.

The lower saddle points, represented by the first term, dominate in the semiclassical expansion. Thus Diaz Dorronsoro *et al.*'s contour recovers the Hartle-Hawking result at leading order in the exponential factor $e^{\frac{12\pi^2}{\hbar\Lambda}}$. However, it also generates the second term in (5.17) which represents a non-perturbative (and exponentially small) correction. This is a minor correction for the background, but it will become problematic when we consider the perturbations.

As we mentioned in the Introduction, there is another way of getting a real solution of the Wheeler-DeWitt equation, from a truly Lorentzian contour, by integrating over purely real $-\infty < N < 0^-$ and $0^+ < N < +\infty$. This combination, representing the real part of our Lorentzian propagator, is equivalent to the continuous contour $\mathcal{C}_2^+$ which avoids the essential singularity at $N = 0$ by passing above it (see the right panel in Fig. 5.1), because the small semicircle above the origin gives a vanishing contribution in the limit as 0^- and 0^+ tend to 0. Note also that $\mathcal{C}_2^+$ only intersects the steepest ascent contours $\mathcal{K}_1$ and $\mathcal{K}_2$, from saddle points 1 and 2. It follows that the integral along $\mathcal{C}_2^+$ equals the sum of the Lefschetz thimbles $\mathcal{J}_2$ and $\mathcal{J}_1$, taken with appropriate signs. Hence the path integral along $\mathcal{C}_2^+$ is twice the real part of the path integral along $\mathcal{C}_1$. In the saddle point approximation, it is given by

$$G_{\mathcal{C}_2^+}[q_1;0] = 2\mathrm{Re}[G_{\mathcal{C}_1}[q_1;0]] \approx 2|c_2|e^{-\frac{12\pi^2}{\hbar\Lambda}}\cos\left(\frac{4\pi^2}{\hbar}\sqrt{\frac{\Lambda}{3}}(q_1-\frac{3}{\Lambda})^{3/2}+\varphi_2\right). \tag{5.18}$$

As explained above, $G_{\mathcal{C}_2^-}[q_1;0]$ and $G_{\mathcal{C}_2^+}[q_1;0]$ provide two independent, real solutions of the Wheeler-DeWitt equation. One might have hoped that one could subtract $G_{\mathcal{C}_2^+}[q_1;0]$ from $G_{\mathcal{C}_2^-}[q_1;0]$ in order to remove the two upper saddle points entirely. Unfortunately, this does not work, because the *entire* Lefschetz thimbles $\mathcal{J}_1$ and $\mathcal{J}_2$ contribute to $G_{\mathcal{C}_2^+}[q_1;0]$, whereas only the two outer "sides" of these thimbles contribute to $G_{\mathcal{C}_2^-}[q_1;0]$. Furthermore, since the thimbles are not perfectly symmetrical, their "outer" and "inner" sides are not identical. This means there is no possible way to cancel the contributions of the upper two thimbles and hence to recover Hartle and Hawking's result.

5.4 Perturbations

We now turn our attention to the perturbations, treated in general relativistic linear perturbation theory. In chapter 4 we showed that the no-boundary causal propagator generates an inverse Gaussian distribution for the perturbations, meaning large perturbations are favored. Here we shall review and extend this treatment to the wavefunction and lapse contour proposed by Diaz Dorronsoro *et al.*, showing that it too includes unsuppressed perturbations. Furthermore, we identify a new and larger source of unsuppressed perturbations coming from a branch cut which the Picard-Lefschetz thimble through the Hartle-Hawking saddles encounters. These results strengthen and generalize our result, allowing us to prove that no possible redefinition of the lapse contour can rescue the no-boundary proposal.

The second order action for a linearized (tensor) perturbation ϕ_l with principal quantum number l is given in terms of the background squared scale factor $q(t)$ as

$$\begin{aligned} S^{(2)}[q,\phi;N] &= \frac{1}{2}\int_0^1 \left[q^2\frac{\dot{\phi}_l^2}{N} - Nl(l+2)\phi_l^2\right]\mathrm{d}t \\ &= \frac{1}{2}\int_0^1 \left[\frac{\dot{\chi}_l^2}{N} + N\left(\frac{\ddot{q}}{q} - \frac{l(l+2)}{q^2}\right)\chi_l^2\right]\mathrm{d}t - \frac{1}{2N}\left[\frac{\dot{q}}{q}\chi_l^2\right]_0^1, \end{aligned} \tag{5.19}$$

where we have re-expressed the dimensionless tensor metric perturbation ϕ_l in terms of the canonically normalized field $\chi_l = q\,\phi_l$. Note that we have orthonormalized the modes on the unit sphere (thus no prefactor of $2\pi^2$ appears in the action). As explained in the introduction, to avoid needless complexity in the equations we only consider a single mode. It is straightforward to amend all the formulae we derive by replacing $l(l+1)(l+2)\phi_l^2$ with $\sum_{lmn} l(l+1)(l+2)\phi_{lmn}^2$ where the $\phi_{1,lmn}$ are the expansion coefficients expressing in the final tensor perturbation in terms of orthonormal tensor spherical harmonics on the three sphere, with quantum numbers l, m, n [158]. Since the treatment of each harmonic proceeds identically we will not write out this sum – one may always think of setting all Fourier coefficients, bar one, to zero on the final three-geometry. For ease of notation, where there is no danger of confusion, we will also usually drop the subscript l. Note that the perturbation of the lapse N is non-dynamical in the absence of matter and may be set to zero.

If we neglect the backreaction of the linear perturbations on the background, such as is reasonable for small final amplitude ϕ_1, then we can evaluate the path integral first for q and then for ϕ, using the classical solution for the background q in the action (5.19) for ϕ. To integrate out the perturbations, we again just find the classical solution and use this to evaluate the classical action. The total semiclassical exponent is then given by $S^{(0)}[q_1;N]+S^{(2)}[q_1,\phi_1;N]$. We perform the final ordinary integral over N using saddle point methods. We shall not calculate any functional determinants in this chapter, although this is perfectly possible. These should not alter any conclusions about the Picard-Lefschetz flow, nor the final semiclassical exponent, in any regime where the semiclassical expansion is valid.

The no-boundary path integral on a contour $\mathcal{C}$ is then given, in this leading semiclassical approximation, by

$$G_{\mathcal{C}}[q_1,\phi_1;0] \propto \int_{\mathcal{C}} \frac{\mathrm{d}N}{\sqrt{D(N,q_1,\hbar)}} e^{i\bar{S}^{(0)}[q_1;N]/\hbar + i\bar{S}^{(2)}[q_1,\phi_1;N]/\hbar}, \tag{5.20}$$

where $\bar{S}^{(0)}[q_1; N]$ is the classical action for the background solution $\bar{q}$ satisfying the final boundary condition $q(1) = q_1$ and the initial, no-boundary condition $q(0) = 0$ (see equation (5.10)). Likewise, $\bar{S}^{(2)}[q_1, \phi_1; N]$ is the classical action for the perturbation, in the background $\bar{q}$, satisfying $\phi(1) = \phi_1$ as well as a second condition we shall define shortly. The quantity $D(N, q_1, \hbar)$ is the functional determinant which is in principle calculable in terms of the classical modes and as a series expansion in $\hbar$ (for a recent review see, *e.g.*, Ref. [118]). However, in this chapter we shall focus on the semiclassical exponent, and shall not consider the functional determinant any further.

5.4.1 Semiclassical path integral over the perturbations

In this section, we shall perform the path integral over the perturbations in the leading semiclassical approximation. That is, we shall fix the perturbation amplitude on the final three-geometry, ϕ_1 and perform the path integral by the saddle point method, *i.e.*, by solving the equations of motion and computing the classical action. The boundary condition on the perturbations at $t = 0$ is delicate because the background geometry is sufficiently singular for a range of real values of the lapse, that the perturbations obey a singular equation of motion. We shall find that, nevertheless, for generic complex N, the criterion of finite classical action selects a unique perturbation mode.

At fixed N, the classical equation for χ following from (5.19) is

$$\ddot{\chi} = \left(\frac{\ddot{\bar{q}}}{\bar{q}} - \frac{N^2 l(l+2)}{\bar{q}^2} \right) \chi \,. \tag{5.21}$$

Near $t = 0$, this becomes

$$\ddot{\chi} \approx -\frac{N^2 l(l+2)}{(q_1 - \Lambda N^2/3)^2} \frac{\chi}{t^2} \equiv \frac{\gamma^2 - 1}{4} \frac{\chi}{t^2} \,, \tag{5.22}$$

from which we see $\chi \sim t^{\frac{1}{2}(1 \pm \gamma)}$, as $t \to 0$. Notice that the equation of motion for χ is singular and this results in some unusual properties of the perturbations, as we explain below.

For small real N, we take γ to be real and positive. Provided N is real and smaller in magnitude than a particular value N_-, then both solutions for χ are monotonic in t and both vanish at $t = 0$. However, only one of them has finite action so it is natural to select that one as the saddle point solution. For real N larger in magnitude than N_- but smaller than another, larger value, N_+, γ is imaginary and the solutions oscillate an infinite

number of times as they approach $t = 0$. In fact, both solutions have a finite regularized action, so the finite action criterion becomes ambiguous for N in this range. Increasing the magnitude of N beyond N_+, while keeping N real, we see that γ becomes real once again. However, as we explain shortly, in this latter regime, there are *no* finite action classical solutions.

The two critical values are given by

$$N_- = \frac{3}{\Lambda}\sqrt{2l(l+2) + q_1\Lambda/3 - 2\sqrt{l(l+2)(l(l+2)+q_1\Lambda/3)}}\,, \tag{5.23}$$

$$N_+ = \frac{3}{\Lambda}\sqrt{2l(l+2) + q_1\Lambda/3 + 2\sqrt{l(l+2)(l(l+2)+q_1\Lambda/3)}}\,, \tag{5.24}$$

with geometric mean $N_\star \equiv \sqrt{N_+N_-} = \sqrt{3q_1/\Lambda}$. It follows that we can take

$$\gamma = \frac{\sqrt{(N_-^2 - N^2)(N_+^2 - N^2)}}{(N_\star^2 - N^2)}, \tag{5.25}$$

defined to be real and positive for small real N and for other values of N by analytic continuation. The branch cuts needed to define the square roots are conveniently placed along the real intervals $-N_+ < N < -N_-$ and $N_- < N < N_+$. In Appendix 5.B, we prove that $\Re[\gamma]$ is positive for all complex N away from these cuts. On the upper side of the cuts, γ is negative imaginary and on the lower sides it is positive imaginary. Away from the cuts, as is evident from (5.19), the action integral converges at $t = 0$ only for the mode behaving as $t^{\frac{1}{2}(1+\gamma)}$ as $t \to 0$. The complete solution of (5.21) with this small t behavior is

$$\chi(t) = \bar{q}(t)^{\frac{1}{2}}\left(\frac{t}{3q_1 + (t-1)N^2\Lambda}\right)^{\frac{\gamma}{2}}\left((3q_1 - \Lambda N^2)(1+\gamma) + 2\Lambda N^2 t\right)\,, \tag{5.26}$$

and the corresponding, correctly normalized classical solution is

$$\phi(t) = \phi_1 \frac{\chi(t)}{\bar{q}(t)}\frac{q_1}{\chi(1)}. \tag{5.27}$$

This solution allows us to calculate the classical action from (5.19). With an integration by parts and using the equations of motion, we find

$$S^{(2)}[q_1, \phi_1; N] = \left[\bar{q}^2\frac{\bar{\phi}\dot{\phi}}{2N}\right]_0^1 = \frac{l(l+2)q_1\phi_1^2}{4N(3l(l+2)+q_1\Lambda)}\left(-3q_1 - N^2\Lambda + \gamma(N_\star^2 - N^2)\right), \tag{5.28}$$

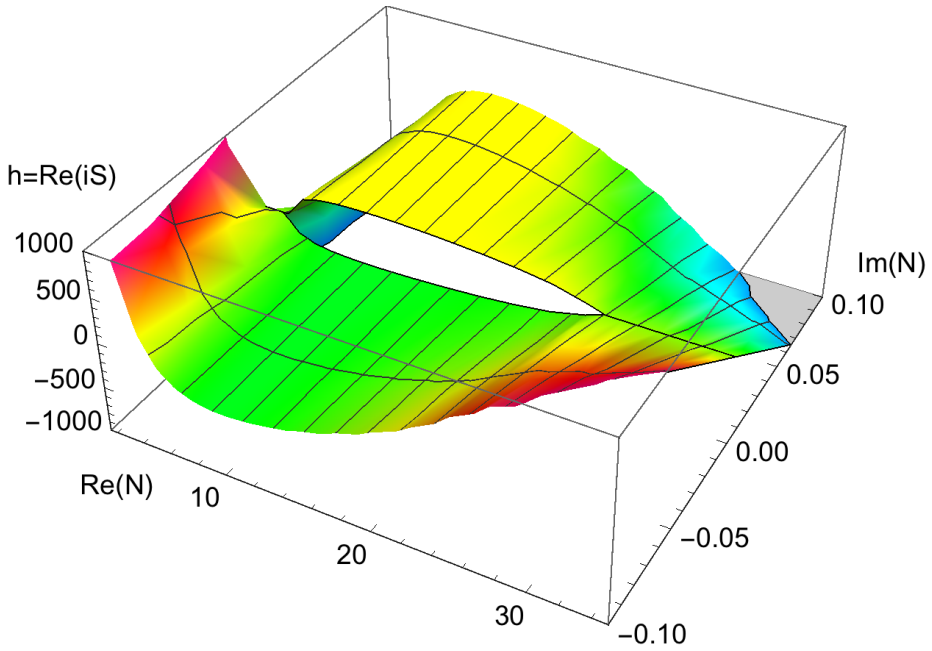


Figure 5.5: The function $h = Re(iS/\hbar)$ around a branch cut, in units where $\hbar = 1$ and for the parameters $\Lambda = 3, q_1 = 101, l = 10, \phi_1 = 1$. At the cut, the h-function reaches its maximum at $N_\star = 10$ coming from the upper half plane, and its minimum also at $N_\star$, though approaching the cut from below.

which is real where γ is real, but gains a negative or positive imaginary part (meaning that the semiclassical exponent $iS/\hbar$ gains a positive or negative real part) as N approaches the real axis from above or below the branch cuts. This behavior is illustrated in Figure 5.5.

There is one additional important consideration: for real N, the background solution for the scale factor (5.9) is real and quadratic in t. For $N > N_\star$ (or $N < -N_\star$) the background solution starts at $q = 0$, then turns negative before crossing $q = 0$ a second time, at $t_s = 1 - \frac{3q_1}{\Lambda N^2}$, to eventually reach q_1 at $t = 1$. Thus there is a second singularity in these real but off-shell-in-N background geometries, as sketched in the right panel of Fig. 5.2. It is obvious from (5.26) that if χ behaves as $t^{\frac{1}{2}(1\pm\gamma)}$ near $t = 0$, then it behaves as $(t_s - t)^{\frac{1}{2}(1\mp\gamma)}$ near $t = t_s$. Thus, for real γ and $N > N_\star$ then if the action integral converges at $t = 0$, it diverges at $t = t_s$, and vice versa. We conclude that for $N > N_+$ or $N < -N_+$ *no* solution of the perturbation equations of motion has finite action. Hence, in performing the integration over N in the last step (5.20) of our calculation, however we deform the contour $\mathcal{C}$, we cannot allow it to cross the real N-axis for real N beyond the outer ends of

the two branch cuts.

Finally, notice that at large $|N|$ in the complex N-plane, the background action $\bar{S}^{(0)} \sim N^3$ – dominates over the perturbation action $\bar{S}^{(2)} \sim N$. The same holds in the small $|N|$ limit, where both the background and the perturbation diverge like $1/N$ (the background and the perturbation action have opposite sign). As a consequence the asymptotic regions of convergence are preserved when we add linearized tensor perturbations.

5.4.2 Integrating over the lapse N

Having determined the classical action for the background (5.10) and for the perturbations (5.28), we are now ready to evaluate the final integral over the lapse N, given in (5.20). We have already explained the principles in the Introduction: here we shall give the details.

The saddle point contribution

In the first approximation, we ignore backreaction from the perturbations on the background and simply evaluate the combined classical action ((5.10) plus (5.28)) at the relevant saddle points for the background. For simplicity, in this section we shall only discuss the saddles in the right half-plane: those in the left half-plane are simply related by symmetry. Assuming the radius of the final three-universe is greater than the de Sitter radius $\sqrt{3/\Lambda}$, the two classical saddles for the background are given by

$$N_s^{\pm} = \frac{3}{\Lambda}\left[\sqrt{\frac{\Lambda}{3}q_1 - 1} \pm i\right]. \tag{5.29}$$

At this saddle points, the parameter γ defined in Section 5.4.1 is precisely equal to $l+1$, meaning that the tensor modes ϕ_l behave as $t^{l/2}$ near the singularity, which means they are regular there. In Appendix 5.C we describe the relevant change of variables which exhibits this property.

The values of the classical action at the upper and lower saddle points respectively are

$$\bar{S}^{(2)}(N_s^{\pm}) = \mp i\frac{\phi_1^2 q_1}{2}\frac{l(l+2)}{l+1 \pm i\sqrt{q_1\Lambda/3-1}}. \tag{5.30}$$

There are two simplifying regimes. If the wavelength on the final three-geometry, $\sim \sqrt{q_1}l^{-1}$ is well within the Hubble radius $\sqrt{3/\Lambda}$, we obtain

$$i\frac{\bar{S}^{(2)}(N_s^{\pm})}{\hbar} \approx \pm\frac{\phi_1^2 q_1}{2\hbar}l, \quad l \gg \sqrt{\frac{\Lambda q_1}{3}}, \tag{5.31}$$

a result which is independent of Λ and which agrees with the result of Appendix 5.C.

In the opposite limit, we obtain the result for the "frozen" modes in the expanding de Sitter spacetime, which have passed out of the de Sitter Hubble radius and ceased to evolve. In this case, we obtain

$$i\frac{\bar{S}^{(2)}(N_s^{\pm})}{\hbar} \approx \pm\frac{3}{2\Lambda}l(l+1)(l+2)\phi_1^2 - i\sqrt{\frac{3q_1}{4\Lambda}}l(l+2)\phi_1^2, \quad \sqrt{\frac{\Lambda q_1}{3}} \gg l. \tag{5.32}$$

For the lower, Hartle-Hawking saddle point, the real part of the exponent exhibits the familiar scale-invariant inflationary power spectrum $\sim l^{-3}$ at large l, so that the real-space variance of the tensor modes is logarithmically divergent. Unfortunately, as we explained in the Introduction, the upper saddle is also relevant and it leads to an inverse Gaussian distribution meaning that the tensor modes are out of control.

In view of this unsettling result, one should ask whether all the assumptions which went into calculating it are really valid. In particular, can we really trust it for large ϕ_1, where the contribution of the upper saddle point outweighs the corresponding lower one? The calculation assumed linear perturbation theory, which requires that the perturbation amplitude is small throughout, *i.e.*, $|\phi(t)| \ll 1$ for all $t \in [0,1]$. However, there is a strong redshifting effect in a de Sitter background, and the amplitude of linearized tensor modes decreases inversely with the scale factor while it is inside the Hubble radius. Thus a mode which has just frozen at some large value of q_1 with amplitude ϕ_1 has a much greater amplitude $\sim \phi_1\sqrt{q_1}$ when it is followed back to the "throat" of de Sitter spacetime. The condition that the mode has just frozen at q_1 reads $l \sim \sqrt{q_1\Lambda}$. The condition that the perturbation contribution to the final semiclassical exponent outweighs the background contribution is that $l^3\phi_1^2$ exceeds unity (assuming l is large). For this to be true, the initial amplitude $\phi_1\sqrt{q_1}$ must exceed $(l\Lambda)^{-\frac{1}{2}}$. This is possible, while maintaining the validity of linear theory at all times, if the frozen mode number l exceeds Λ^{-1}. That requires that the de Sitter spacetime has undergone expansion by a factor $\Lambda^{-\frac{3}{2}}$, *i.e.*, that q_1 exceeds this factor, which is a rather modest condition. Our conclusion is that it is perfectly possible to have the perturbations dominate in the semiclassical exponent, while remaining consistent with linear perturbation theory throughout the evolution of the perturbation modes. This is confirmed by the numerical calculations we shall report in Section 5.5. In fact, those calculations show that nonlinear effects further enhance the discontinuity in the effective action across the real N-axis, created by integrating out the perturbations .

The branch cut contribution

Before turning to the detailed implications of the various contours of integration discussed in section 5.3, we study the branch cut. We will specialize to $\mathrm{Re}(N) > 0$, but analogous considerations apply for $\mathrm{Re}(N) < 0$. As discussed above, the branch cut represents an impenetrable barrier to the integral over the lapse, since traversing it would mean running into regions where the perturbations are not well defined. Moreover, as shown in Fig. 5.3, some of the Lefschetz thimbles intersect the branch cuts. This means that when evaluating the path integral, in some cases we are forced to distort the contour of integration around the branch cut. We thus need to know the contribution of the branch cut to the integral over the lapse. Note that because the perturbative action is infinite on the real N-axis outside of the cut (, *i.e.*, for $N > N_+$), we are forced to deform the contour to pass on the inside of the cut.

Let us focus on a mode that has just frozen, *i.e.*, a mode for which $l \sim \sqrt{q_1 \Lambda}$, and which has a large amplitude (but within the limits of perturbation theory) , *i.e.*, $\phi_1 \sim l^{-\frac{1}{2}}$. We work in the limit of large final scale factor. Similar calculations can be performed for other wavenumbers and amplitudes.

We approximate the integral with the integration contour going around the branch cut in a clockwise direction. As we saw above, the h-function is much higher above the cut than below, hence it is sufficient to consider the integral running just above the branch cut on the real N-axis, see also Fig. 5.5. The maximum of the h-function occurs at $N_\star^+$, which is the location of the saddle point of the perturbative action (5.28) evaluated on the upper side of the branch cut (note that the h-function of the background action is zero on the real N-axis). The total exponent $iS/\hbar$ and its first two derivatives evaluated at $N_\star^+$, keeping the leading real and imaginary terms in the limit of large q_1, are given by

$$\frac{i}{\hbar} S(N_\star^+) = -i\sqrt{\frac{4\Lambda q_1^3}{3\hbar^2}} + \frac{3}{2\hbar\Lambda}(l(l+2))^{3/2}\phi_1^2\,, \tag{5.33}$$

$$\frac{i}{\hbar} S_{,N}(N_\star^+) = 3\frac{i}{\hbar}\,, \tag{5.34}$$

$$\frac{i}{\hbar} S_{,NN}(N_\star^+) = -\frac{i}{\hbar}\sqrt{\frac{\Lambda}{12 q_1}}\, l(l+2)\phi_1^2 - \frac{\Lambda}{6\hbar}(l(l+2))^{1/2}\phi_1^2\,. \tag{5.35}$$

At large scale factor, the phase in the last expression can be dropped. We can approximate the integral running just above the cut from left to right as an integral over real x along

the path $N_\star^+ + x$,

$$e^{\frac{i}{\hbar}S(N_\star^+)}\int_{-\infty}^{0}\mathrm{d}x e^{-\frac{\Lambda}{12\hbar}(l(l+2))^{1/2}\phi_1^2 x^2} = \frac{\sqrt{3\pi\hbar}}{\sqrt{\Lambda}(l(l+2))^{1/4}\phi_1}e^{\frac{i}{\hbar}S(N_\star^+)} \sim \sqrt{\frac{\hbar}{\Lambda}}e^{-i\sqrt{\frac{4\Lambda q_1^3}{3\hbar^2}}+\frac{3}{2\hbar\Lambda}(l(l+2))^{3/2}\phi_1^2}\,. \tag{5.36}$$

From the point $N_\star$ the path of steepest descent runs straight up, and we may check that including this contribution does not significantly affect the integral arising from the cut itself. For this case we add an integration along a path $N_\star^+ + iy$ with positive real y, obtaining essentially the same result as above,

$$e^{\frac{i}{\hbar}S(N_\star^+)}\left[\int_{-\infty}^{0}\mathrm{d}x e^{-\frac{\Lambda}{12\hbar}(l(l+2))^{1/2}\phi_1^2 x^2} + i\int_0^{+\infty}\mathrm{d}y e^{-\frac{3}{\hbar}y}\right] = \left(\frac{\sqrt{3\pi\hbar}}{\sqrt{\Lambda}(l(l+2))^{1/4}\phi_1}+\frac{i\hbar}{3}\right)e^{\frac{i}{\hbar}S(N_\star^+)} \sim \sqrt{\frac{\hbar}{\Lambda}}e^{-i\sqrt{\frac{4\Lambda q_1^3}{3\hbar^2}}+\frac{3}{2\hbar\Lambda}(l(l+2))^{3/2}\phi_1^2}\,. \tag{5.37}$$

Hence the integral around the branch cut yields unsuppressed fluctuations, with weighting

$$\sqrt{\frac{\hbar}{\Lambda}}e^{+\frac{3}{2\hbar\Lambda}(l(l+2))^{3/2}\phi_1^2}\,. \tag{5.38}$$

Integrating N over positive values

For the no-boundary proposal, defined in terms of the propagator, we integrate over positive lapse $\mathcal{C}_1 = (0^+, \infty)$. According to Picard-Lefschetz theory, we should distort the contour integral over N to the relevant Picard-Lefschetz thimble $\mathcal{J}_1$ at the first stage of the calculation. We then need the solution for the classical background and perturbations, given in previous sections. Using the background action (5.10) and the perturbation action (5.32), we obtain

$$G_{\mathcal{C}_1}[q_1, \phi_1; 0] \propto e^{iR(q_1,\phi_1)/\hbar}e^{-\frac{12\pi^2}{\hbar\Lambda}+\frac{3}{2\hbar\Lambda}l(l+1)(l+2)\phi_1^2}\,, \tag{5.39}$$

with the phase given by the real part of the classical action

$$R(q_1, \phi_1) = -4\pi^2\sqrt{\frac{\Lambda}{3}}\left(q_1 - \frac{3}{\Lambda}\right)^{3/2} - \sqrt{\frac{3q_1}{4\Lambda}}l(l+2)\phi_1^2\,. \tag{5.40}$$

This is the result described in chapter 4, where the background is suppressed as $e^{-12\pi^2/(\hbar\Lambda)}$, but the fluctuations are unsuppressed, so they are out of control.

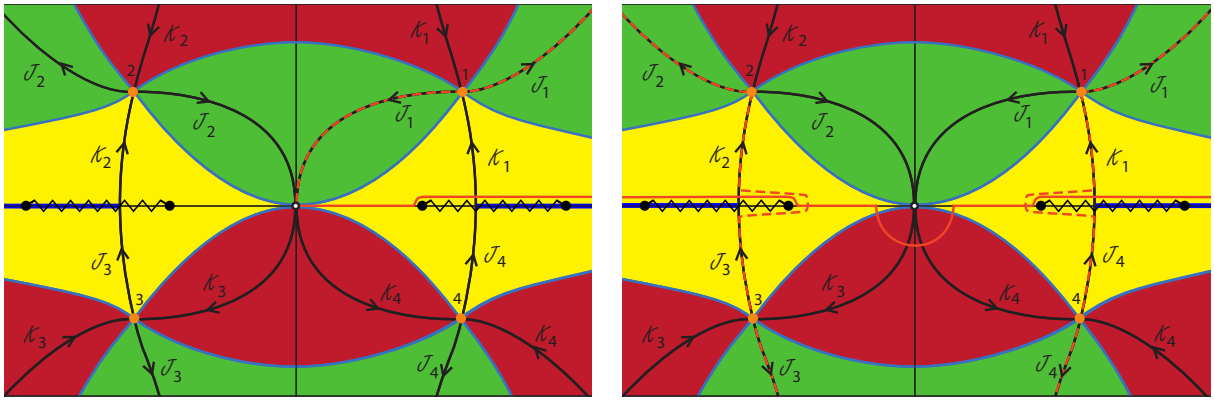


Figure 5.6: Picard-Lefschetz theory for a Λ-dominated universe with gravitational waves. The solid orange and dashed orange lines are the original and deformed integration contours respectively, while the zigzag lines denote the branch cuts. The lines denoted by $\mathcal{J}_i$ are lines of steepest descent and the lines denoted by $\mathcal{K}_i$ are lines of steepest ascent. Left panel: the integration path for the Lorenztian propagator, deformed to run above the cut. Right panel: the integration domain prescribed by Diaz Dorronsoro *et al.* [104] with the original integration domain above the branch cuts. Note that the contour must be deformed to partially encircle the branch cuts in order to reach the Lefschetz thimbles.

Integrating N from $-\infty$ to $+\infty$

The integration domain $\mathcal{C}_2^+$, deformed to pass above the essential singularity at the origin $N = 0$, gives twice the real part of the half line contour given (5.39) and (5.40) above. Its implications are immediately obtained from Eq. (5.39) above. We do not need to discuss this contour further.

Diaz Dorronsoro *et al.* propose to use the integration domain $\mathcal{C}_2^-$, passing below the essential singularity at the origin $N = 0$. As discussed above, asymptotically the contour must be deformed to run above the real N-axis to yield a convergent path integral, hence we will adopt this definition here. Picard-Lefschetz theory implies the relevance of all four saddle points, since the integration contour is intersected by all lines of steepest ascent (see the right panel in Fig. 5.6). For the background, the lower saddle points dominate over the upper saddle points leading to the Hartle-Hawking result. Obtaining this result for the background appears to have been the main goal in choosing this complex contour. When including the perturbation action $\bar{S}^{(2)}$, we need to deform the contour to avoid the branch cut on the side of the origin, as shown in the figure. Up to the first sub-leading order in

the saddle point approximation, the path integral evaluates to

$$G_{\mathcal{C}_2^-}[q_1,\phi_1;0] \approx 2\Re\left[e^{i\frac{R(q_1,\phi_1)}{\hbar}}\left(C_4 e^{\frac{12\pi^2}{\hbar\Lambda}-\frac{3}{2\hbar\Lambda}l(l+1)(l+2)\phi_1^2}+(C_b+C_1e^{-\frac{12\pi^2}{\hbar\Lambda}})e^{+\frac{3}{2\hbar\Lambda}l(l+1)(l+2)\phi_1^2}\right)\right]. \tag{5.41}$$

The functional determinants corresponding to the Hartle-Hawking and Picard-Lefschetz saddle points are C_4 and C_1. The term C_b is the prefactor of the integral along the branch cut. The overall phase is again given by the real part of the classical action (see equation (5.40)).

The lower saddle points alone would have given the standard Bunch-Davies vacuum state. However, the branch cut and the upper saddle points lead to non-perturbative corrections, suppressed by one and two powers of the nonperturbative factor $e^{-12\pi^2/(\hbar\Lambda)}$ respectively. However, as both the perturbation amplitude ϕ_1 and the wavenumber l increase, as discussed below Eq. (5.32), these corrections can dominate. The consequence is that the Bunch-Davies vacuum obtains corrections which are nonperturbative in the semi-classical ($\hbar$) expansion that are so large that the theory does not admit a sensible vacuum any more. Put differently, the no-boundary proposal does not imply the Bunch-Davies vacuum for perturbations, as was until recently believed. Rather, increased fluctuations receive an ever larger weighting, leading to a breakdown of the model.

The instability can be related to the existence of the branch cuts on the real line in the perturbative action (5.19). In the absence of such a singularity, the h-function on the real line is strictly zero, and Picard-Lefschetz theory implies that any contour defined on the real line would have to flow down to lower values to be expressible as a manifestly convergent integral. In this case it would be impossible for the total weighting to become positive, and the fluctuations would not be able to surpass the background. In fact this makes perfect sense: quantum effects are suppressed compared to classical evolution, which would occur with probability 1 (, *i.e.*, weighting 0). The branch cut changes this. The h-function no longer tends to zero as one approaches the real N-axis from above or below the cut – rather it has a discontinuity leading to the instability discussed above. Another aspect of the problem is to notice that the resulting amplitude violates the correspondence principle, *i.e.*, classical physics is no longer recoverable in the limit $\hbar \to 0$. For large fluctuations, the propagators (5.39), (5.41) do not satisfy this condition. Whichever point of view one prefers, the conclusion in all cases is that the no-boundary proposal becomes untenable and that the idea of a smooth semi-classical beginning of the universe fails.

5.5 Backreaction

The inverse Gaussian distribution of the tensor perturbations, described in the previous section, arises already within the limits of validity of general relativistic linear perturbation theory, signalling a clear problem with the no-boundary proposal. However, one may wonder whether backreaction of the gravitational waves might be significant in the regime where the upper saddle points start to dominate over the lower saddle points (assuming the contour $\mathcal{C}_2^-$ passing below the origin). To settle this question we studied the backreaction numerically, in two representative situations of interest: for the lowest modes, *i.e.*, for the $l = 2$ modes, we have evaluated the full Einstein equations numerically. For the higher l modes we have solved the linear equation of motion for ϕ and included its backreaction at second order in the equation of motion for the scale factor q. As we will discuss below, these studies serve to reinforce the conclusions drawn in linear perturbation theory.

5.5.1 The $l = 2$ mode

The $l = 2$ modes are particularly interesting as a possible non-linear completion of the metric exists, in the form of the Bianchi IX line element

$$ds^2_{IX} = -N_p^2(t)dt_p^2 + \sum_m \left(\frac{l_m(t)}{2}\right)^2 \sigma_m^2 \,, \tag{5.42}$$

where N_p is the physical lapse function and $\sigma_1 = \sin\psi d\theta - \cos\psi \sin\theta d\varphi$, $\sigma_2 = \cos\psi d\theta + \sin\psi \sin\theta d\varphi$, and $\sigma_3 = -(d\psi + \cos\theta d\varphi)$ are differential forms on the three sphere such that $0 \leq \psi \leq 4\pi$, $0 \leq \theta \leq \pi$, and $0 \leq \phi \leq 2\pi$. For ease of notation we will denote a derivative w.r.t. physical time t_p by an overdot in this section (and only in this section). Employing the original definition of Misner [254], we can re-write the three scale factors as

$$l_1(t_p) = a(t_p) \exp\left[\frac{1}{2}\left(\beta_+(t_p) + \sqrt{3}\beta_-(t_p)\right)\right] \,, \tag{5.43}$$

$$l_2(t_p) = a(t_p) \exp\left[\frac{1}{2}\left(\beta_+(t_p) - \sqrt{3}\beta_-(t_p)\right)\right] \,, \tag{5.44}$$

$$l_3(t_p) = a(t_p) \exp\left[-\beta_+(t_p)\right] \,, \tag{5.45}$$

which makes it clear that a is the average scale factor while the βs quantify anisotropic perturbations. In these coordinates the action becomes

$$S = 2\pi^2 \int dt_p N_p a \left[\frac{1}{N_p^2} \left(-3\dot{a}^2 + a^2 \left(\frac{1}{2}\dot{\phi}^2 + \frac{3}{4}\dot{\beta}_+^2 + \frac{3}{4}\dot{\beta}_-^2 \right) \right) - \left(a^2 V(\phi) + U(\beta_+, \beta_-) \right) \right] , \tag{5.46}$$

where the full non-linear potential is given by

$$U(\beta_+, \beta_-) = -2 \left(e^{2\beta_+} + e^{-\beta_+ - \sqrt{3}\beta_-} + e^{-\beta_+ + \sqrt{3}\beta_-} \right) + \left(e^{-4\beta_+} + e^{2\beta_+ - 2\sqrt{3}\beta_-} + e^{2\beta_+ + 2\sqrt{3}\beta_-} \right) \tag{5.47}$$

$$= -3 + 6\beta_+^2 + 6\beta_-^2 + \ldots \tag{5.48}$$

Varying with respect to the lapse N_p (and working in the gauge $\dot{N}_p = 0$) we obtain the Friedman constraint equation

$$3\dot{a}^2 = a^2 \left(\frac{3}{4}\dot{\beta}_+^2 + \frac{3}{4}\dot{\beta}_-^2 \right) + N_p^2 U(\beta_+, \beta_-) \, , \tag{5.49}$$

while the equations of motion for a, β_+, β_- are given by

$$\frac{\ddot{a}}{a} + \frac{1}{2}\frac{\dot{a}^2}{a^2} + \frac{3}{8}\left(\dot{\beta}_+^2 + \dot{\beta}_-^2 \right) - \frac{N_p^2}{6a^2} U(\beta_+, \beta_-) = 0 \, , \tag{5.50}$$

$$\ddot{\beta}_\pm + 3\frac{\dot{a}}{a}\dot{\beta}_\pm + \frac{2}{3}\frac{N_p^2}{a^2} U_{,\beta_\pm} = 0 \, . \tag{5.51}$$

Expanding the last equation to linear order we obtain

$$\ddot{\beta}_\pm + 3\frac{\dot{a}}{a}\dot{\beta}_\pm + 8\frac{N_p^2}{a^2}\beta_\pm = 0 \, . \tag{5.52}$$

A comparison with Eq. (5.21) confirms that the βs are non-linear versions of the $l = 2$ modes – more specifically, they are non-linear versions of two $l = 2$ modes which are such that they preserve the Bianchi IX symmetry [173]. To match with our earlier normalization conventions, one has to re-scale

$$\beta_\pm = \frac{1}{\sqrt{3}\pi}\phi_\pm \, , \tag{5.53}$$

where $\phi_\pm$ denote two separate $l = 2$ modes. The structure of the potential U shows that when going beyond linear order, the equations of motion lead to direct couplings between these two $l = 2$ modes.

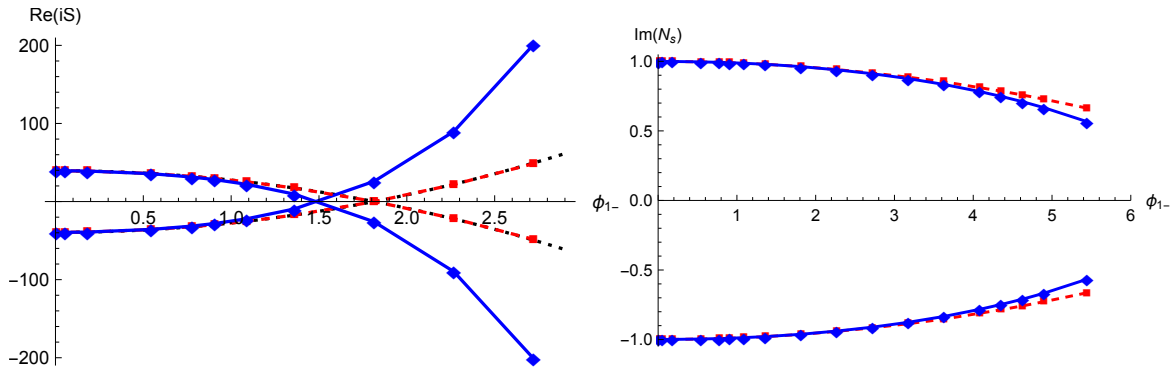


Figure 5.7: These graphs show the weighting at the saddle points (left panel) and the imaginary part of the saddle point locations (right panel) as a function of the $l = 2$ anisotropy mode amplitude ϕ_{1-}, for $\Lambda = 3$. In the plot of the action (left) the line starting at $+12\pi^2/\Lambda = +4\pi^2$ for $\phi_{1-} = 0$ corresponds to the saddle points in the lower half N plane, while the line starting at $-4\pi^2$ corresponds to the saddle points in the upper half plane. In black (mostly hidden behind the red line) are the linear results without backreaction, in red the results including backreaction but still in linear perturbation theory, and in blue the results stemming from solving the fully backreacted Einstein equations. For values of ϕ_{1-} below 1 the linear and non-linear results agree to high precision, while one can see that at larger values of the anisotropy the non-linear corrections enhance the instability of the fluctuations, and move the saddle points further towards the real N-axis. Note that the weighting of the upper saddle points surpasses that of the lower ones when backreaction is still entirely negligible. Moreover, the non-linear effects of the full Einstein equations imply that the (unstable) upper saddle points come to dominate already for smaller amplitudes of the fluctuations.

In the present section we work in a gauge where $N_p = 1$ and where one then has to determine the value of the time coordinate of the final hypersurface on which the boundary conditions $q_0 = 0, q_1 = a_1^2, \phi_\pm = \phi_{1\pm}$ are satisfied. This is done using the shooting method discussed in [55]. In this method, the (generally complex valued) second time derivatives of $\phi_\pm$ at the no-boundary point $a = 0$ are adjusted using an optimization algorithm such that at a final time t_f the desired real values $q_1, \phi_{1\pm}$ are simultaneously reached. The total time interval $\int N_p \mathrm{d}t_p = t_f$ can then also be related to the lapse function N using the change of coordinates $N_p \mathrm{d}t_p = N q^{-1/2} \mathrm{d}t$,

$$N = \int_0^1 N \mathrm{d}t = \int_0^{t_f} a(t_p) \mathrm{d}t_p \,. \tag{5.54}$$

Our results are shown in Fig. 5.7. For ease of comparison with linear perturbation theory, we only show results for the case where a single $l = 2$ mode (here chosen to be ϕ_{1-}) takes on a non-trivial value on the final hypersurface. The left panel shows how the weighting of the saddle point solution (for saddles in the upper half plane) increases as the perturbation amplitude is increased. The opposite behavior is seen for the saddle points in the lower half plane. As is evident from the figure, backreaction at second order in perturbation theory is utterly negligible. Even more importantly, the effects of the instability are even stronger when non-linear terms are included, and the dominance of the upper saddle point over the lower ones occurs already for smaller values of ϕ_{1-} than in the linear theory. Also, as shown in the right panel, the saddle point moves faster towards the real N-axis in the non-linear theory. These results consolidate our analytic results, and indicate that the inclusion of non-linear backreaction only reinforces the instability that we have identified.

5.5.2 Backreaction in ϕ of higher l modes

To quadratic order in the gravitational wave modes, the equations of motion corresponding to the total action $S = S^{(0)} + S^{(2)}$ are

$$0 = \ddot{q} - \frac{2N^2}{3}\Lambda + \frac{\dot{\phi}^2}{3\pi^2}q\,, \tag{5.55}$$

$$0 = \ddot{\phi} + 2\frac{\dot{q}}{q}\dot{\phi} + \frac{N^2}{q^2}l(l+2)\phi\,. \tag{5.56}$$

The term $\frac{\dot{\phi}^2}{3\pi^2}q$ encodes the backreaction of the perturbations ϕ on the scale factor q, ignored in the analytic calculations of the previous sections.

We use a numerical shooting method to solve the equations of motion (5.55) and (5.56) with "no-boundary" boundary conditions $q(0) = \phi(0) = 0$, $q(1) = q_1$, and $\phi(1) = \phi_1$ for a given spherical wavenumber l and lapse N. We start with the analytic solution (5.9) of the scale factor ignoring backreaction. We then solve (5.56) numerically for an arbitrary $\phi(1)$, and normalize the solution to enforce $\phi(1) = \phi_1$. The resulting approximate solution for ϕ is used in the equation of motion (5.55) for $q(t)$, likewise solved as a one dimensional shooting problem. The procedure is then iterated until both solutions converge. Fig. 5.8 compares the results with and without backreaction included, for a saddle point solution with $\phi_1 = 3$, for which the perturbation contribution to the h-function at the saddle outweighs that for the unperturbed background.

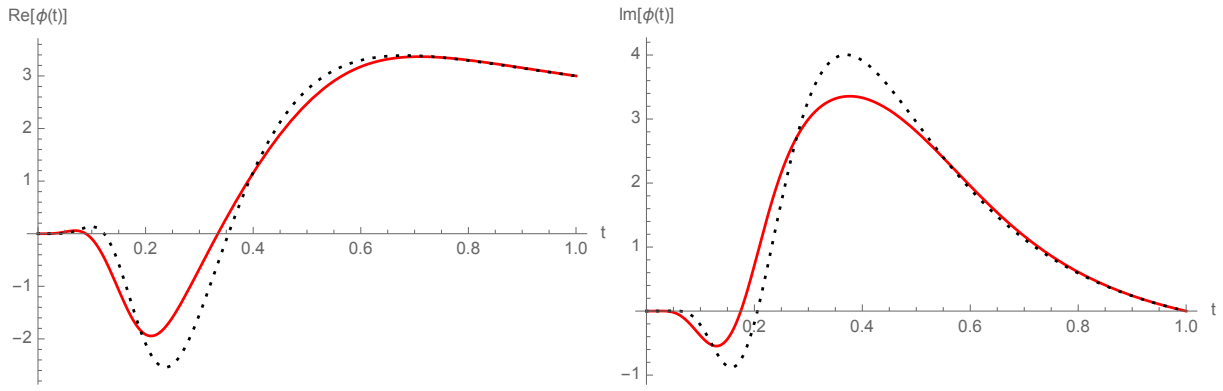


Figure 5.8: The real and imaginary parts of the mode function $\phi(t)$ in the saddle point N_1 with the boundary conditions $q_0 = 0, \phi_0 = 0$, $q_1 = 101$, and $\phi_1 = 3$ for mode $l = 10$. The black dashed lines correspond to the analytic result without backreaction. The red lines correspond to the numerical analysis with backreaction. We observe that the backreaction for $\phi_1 = 3$, which violates the perturbation theory condition $|\phi(t)| > 1$, leads to a change in the mode functions of approximately 10%, with no qualitative change. For the boundary condition $0 \leq \phi_1 \leq 0.8$, which do satisfy the condition $|\phi(t)| < 1$ for all $t \in [0, 1]$, the correction due to backreaction is completely negligible.

For more modest values of the final perturbation, $0 \leq \phi_1 \leq 1$, we find that the backreaction is small and does not significantly affect the location of the saddle points of the h-function in the complex N-plane. The saddle points move the most for the high l modes. For $l = 10$, $q_1 = 101$ and $\phi_1 = 1$, for example, the saddle points are located at $N_s = \pm 10.0232 \pm 0.97904i$, compared to the background saddle points $N_s = \pm 10 \pm i$. In agreement with our analytic arguments, the value of the h-function at the saddle points does, however, change significantly as ϕ_1 is increased (see Fig. 5.9). The lines starting from $h = -4\pi^2$ and $h = +4\pi^2$ correspond respectively to the upper and the lower saddle points. For the modes $l = 2, 3$ and $l = 4$ the lower saddle point always dominates over the upper saddle point in the regime of validity of linear perturbation theory. This coincides with the analysis of the $l = 2$ mode using full Einstein equations, discussed above, which showed that the upper saddle only dominates at values of ϕ_1 greater than ~ 1.5). For $l \geq 5$, the upper saddle points dominate over the lower saddle points within the regime of validity of linear theory.

As explained in section 5.4, it is possible for the upper saddles to dominate (and even to acquire a positive real exponent) due to the existence of branch cuts on the real line in the effective action for N, given in (5.19) and (5.28). Fig. 5.10 shows that including

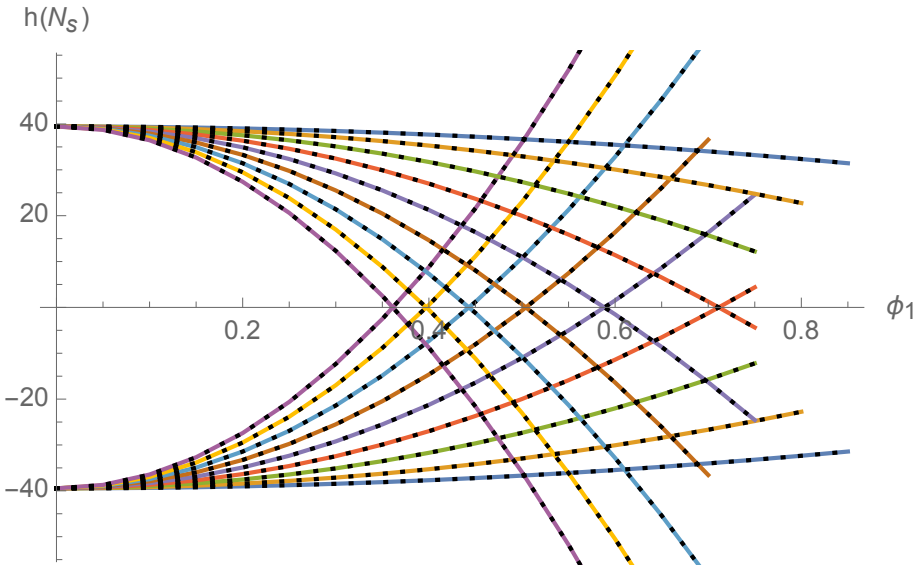


Figure 5.9: The h-function evaluated at the saddle points $h(N_s)$ as a function of the boundary condition ϕ_1 for which $|\phi(t)| \leq 1$ for all t. The lines starting from $h = -4\pi^2$ and $h = 4\pi^2$ correspond to the upper and the lower saddle points. The blue to purple lines correspond to $l = 2, 3, \ldots, 10$. The colored lines are the numerical simulations with backreaction. The dotted lines are the analytic results without backreaction. We observe that the backreaction of the perturbation ϕ on the scale factor q is very small for the boundary condition $0 \leq \phi_1 \leq 0.8$. Consequently, the analytic calculations which neglect backreaction are accurate when the upper saddle point starts to dominate over the lower saddle point for $l \geq 5$.

the effects of backreaction does not remove this branch cut. As we approach the point $N_\star$ from above and below the real N-axis, one can see the jump in the h-function at $N_\star$ is maintained.

In general, we find that the effects of nonlinear backreaction are significant only near the real N-axis. This is perhaps not surprising, since we know the theory is quite singular there: for example it is well known that the Bianchi IX model studied in the previous section exhibits chaotic behavior for real metrics as the singularity $q = 0$ is approached. However, we find that nonlinear backreaction is insignificant along the Lefschetz thimble $\mathcal{J}_1$ associated with the upper saddle point, *i.e.*, the thimble relevant to the strictly Lorentzian path integral. The upper panel of Fig. 5.11 illustrates the first quadrant of the complex N plane. In the white region, defined by $\mathrm{Re}[\gamma] > 1$, the finite action condition forces the mode function to vanish on the initial boundary, *i.e.*, $\phi_0 = 0$. In this region the shooting method

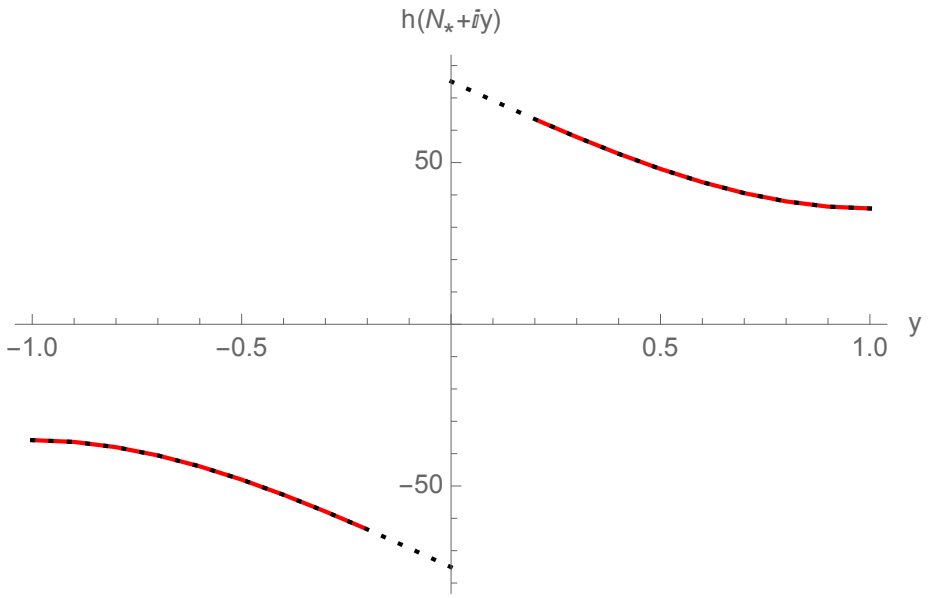


Figure 5.10: The function h evaluated at $N = N_\star + iy$ with the boundary conditions $\Lambda = 3, q_1 = 101, \phi_1 = 0.5$ for the mode $l = 10$. The dashed line is the analytic result without backreaction. The red line is the numerical calculation including backreaction. We plot the backreaction calculation for the points for which $|\phi(t)| < 1$ (, *i.e.,* for which linear theory is reliable) and the backreaction on the q variable is small, *i.e.*, $\frac{\dot{\phi}^2 q}{3\pi^2} < \left|\frac{2N^2\Lambda}{3}\right|$ for all t. Note that the backreaction is extremely small near the saddle point for these boundary conditions. The jump at $y = 0$ illustrates the branch cut. The backreaction becomes significant when N approaches the real axis. However, the backreaction does not appear to remove the branch cut.

described above may be used. In contrast, in the shaded region, defined by $\mathrm{Re}[\gamma] \leq 1$, the finite action condition selects a mode function ϕ which diverges on the initial boundary and a different method should be used.

Note that the part of the Lefschetz thimble shown in the Figure lies entirely in the white region. In order to study the significance of backreaction, we selected 32 regularly spaced points along the thimble. In the lower two panels we plot the h-functions given by the analytic calculations and the numerical calculations including backreaction, respectively, for each of the background and fluctuation actions. We observe that backreaction is negligible along the Lorentzian Lefschetz thimble $\mathcal{J}_1$. Thus, deforming the contour from real fields to the Lefschetz thimble appears to render the path integral significantly more

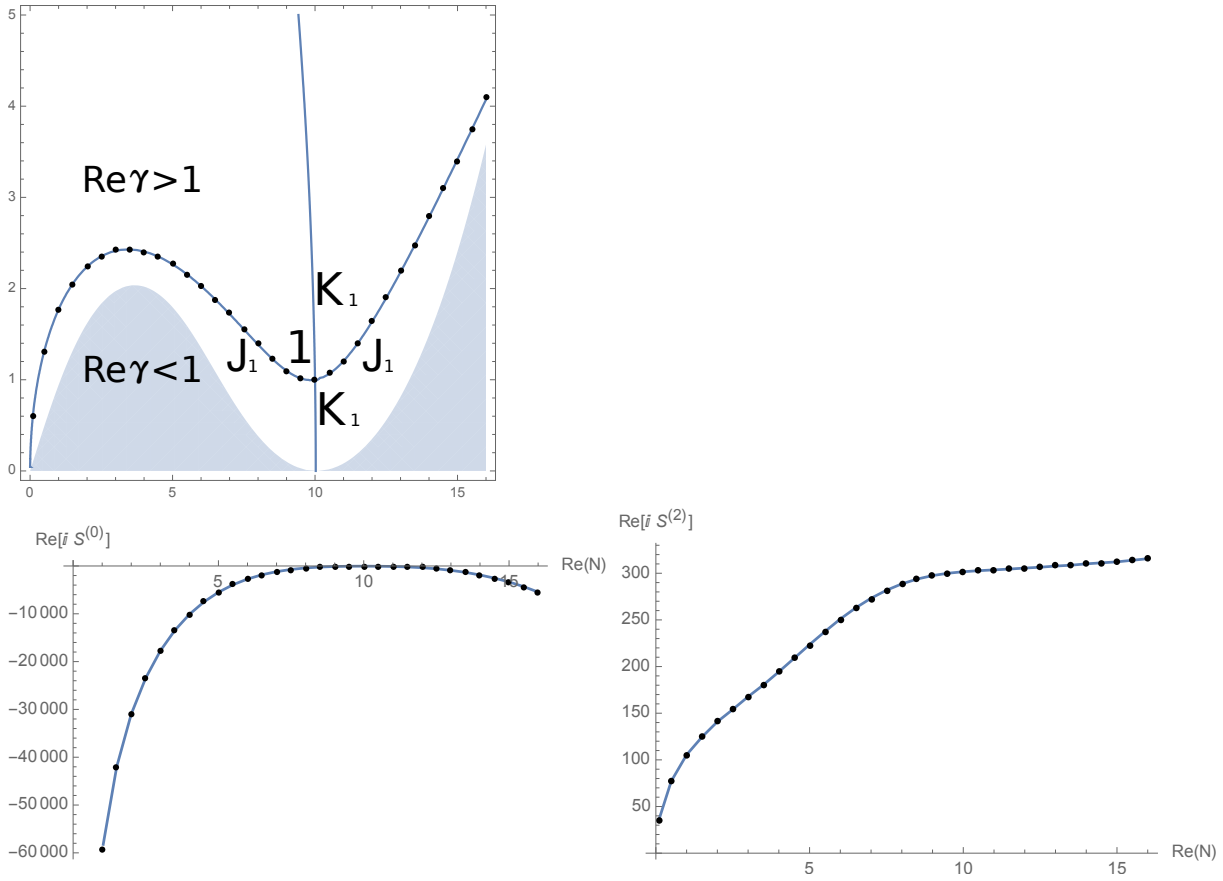


Figure 5.11: The h-function along the thimble for the boundary conditions $\Lambda = 3, q_1 = 101, \phi_1 = 1$ for the mode $l = 10$. For these boundary conditions the action of the background and perturbations are comparable in the saddle points. Upper panel: first quadrant of the complex N-plane with the lines of steepest ascent and descent of the upper saddle point. The shaded/white regions, defined by Re[γ] less/greater than 1 respectively, denote the regions in which the finite action condition selects a mode which diverges/vanishes on the initial boundary. The points on the thimble $\mathcal{J}_1$ indicate the values of the lapse N at which we have evaluated the importance of backreaction in the lower panels. Lower panels: the h-functions for the background and perturbation actions $\bar{S}^{(0)}$ and $\bar{S}^{(2)}$. The blue lines denote the analytic h-function while the black dots denote the numerical results including backreaction. Note that nonlinear backreaction is completely negligible along the thimble.

calculationally tractable.

5.5.3 Backreaction in χ

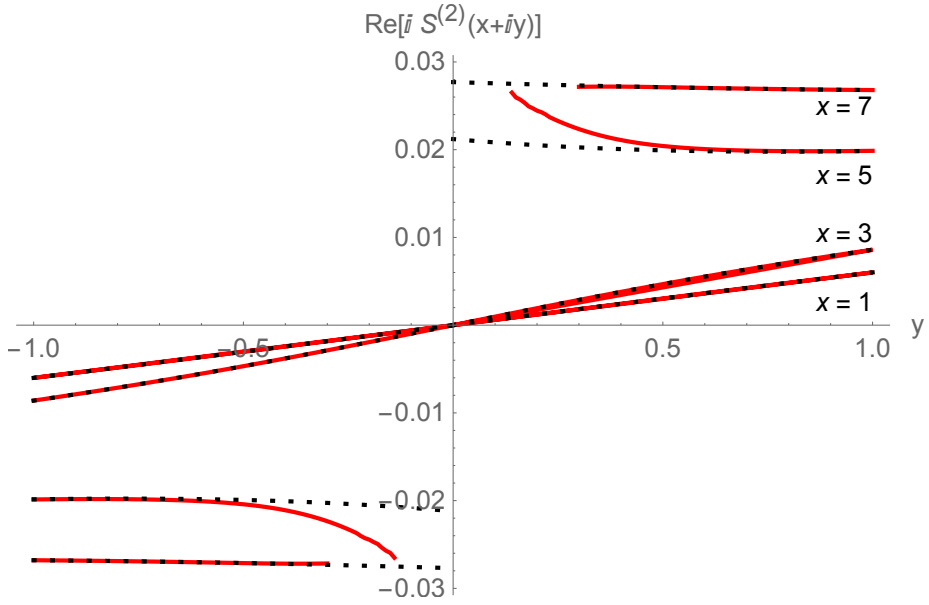


Figure 5.12: The h-function for the perturbations is plotted along a line $N = x + i\,y$ in the complex N-plane, crossing the real N-axis $y = 0$ in the vertical direction. The parameters and boundary values are $\Lambda = 3, q_0 = 0, q_1 = 101$, $l = 10$, $\chi_0 = 0$, and $\chi_1 = 1$. The black dashed line is the analytic result of perturbation theory without backreaction. The red line is the numerical result including backreaction up to second order in the perturbations.

In the previous sections we studied the significance of backreaction in terms of the mode function ϕ. As can be seen in the upper panel of Fig. 5.11, the shooting method for ϕ only works for part of the complex N plane. Near the real line, in particular, the shooting method for ϕ breaks down. In order to study the effects of backreaction near the real N-axis to the left of the branch cut we change coordinates to $\chi = q\phi$ as was discussed in section 5.4. To quadratic order in the gravitational wave modes, the equations of motion corresponding to the action $S = S^{(0)} + S^{(2)}$ are

$$0 = \ddot{q} - \frac{2N^2}{3}\Lambda + \frac{q}{2\pi^2}\left(\frac{\dot{\chi}}{q} - \frac{\dot{q}}{q^2}\chi\right)^2 \tag{5.57}$$

$$0 = \ddot{\chi} - \left(\frac{\ddot{q}}{q} - \frac{N^2 l(l+2)}{q^2}\right)\chi \tag{5.58}$$

By solving these equations with the shooting method described in the previous section we can study the effects of backreaction in the shaded regions of the upper panel of Fig. 5.11 since the finite action condition implies $\chi_0 = 0$ (see Fig. 5.12). Observe that for $\mathrm{Re}[N] = 1, 3$ the h-function does not appear to be discontinous across the real N-axis (, *i.e.,* it suffers no branch cut) and backreaction remains negligible. When crossing the real line further to the right, for $\mathrm{Re}[N] = 5, 7$, we do find a branch cut and likewise notice significant backreaction. These results indicate that the presence or absence of a branch cut in the effective action for N, on the real N-axis, appears at least qualitatively consistent with indications from perturbation theory. This finding is significant for the discussion of the next section.

5.6 No contour works

In Section 5.2 we reviewed the physical principles for integrating over positive lapse ($\mathcal{C}_1$) when defining the Lorentzian propagator. In contrast, Diaz Dorronsoro *et al.* [104] proposed a different, intrinsically complex contour $\mathcal{C}_2^-$ running below the origin over $-\infty < N < \infty$, with the motivation of obtaining a real wavefunction satisfying the Wheeler-DeWitt equation. (For an earlier discussion of various lapse contours for the background, see, *e.g.*, Ref. [186]). We have already shown that with either choice, one cannot avoid contributions from saddles 1 and 2 (shown in Fig. 5.13), both of which yield an unsuppressed, inverse Gaussian distribution for the perturbations. The goal of this section is to extend this analysis and show that *no* integration contour avoids this problem.

Before passing to the proof of this general result, let us also mention the new complication we have detailed in this chapter. When the condition of finite action is imposed upon the perturbation modes, a large part of the real N-axis must be excluded from the definition of any possible contour, due to the presence of branch cuts and the non-existence of finite action perturbation modes at large $|N|$.

5.6.1 Time reparametrization invariance

Let us consider a generalized path integral, where the lapse N is integrated over some complex contour $\mathcal{C}$ in the complex N-plane. Time reparametrization invariance severely limits the points at which $\mathcal{C}$ may start and end. Defining $\langle 1|0\rangle_{\mathcal{C}} \equiv \int_{\mathcal{C}} dN \langle 1|e^{-iHN/\hbar}|0\rangle$, as a natural generalization of (5.3), it follows that $H\Psi = \int_{\mathcal{C}} i\hbar(d/dN)\langle 1|e^{-iHN/\hbar}|0\rangle$, *i.e.*, one obtains surface terms at the start and end of $\mathcal{C}$. If one insists that the Hamiltonian

annihilates the amplitude, it follows that all surface terms must vanish: that is, $\mathcal{C}$ must start and end at a point where the function $h(N)$ tends to minus infinity (we assume the $h(N)$ dependence dominates over the prefactor). There is a subtlety at the point $N = 0$ because, as discussed above Eq. (5.5), the integrand $\langle 1|e^{-iHN/\hbar}|0\rangle$ is singular in the small N limit, with the prefactor and the exponent diverging in just such a way that the combined surface term generates a delta function. This is a universal property of quantum propagators, ensuring that they are Green's functions. As we reviewed in Section 5.2, for geometrodynamics this introduces a primitive notion of causality. Nevertheless, it remains true that the only acceptable contours for $\mathcal{C}$ are those which start and end at points where the h-function tends to minus infinity. Furthermore, if the integral over N is to converge, $\mathcal{C}$ must approach these singular points along paths which can be deformed into paths of steepest descent. Near $N = 0$, the background action (5.10) diverges as $-A/N$ with A positive. It follows that $\mathcal{C}$ must approach the origin along a path which can be deformed to run into the origin along the positive imaginary axis, since that is the unique curve of steepest descent which ends at $N = 0$. In contrast, when we consider the singular point at $N = \infty$, the classical action is proportional to AN^3, with A positive. So the point at infinity may be approached in three inequivalent ways: the exponent iN^3 yields steepest descent paths along $N = |N|e^{i\theta}$ with $\theta = \pi/6$, $5\pi/6$ or $3\pi/2$. (As we shall discuss later, in general there is a branch cut emanating from the origin. In this case we must distinguish the angle $3\pi/2$ from $-\pi/2$, relative to the positive N-axis.)

5.6.2 Picard-Lefschetz theory

Let us begin by ignoring the branch cuts on the real N-axis. Then it is a basic result of Picard-Lefschetz theory that any contour $\mathcal{C}$ running between singular points of the h-function and yielding a finite integral may be deformed, using Cauchy's theorem, into a sum of Picard-Lefschetz "thimbles". In order to remove any degeneracies between thimbles (see section 5.2 of chapter 3), we introduce a small deformation so that thimble 3 is completed by passing just below the steepest descent curve connecting saddle point 2 to the origin, and similarly for thimble 4 and thimble 1. These thimbles are illustrated in Fig. 5.13. Any contour $\mathcal{C}$ connecting singularities of the h-function, as explained above, may now be deformed into a sum of thimbles, taken with appropriate signs. Notice that, with this definition of the thimbles, in the limit that the deformation is taken to zero, thimble 4 and thimble 1 share that part of thimble 1 which connects saddle point 1 to the origin. And the thimble 3 shares a similar part of thimble 2. In this way *every thimble includes a contribution from saddle point 1 or 2*, both of which yield unsuppressed perturbations. This is the first part of the proof.

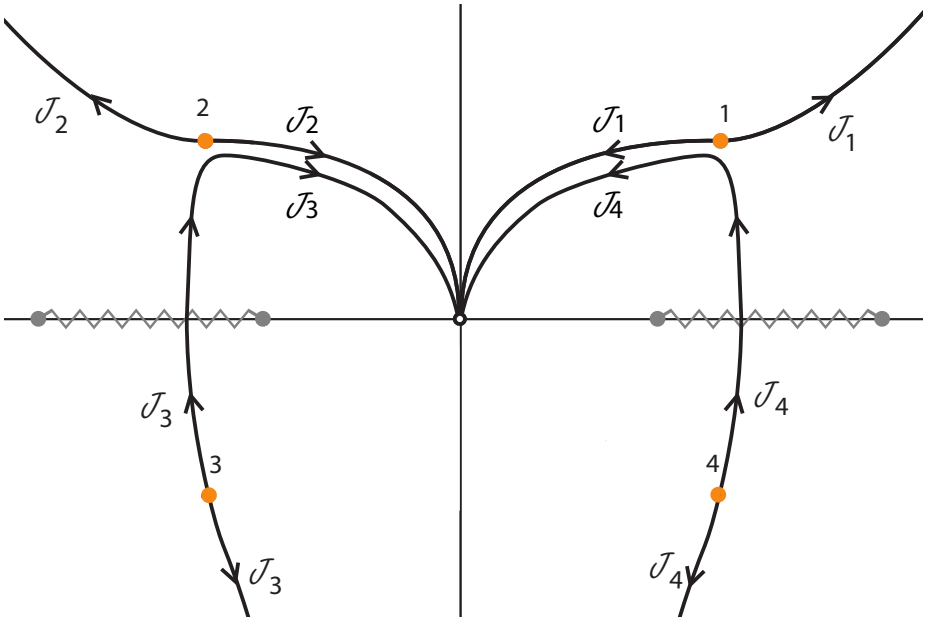


Figure 5.13: The Picard-Lefschetz thimbles (solid dark lines, labelled $J_1 - J_4$) for our problem, in the complex N-plane. Any convergent integral starting and ending at a singularity of the h-function can be represented as a sum of these thimbles. Also shown are the branch cuts (jagged grey lines) which the no-boundary perturbations introduce on the real N-axis. As we explain later, the contours $\mathcal{J}_3$ and $\mathcal{J}_4$ need to be deformed around the branchcuts (to the inside) when including the perturbations.

For example, our Lorentzian propagator is given by thimble 1. Its real part is given by the sum of thimble 1 and thimble 2, both taken 'left to right,' and its imaginary part is proportional to the difference. The contour of Diaz Dorronsoro *et al.* is deformable to thimbles 1,2,3 and 4 taken with signs such that the steepest descent contours connecting saddle points 1 and 2 to the origin are cancelled. But the steepest descent contours connecting saddle points 1 and 2 to infinity are included. Other contours include thimbles 3 or 4, either taken alone or combined into a contour which 'runs around' the origin, starting from negative imaginary values. Likewise we can combine thimble 3 and thimble 2, taken with a sign such that the steepest descent contour connecting saddle point 2 to the origin is cancelled. And so on.

Now let us include the two branch cuts on the real N-axis. Clearly, their introduction affects thimbles 3 and 4. Any contour which gives contributions from saddles 3 or 4 – the Hartle-Hawking saddles – as, for example, that of Diaz Dorronsoro *et al.*, necessar-

ily involves contributions from either or both branch cuts, which also give unsuppressed perturbations.

Finally, we note in passing that in general the singularity at $N = 0$ may also be the terminus of a branch cut extending to infinity. In pure de Sitter, the path integral over the background yields a square root of N and the perturbations, as noted above, yield a functional determinant. So one can consider a nontrivial contour for the integral over N coming in from infinite negative imaginary values and completely encircling this branch cut. It will be deformable to a combination of thimbles 1,2, 3 and 4. Again, it necessarily acquires contributions from both saddles 1 and 2, and the branch cuts. This completes our proof that no choice of contour for N avoids the problem of unsuppressed perturbations.

5.7 Discussion

The introduction of a new, more precise formulation of semi-classical quantum gravity should, we believe, have wide implications.

Among the questions which now appear accessible is whether or not there is a quantum version of de Sitter spacetime. In Refs. [277, 278], Polyakov has expressed doubts, claiming in particular that there is strong interplay between IR and UV effects leading to various divergences. While there are considerable differences between our approaches, his words certainly resonate with our findings. Namely, our predicted distribution of the tensor modes diverges, a problem which worsens in the UV, as a result of a nonlinear (and completely nonlocal) interplay between these modes and the quantized background, *i.e.*, the IR. Because the background has the "wrong sign" kinetic term, convergence of the Feynman path integral over backgrounds in effect chooses the "wrong sign" Wick rotation for the perturbations, giving them an inverse Gaussian distribution and implying they are completely out of control. The general theorem we proved in chapter 5 identified the fundamental topological nature of this problem.

Our work indicates that a "no-boundary" formulation of quantum de Sitter spacetime does not exist. However, the question remains whether there is any other viable formulation. In particular, one might try to define it by considering the Lorentzian propagator between two real classical three-geometries, both large and roughly spherical, so that there would be two possible intervening classical four-geometries, one a "bounce" including the throat of de Sitter and the other not including the throat. In the former case, one may have to worry about strong nonlinear backreaction, in that the quantum mechanical perturbations present in the contracting phase of global de Sitter spacetime blueshift as the

size of the three-sphere shrinks, potentially causing a "big crunch" at the throat and terminating the semiclassical description. The methods reported here, especially concerning semiclassical backreaction, seem ideal for approaching these questions. In particular, we are very interested in understanding implications for the future quantum evolution of our universe, which now appears to be entering a potentially eternal de Sitter phase. Is it possible to describe such a phase semiclassically? Is such a phase ultimately stable against quantum mechanical perturbations? Finally, given that our universe emerged from the big bang filled with radiation and matter, how does the presence of these other forms of energy affect the quantum mechanical behaviour of the background and the perturbations? The work of Refs. [163, 164] on a radiation-dominated "perfect bounce" indicated no problems with unsuppressed perturbations, such as those identified here. We are therefore keen to explore further the possible advantages of such a scenario.

Second, our work may well have implications for cosmological inflation. In the usual descriptions, one treats the background classically, and quantizes only the perturbations. However, as is well known, it is not possible to consistently couple classical degrees of freedom with quantum degrees of freedom, so this usual description can at best be only an approximation. Second, the usual treatment of inflation *assumes* the quantum mechanical perturbations start out in their local adiabatic vacuum – the so-called Bunch-Davies state. Here too, this is at best an approximation since it applies only to short wavelength modes. There is the additional problem that such modes are generically sub-Planckian deep in the inflationary era, so without a clear treatment of quantum gravitational phenomena, they cannot be precisely specified.

The promise of the no-boundary proposal was to present a "completion" of the inflationary scenario, along with the hope that all these questions could be resolved within a low-energy, effective description of Einstein gravity coupled to quantum fields. We believe that our work has now excluded that option. Therefore, it raises fundamental questions for inflation: How is the theory to be completed? How does inflation avoid non-perturbative corrections to the Bunch-Davies vacuum, of the type we have shown to exist for the contour proposed by Diaz Dorronsoro *et al.* in their attempt to rescue Hartle and Hawking's proposal?

Although the main outcome of this work has so far been negative, we find it very exciting that we can at last formulate semi-classical quantum cosmology in a precise enough way to identify clear problems. In our view, the bigger the problem, the more instructive the clue it provides. It is particularly surprising to find that quantum effects on the universe on large scales can be so significant and can so drastically influence the description of phenomena on the smallest scales, as we found. One cannot help but hope that herein lies a clue to understanding the dark energy and resolving the biggest "fine tuning" puzzle

in physics. Quantum field theory tells us that the dark energy is dominated by the UV vacuum energy, yet it is also what ultimately sets the IR cutoff by limiting our causal horizon and thus the largest scale anyone shall ever see. Again, our work provides a clue as to how the UV and the IR are connected in quantum gravity, in a way we have yet to fully unravel.

Finally, we mentioned in the opening of this chapter that the no-boundary concept has been very fruitful in mathematical physics, in contexts such as holography, as well as conformal and topological field theory. We believe our more precise formulation should also be useful in these contexts.

5.A Illustration of higher dimensional Picard-Lefschetz theory

In this appendix, we use a simple two-dimensional, conditionally convergent, oscillatory integral to illustrate some features of Picard-Lefschetz theory. The integral may be computed iteratively in two ways: by integrating first over one variable and then the other, or vice versa. In the former case, as we shall see, we obtain a one dimensional conditionally convergent integral with a nonsingular integrand. In the latter case, the first integration yields an integrand which is singular on the original contour for the second variable. In order to give the integral meaning we must exclude the singular point or, equivalently, deform the original contour to avoid the singularity. An alternative to making sense of the original integral is to use higher dimensional Picard-Lefschetz theory and Cauchy's theorem to distort the two original integration contours so that the two-dimensional integral becomes absolutely convergent *before* either integral is performed. As discussed in section 5.2, this ensures that the result of the integral is independent of the order in which the two integrals are then taken. More than this, as we shall see, distorting the contours to a two-dimensional steepest descent surface ensures that no one dimensional integral generates a singularity which might then complicate subsequent integrals.

We consider the integral

$$I = \int_{-\infty}^{\infty} dN \int_{-\infty}^{\infty} dz\, e^{i(N-1)z^2+iN^2}, \tag{5.59}$$

where both N and z are taken over all real values. To begin with, we compute I by

integrating first over N. This is a simple Gaussian integral which yields

$$I = e^{i\pi/4}\sqrt{\pi}\int_{-\infty}^{\infty} dz\, e^{-i(z^2+\frac{1}{4}z^4)} = \frac{1}{2}e^{i(\frac{1}{2}-\frac{\pi}{8})}\pi^{\frac{3}{2}}H^{(2)}_{-\frac{1}{4}}\left(\frac{1}{2}\right), \tag{5.60}$$

the last being a standard result. Alternatively, we may compute I by integrating first over z. But there is an immediate problem: the z integral, taken along the real z-axis, fails to converge at $N = 1$. So we can exclude the point at $N = 1$, which is of zero measure, to obtain

$$I = e^{i\pi/4}\sqrt{\pi}\int_{1}^{\infty} \frac{dN}{\sqrt{N-1}}e^{iN^2} + e^{-i\pi/4}\sqrt{\pi}\int_{-\infty}^{1} \frac{dN}{\sqrt{1-N}}e^{iN^2}. \tag{5.61}$$

Equivalently, we may define the integral over a continuous contour for N following the real axis from $-\infty$ to ∞ but avoiding $N = 1$ on an infinitesimal semicircle passing above it (only if we pass above is the z integral in (5.59) convergent). In the limit that the semicircle radius is taken to zero, we also obtain (5.61). One can readily check that (5.61) agrees with (5.60).

While both derivations are correct, the second one is complicated by the occurrence of the singularity, requiring a contour for N which is either discontinuous or moves off the real axis. A similar phenomenon occurs in our more complicated path integral examples where the integration over the perturbations generates singularities in the resulting effective action for the lapse N which likewise occur on the original, defining integration contour. In that case, the singularities are more severe, occurring in the *exponent* of the integrand, as explained in the Introduction. In both cases, however, if we wish to calculate the integral along the original N contour having performed partial integrations, care is needed to determine exactly how the contour should avoid the singularities.

Let us now see how Picard-Lefschetz theory enables one to deform the integration contour to steepest descent contours *before* either integral is performed, in order to avoid such problems. As we shall see in this example, this avoids singularities of the type just discussed being generated, leaving one with a completely unproblematical absolutely convergent integral which can be calculated iteratively, in any convenient order. Examining (5.59), we see that we can make the integral over N convergent by rotating the N contour, setting $N = e^{i\frac{\pi}{4}}\sqrt{2}n = (1+i)n$, with n real (the factor of $\sqrt{2}$ is merely to simplify the algebra). Note that this contour for N never passes through the point $N = 1$. Turning now to the z integral, we set $z = x + iy$ and find the steepest descent contour $y(x)$ passing through the saddle point at $z = 0$. As explained, *e.g.*, in section 5.2 of see chapter 3), the imaginary part of the exponent is constant along this contour. This immediately yields

the equation $(n-1)(x^2-y^2)-2nxy=0$. One of the two solutions gives the steepest descent contour, in which the real part of the exponent (the h-function) is monotonically decreasing away from the saddle (the other solution, which we ignore, gives the steepest ascent contour). Setting $n=\nu+\frac{1}{2}$, the steepest descent contour is

$$y=\frac{1+2\nu-\sqrt{2+8\nu^2}}{1-2\nu}x \tag{5.62}$$

and the h-function on this contour is

$$h=-2(1+4\nu^2)\frac{\sqrt{2+8\nu^2}-1-2\nu}{(1-2\nu)^2}x^2-2\left(\nu+\frac{1}{2}\right)^2. \tag{5.63}$$

Notice that the slope of the contour in the complex z-plane changes from a negative to a positive value as ν (or n) runs from $-\infty$ to ∞, so that the coefficient of x^2 is always negative, whatever the (real) value of ν.

Having found the steepest descent contour in the complex z-plane, we may now integrate over z, along the steepest descent contour (5.62), *without* generating any singularity in ν:

$$\begin{aligned} I &= \sqrt{2}e^{i\frac{\pi}{4}}\int_{-\infty}^{\infty}dx\left(1+i\frac{dy}{dx}\right)\int_{-\infty}^{\infty}d\nu e^{-2[(1+4\nu^2)\frac{\sqrt{2+8\nu^2}-1-2\nu}{(1-2\nu)^2}x^2+(\nu+\frac{1}{2})^2]} \\ &= \sqrt{\pi}e^{i\frac{\pi}{4}}\int_{-\infty}^{\infty}d\nu\left(1+i\frac{1+2\nu-\sqrt{2+8\nu^2}}{1-2\nu}\right)\frac{|1-2\nu|}{\sqrt{1+4\nu^2}\sqrt{\sqrt{2+8\nu^2}-1-2\nu}}e^{-2(\nu+\frac{1}{2})^2} \end{aligned} \tag{5.64}$$

While this expression is hardly elegant, it is completely nonsingular along the integration contour for ν. It is easy to integrate numerically, for example, and yields a value identical to that of (5.60).

The key point illustrated by this example is that when we use higher-dimensional Picard-Lefschetz theory to convert the original integral, which is only conditionally convergent, into one which is absolutely convergent, partial integrals do not typically generate any singularities. However, if we instead leave N fixed while performing the z integral, we obtain singularities on the real N-axis. Care is then needed to determine how the N integration contour should be taken around them.

5.B Proof that $\Re[\gamma]>0$ almost everywhere in the cut N-plane

A crucial role in our analysis is played by the constant γ describing the behavior of the perturbation modes near $t=0$. Near $t=0$, we have $\phi\sim t^{\frac{\gamma-1}{2}}$ and the Lagrangian density

$\mathcal{L}(t) \sim t^{\gamma-1}$. So, if the real part of γ is positive, the singularity at $t = 0$ is integrable. We shall now prove that $\Re[\gamma] > 0$ almost everywhere in the complex N-plane, the exception being the closed real intervals $N_- \leq N \leq N_+$ and $-N_+ \leq N \leq -N_-$, where N_- and N_+ are positive constants, defined in (5.24).

In Section 5.4, we show that

$$\gamma = \sqrt{\frac{N_-^2 - N^2}{N_\star^2 - N^2}}\sqrt{\frac{N_+^2 - N^2}{N_\star^2 - N^2}} \equiv \sqrt{\frac{Z_- - Z}{Z_\star - Z}}\sqrt{\frac{Z_+ - Z}{Z_\star - Z}}, \tag{5.65}$$

where $N_- < N_\star < N_+$, and we define $Z = N^2$. Consider the first factor. It is the square root of a Möbius map, namely a one-to-one mapping of the complex Z-plane onto the complex Y-plane, where $Y = (Z_- - Z)/(Z_\star - Z)$. This map takes the real interval $Z_- < Z < Z_\star$ onto the negative real Y-axis, the upper half Z-plane to the lower half Y-plane and vice versa. Setting $S = \sqrt{Y}$ takes the entire Z-plane to the right half of the S-plane, *i.e.*, $\Re[S] \geq 0$, with the upper half Z-plane being mapped to the lower right quadrant of the S-plane, and the lower half Z-plane to the upper right quadrant of the Z-plane. Except for the closed interval $Z_- \leq Z \leq Z_\star$, which is mapped to the imaginary S-axis, every point in the complex Z-plane is mapped to a point with $\Re[S] > 0$, with the sign of $\Im[S]$ being opposite to the sign of $\Im[Z]$. Now consider the second factor, written as the complex number T. Similar arguments show that, except for the real interval $Z_\star \leq Z \leq Z_+$, every point in the complex Z-plane is mapped to $\Re[T] > 0$, with the sign of $\Im[T]$ being the same as the sign of $\Im[Z]$. Therefore $\Re[ST] = \Re[S]\Re[T] - \Im[S]\Im[T] > 0$ except on $Z_- \leq Z \leq Z_+$, where ST is pure imaginary. This proves our claim.

5.C No-boundary amplitude for perturbed S^3 with $\Lambda = 0$

In subsection 5.1.4 of the Introduction, we discussed a very simple example of a "no-boundary" gravitational path integral. Namely, we set $\Lambda = 0$ in the action (5.8) governing the background and impose a final three-geometry consisting of a sphere of radius squared q_1. The background solution is $q = q_1 t$, and the classical background action is $2\pi^2(-3q_1^2/(4N) + 3N)$. This has two saddle points, at $N = N_s^\pm = \pm i q_1/2$. This means that the saddle points are actually on the imaginary axis. For either saddle, the line element is $q_1(dt^2/(4t) + t d\Omega_3^2) = q_1(dr^2 + r^2 d\Omega_3^2)$, where $r \equiv \sqrt{t}$, *i.e.*, just that for a flat Euclidean ball of radius $R \equiv \sqrt{q_1}$.

Thus the saddle in the upper (lower) half plane has semiclassical exponent $-6\pi^2R^2/\hbar$ (or $+6\pi^2R^2/\hbar$) respectively, where R is the radius of the ball in reduced Planck units. By studying the flow, one can easily see that the upper saddle is on a Picard-Lefschetz thimble connecting the origin $N = 0$ (approached from above) to $N = +i\infty$. Therefore this saddle point is relevant to the *Euclidean* path integral for quantum gravity, in the absence of a cosmological constant. The reason the Euclidean path integral is meaningful is that if $\Lambda = 0$, the positive imaginary N-axis is in fact a direction of steepest descent. Note, however, that since the action is odd in N one can only take N to run over the positive imaginary axis, not the entire imaginary axis. The same saddle is *also* relevant to the *Lorentzian* path integral, with a defining contour $0^+ < N < \infty$, which also makes sense. The steepest ascent flow from the saddle is easily shown to be a circle: setting $N = x + iy$, the flow is $y^2 = \frac{1}{4}q_1^2 - x^2$. So the steepest ascent contour meets the real N-axis at $N = q_1/2$, where the background classical action is real.

Returning to our example of n subsection 5.1.4 of the Introduction, we analyzed the perturbation equations of motion. From the discussion below equation (5.2), it follows that the scaling exponent governing the perturbation $\gamma = l + 1$, so that the tensor modes are proportional to $t^{l/2} = r^l$. This is exactly the right scaling with r to make them analytic at $r = 0$. (If a tensor quantity has l indices and is expressible in terms of a tensor product of l Cartesian coordinates, it must necessarily scale as r^l.) The perturbation exponent is easily calculated from equation (5.2).

Putting everything together, we find the causal propagator to create a perturbed three-sphere of radius R in reduced Planck units "from nothing," when $\Lambda = 0$, is

$$G^{\Lambda=0}[q_1, \phi_1; 0] \sim e^{(-6\pi^2 + l\phi_1^2/2)R^2/\hbar}, \tag{5.66}$$

up to prefactors. As discussed in the Introduction, performing the semiclassical path integral over the perturbations generates a branch cut on the real N-axis, over $q_1/(2\sqrt{l(l+2)}) < N < \infty$ (and similarly for negative N). On the upper side of this branch cut, $\gamma = -i\sqrt{4l(l+2)N^2/q_1^2 - 1}$ and the real part of the semiclassical exponent (the h-function) is positive: one finds $\Re[iS^{(2)}(N)/\hbar] = \frac{1}{2}(q_1\phi_1^2/\hbar)\sqrt{l(l+2) - 4}$ at $N = q_1/2$, where the steepest ascent contour from the background saddle meets the N-axis. This is greater than the h-function from perturbations at the saddle, for all ϕ_1 and l. As in the de Sitter example, we see that when we integrate over the perturbations the net semiclassical exponent can become positive, at least when backreaction is ignored. Based on the results of Section 5.5, we do not expect the inclusion of nonlinear backreaction to change this conclusion.

The fact that this calculation, like that for de Sitter, yields an unacceptable inverse Gaussian distribution for the perturbations implies, we believe, that one should not consider "no-boundary" amplitudes of this type, even for $\Lambda = 0$.

5.D Behavior of solutions and classical action as $N \to N_\star$

At $N_\star = \sqrt{3q_1/\Lambda}$ the classical background solution (5.9) takes the simple form $\bar{q}(t) = q_1 t^2$. The equation of motion of the canonically normalised perturbation $\chi \equiv \bar{q}(t)\phi$, equation (5.21), has two solutions

$$\chi_{\star\pm} = t\left(t \mp iN_\star\sqrt{l(l+2)}/q_1\right) e^{\pm iN_\star\sqrt{l(l+2)}/(q_1 t)}. \tag{5.67}$$

The choice of mode is fixed by the finite action condition and depends on how we approach the point $N_\star$. When approached from above, the finite action solution is $\phi(t) = \phi_1\chi_{\star+}(t)/(t^2\chi_{\star+}(1))$. The corresponding classical action is

$$\bar{S}^{(2)}(N_\star) = \left.\frac{\bar{q}^2\bar{\phi}\dot{\bar{\phi}}}{2N}\right|_{t=0}^{1} = -\frac{iN_\star}{i+\sqrt{l(l+2)}\,N_\star/q_1}\frac{l(l+2)}{2}\phi_1^2\,. \tag{5.68}$$

When approached from below, the perturbation and action are given by the complex conjugate expressions. In both cases, the action agrees with $\bar{S}^{(2)}[q_1;\phi_1;N]$, given in equation (5.28), evaluated just above or below $N = N_\star$ in the complex N-plane, taking into account the analytic properties of $\gamma(N)$ explained in Appendix 5.B.

Chapter 6

Inconsistencies of the new no-boundary proposal

> The aim of physics is to provide a mathematical model of the universe which will agree with all observations that have been made so far and which will predict the result of further observations. Our present models would consist of two parts: (i) a set of differential equations that govern the variables in the theory. These are normally derived from an action principle; (ii) boundary conditions for the differential equations, or for the fields that are considered in the action principle. We have made a lot of progress on the first part of the problem in recent years and it now seems possible that we might find a fully unified field theory within the not-too-distant future. However, we shall not have a complete model of the universe until we can say more about the boundary conditions than that they must be whatever would produce what we observe.
>
> Stephen Hawking

Abstract

In previous chapters, we have demonstrated that the path integral for *real, Lorentzian* four-geometries in Einstein gravity yields sensible results in well-understood physical situations, but leads to uncontrolled fluctuations when the "no-boundary" condition proposed by Hartle and Hawking is imposed. In order to circumvent our result, new definitions for the gravitational path integral have been sought, involving specific choices for a class

of *complex* four-geometries to be included. In their latest proposal, Diaz Dorronsoro *et al.* [117] advocate integrating the lapse over a complex circular contour enclosing the origin. In this note we show that, like their earlier proposal, this leads to mathematical and physical inconsistencies and thus cannot be regarded as a basis for quantum cosmology. We also comment on Vilenkin and Yamada's recent modification of the "tunneling" proposal, made in order to avoid the same problems. We show that it leads to the breakdown of perturbation theory in a strong coupling regime.

6.1 Introduction

The no-boundary proposal of Hartle and Hawking [188] has been an influential idea in theoretical cosmology for more than three decades, and with good reason: it puts forth a proposal for the initial state of the universe, from which – assuming some set of physical laws – everything else is supposed to follow. If true, it would do no less than explain the origin of space and time. What is more, the proposal involves only semi-classical gravity, *i.e.*, a theoretical framework already within reach of contemporary physics, without requiring the development of a full theory of quantum gravity. Given the promise and magnitude of this claim, it should be analyzed with great care. In previous chapters 3 and 4, we attempted to put the no-boundary proposal on a sound mathematical footing by defining the gravitational path integral more carefully. Unfortunately, we found as a consequence that the no-boundary proposal leads to a universe with fluctuations which are out of control. Our work led Diaz Dorronsoro *et al.* to propose a new definition of the no-boundary proposal, involving an inherently complex contour in the space of four-metrics, *i.e.*, one which cannot be deformed to an integral over real four-metrics and hence has no geometrical interpretation. In particular, they chose to integrate the lapse N over a complex contour running below the origin in the complex N-plane, from negative to positive infinite real values [104]. In chapter 5, we demonstrated the inconsistency of this proposal. Very recently, Diaz Dorronsoro *et al.* have proposed yet another definition of the no-boundary proposal, this time in a particular truncation of Einstein gravity and taking instead a complex contour for the lapse which encircles the origin [117]. In this chapter we show that this latest incarnation of the no-boundary idea also leads to physical and mathematical inconsistencies.

The instability that we demonstrated applies equally well to the tunneling proposal developed by Vilenkin starting around the same time as the no-boundary proposal [325, 323, 329]. After we posted the original preprint version of the paper corresponding to this chapter, Vilenkin and Yamada proposed a modification of the tunneling proposal

in an attempt to rescue it [330]. Their new tunneling proposal involves the addition of a boundary term to the action, which has the consequence of selecting a different perturbation mode. As we explain in appendix 6.B, this amendment unfortunately introduces a strong coupling problem and, at present, its purported consequences cannot therefore be trusted.

6.2 Physical motivation

The path integral over four-geometries provides a well-motivated framework for the study of semi-classical quantum gravity. In analogy with Feynman's path integral formulation of quantum mechanics, one attempts to define transition amplitudes between two three-geometries $h^{(0)}_{ij}, h^{(1)}_{ij}$ by summing over all four-geometries that interpolate between the initial $h^{(0)}_{ij}$ and final boundary $h^{(1)}_{ij}$, *i.e.*,

$$G[h^{(1)}_{ij}; h^{(0)}_{ij}] = \int_{\partial g = h^{(0)}_{ij}}^{\partial g = h^{(1)}_{ij}} \mathcal{D}g e^{iS[g]/\hbar}\,, \tag{6.1}$$

where g denotes the four-metric. In this note, as in the work of Diaz Dorronsoro *et al.*, we study a simplified model in which $S[g]$ is taken to be the usual action for Einstein's theory of gravity plus a positive cosmological constant Λ.

In the previous chapters 3 and 5 we demonstrated that, somewhat to our surprise, the path integral, over *real, Lorentzian* four-geometries yields well-defined and unique results *as it stands*, when evaluated semiclassically and in cosmological perturbation theory, *i.e.*, when we treat the four-geometry as a homogeneous, isotropic background with small, but otherwise generic, perturbations. In contrast, we found the path integral over Euclidean four-geometries (as originally advocated by Hartle and Hawking [188]), even at the level of the homogeneous, isotropic background, to be a meaningless divergent integral. The key to our work was the use of Picard-Lefschetz theory, a powerful mathematical framework that allows one to rewrite highly oscillatory and only conditionally convergent integrals (such as (6.1) turns out to be) as absolutely convergent integrals. To do so, one regards the integral (6.1) as being taken over the subspace of real, Lorentzian metrics in the space of complex four-metrics. Cauchy's theorem, and Picard-Lefschetz theory, then allows one to deform the original, real integration domain into a complex domain consisting of one or more steepest descent thimbles. Each of these yields an absolutely convergent integral: their sum equals the original integral. Note that the analytical continuation to complex metrics and the deformation to steepest descent thimbles, are merely a convenient calculational tool, used to evaluate the original, uniquely defined but only conditionally convergent integral using

steepest descent methods. To make this point very clear, we provide in appendix 6.A an explicit proof of the convergence of integrals of the type we encounter, taken over real values of the lapse function N. Thus, one can prove the original integrals exist, and only subsequently use Picard-Lefschetz theory to evaluate them.

One frequently raised question is the range over which the lapse N should be integrated over in the path integral. The Lorentzian four-geometries we consider may be parameterized with the line element $-N^2(t,x)dt^2 + h_{ij}(t,x)dx^i dx^j$, where $0 \leq t \leq 1$ is a good time-like coordinate, *i.e.*, a one to one, invertible map from the manifold into the closed unit interval. The lapse N accounts, for example, for the proper time interval τ between two spacetime points (t_1, x^i) and (t_2, x^i), both at fixed x^i: one has $\tau = \int_{t_1}^{t_2} N(t,x^i)dt$. Note that the coordinate t already defines an orientation for the integral: the lapse N is simply a local rescaling, which must therefore be taken strictly positive as long as the coordinate chart and the manifold are both nonsingular. Stated more generally, assigning a non-singular coordinate system to a four-manifold already introduces an orientation, allowing one to define integrals such as the action or measures of volume, area or length. Writing the metric as usual by $g_{\mu\nu} = e^A_\mu e^B_\nu \eta_{AB}$, with e^A_μ the frame field and η_{AB} the Minkowski metric, only one continuously connected component of non-singular frame fields e^A_μ – for example the component with strictly positive eigenvalues – is needed in order to describe a general, nonsingular four-geometry. To sum over additional components (for example to sum over both positive and negative lapse functions N while taking the determinant h to be positive) is not only unnecessary, it represents an overcounting which is unjustified from a geometrical point of view. Furthermore, although arbitrarily small N should be allowed, one should *not* include the point $N = 0$ in the sum since it does not describe a four-geometry. Finally, integrating over all Lorentzian four-geometries requires only *real* (and positive) values of N. If that fundamental, geometrical definition can be deformed into a mathematically equivalent integral over complex metrics which is easier to calculate, as Picard-Lefschetz theory and Cauchy's theorem allow, that is all well and good. But it makes little geometrical sense to take an integral over complex lapse functions N as a fundamental definition of the theory.

In their most recent paper, Diaz Dorronsoro *et al.* [117] misrepresent our work by stating that we "have recently advanced a larger class of wavefunctions that extend the original" no-boundary wavefunction. Quite to the contrary, what we explained in our earlier papers is that the integral over Lorentzian four-geometries is actually *unique*! This allowed us to compute *the only* geometrically meaningful "no-boundary wavefunction." The fact that calculation failed to give an observationally acceptable result is not the fault of the path integral for gravity, but rather that of imposing the "no-boundary" idea in this particular model, attempting to describe the beginning of the universe in the context of inflationary

scenarios.

In fact, it is Diaz Dorronsoro *et al.*, not us, who are "advancing a larger class of wavefunctions" in an attempt to rescue the no-boundary proposal. As we have explained, there is no geometrical justification for taking an integral over complex metrics as a starting point for the theory. Yet this is exactly what they propose [117]. They consider metrics of the biaxial Bianchi IX form

$$ds^2 = -\frac{N^2}{q}dt^2 + \frac{p}{4}(\sigma_1^2 + \sigma_2^2) + \frac{q}{4}\sigma_3^2\,, \tag{6.2}$$

where $p(t), q(t)$ are time dependent scale factors and $\sigma_1 = \sin\psi d\theta - \cos\psi\sin\theta d\varphi$, $\sigma_2 = \cos\psi d\theta + \sin\psi\sin\theta d\varphi$, and $\sigma_3 = -(d\psi + \cos\theta d\varphi)$ are differential forms on the three sphere with $0 \leq \psi \leq 4\pi$, $0 \leq \theta \leq \pi$, and $0 \leq \varphi \leq 2\pi$. For real $N > 0$, this metric describes Bianchi IX spacetimes on the axes of symmetry. In the well-known notation of Misner [254] this corresponds to the line $\beta_- = 0$. Diaz Dorronsoro et al. now propose to define the gravitational path integral as a sum over real values of p and q, supplemented by a sum over values of the lapse function N, taken along a *complex* circular contour enclosing the origin.

In our view this proposal is quite arbitrary, as it is not motivated by any fundamental physical principle. What does it mean to integrate over metrics with complex proper time intervals? In [117], this sum over specific complex metrics is regarded not merely as a calculational device, but as the starting definition of the theory. Furthermore, this definition seems context dependent. Such a definition will neither allow one to calculate meaningful transition amplitudes between two large three-geometries nor to understand how quantum field theory on curved spacetime emerges when the scale factor evolves classically. Given its poor motivation, we find it unsurprising that this definition ultimately leads to mathematical and physical inconsistencies, as we shall explain in the remainder of this note.

Before doing so, it may be useful to briefly comment on the relation of the path integral to the propagator and the Wheeler-DeWitt equation. For concreteness, consider a simple relativistic path integral such as is encountered for a relativistic particle. Formally, the starting point is the relativistic propagator,

$$G[q_1, q_0] = \langle q_1| \int_{0^+}^{\infty} \mathrm{d}N e^{-i\hat{\mathcal{H}}N} |q_0\rangle, \tag{6.3}$$

where $\hat{\mathcal{H}}$ is the Hamiltonian, where the rhs can be expressed as a path integral in the usual way. Applying the operator $\hat{\mathcal{H}}$ and rewriting $\hat{\mathcal{H}}e^{-i\hat{\mathcal{H}}N} = i(\mathrm{d}e^{-i\hat{\mathcal{H}}N}/\mathrm{d}N)$, the N integral

becomes a boundary term at $N = \infty$, which may be taken to vanish, minus another at $N = 0^+$, proportional to $\langle q_1 | q_0 \rangle = \delta(q_1 - q_0)$. In this way one obtains

$$\hat{\mathcal{H}} G[q_1, q_0] = -i\delta(q_1 - q_0). \tag{6.4}$$

The proposal of [117] is to instead obtain a homogeneous solution of the Wheeler-DeWitt equation $\hat{\mathcal{H}}\Psi = 0$ from a similar formula

$$\Psi[q_1] = \langle q_1 | \int_{\mathcal{C}} \mathrm{d}N e^{-i\hat{\mathcal{H}}N} | \chi \rangle, \tag{6.5}$$

where χ is any state, and $\mathcal{C}$ is a contour which yields no endpoint contributions. For example, $\mathcal{C}$ may start and end at infinity, or it may be closed. Whereas the propagator (6.3) is uniquely defined, (6.5) in principle depends both on the state $|\chi\rangle$ and the contour $\mathcal{C}$. This infinite ambiguity is related to the fact that there are infinitely many homogeneous solutions of the Wheeler-DeWitt equation. In order to define a "wavefunction of the universe," some other physical or mathematical principles are needed. In the model they study, Dorronsoro *et al.* take for $\mathcal{C}$ a small contour in the complex N-plane enclosing the origin. As already noted, there is little justification for this choice. Furthermore, it immediately leads to a problem with the path integral. Since there is no singularity in N in the integrand of (6.5), at fixed $p(t)$ and $q(t)$, there is no obstruction to shrinking the N contour away. This means that if one performs the N integral first, the answer is zero! Diaz Dorronsoro *et al.* do not notice this because they perform the path integrals over $p(t)$ and $q(t)$ first, generating a pole in N from the corresponding prefactors. Then the N integral, taken on a closed contour enclosing the origin, extracts the residue. Clearly, their result depends on the order in which the partial integrals of the path integral are taken. While Dorronsoro *et al. do* generate a solution to the Wheeler-DeWitt equation this way, any connection to the original path integral is clearly on shaky ground. There are additional (and related) problems with their definition to which we return at the end of the next section.

6.3 Normalizability

With the metric (6.2), the action for gravity plus a cosmological constant Λ, in units where $8\pi G = 1$, is given by

$$S/(2\pi^2) = \int_0^1 \mathrm{d}t \left[-\frac{1}{4N} \left(\frac{q\dot{p}^2}{p} + 2\dot{p}\dot{q} \right) + N \left(4 - \frac{q}{p} - p\Lambda \right) \right], \tag{6.6}$$

where the integrals over angular directions yield a factor of $16\pi^2$. In this section we evaluate the classical action. Then we apply Picard-Lefschetz theory to identify the relevant saddles and deform the N integral to render it absolutely convergent. We also discuss the normalizability of the resulting "wavefunction".

6.3.1 The classical action

The equations of motion corresponding to the variations of q and p are given by

$$2p\ddot{p} - \dot{p}^2 = 4N^2\,, \qquad \ddot{q} + \frac{\dot{p}}{p}\dot{q} = N^2\left(2\Lambda - \frac{4q}{p^2}\right)\,, \tag{6.7}$$

and the constraint following from the variation of N is given by

$$\frac{1}{4}\left(\frac{q}{p}\dot{p}^2 + 2\dot{p}\dot{q}\right) + N^2\left(4 - \frac{q}{p} - p\Lambda\right) = 0\,. \tag{6.8}$$

Regular solutions to these equations, behaving as $p(t) \sim q(t) \sim \pm 2iNt$ as $t \to 0$, correspond to (part of) Taub-NUT-de Sitter spacetime. The corresponding complex, regular geometries are considered to be of the no-boundary type [117]. We focus on these in what follows.

Since the Lagrangian in (6.6) is linear in $q(t)$, the path integral over $q(t)$ enforces a functional delta function for first equation of motion in (6.7). Therefore the only paths $p(t)$ which contribute to the path integral are those which satisfy this equation. Using it, the action reduces to:

$$S/(2\pi^2) = \int_0^1 \mathrm{d}t\left(-\frac{1}{2N}\frac{\mathrm{d}}{\mathrm{d}t}\left(\dot{p}q\right) + N\left(4 - p\Lambda\right)\right), \tag{6.9}$$

so that the classical action depends on $q(t)$ only through its boundary values.

In order to implement the no-boundary proposal, as explained above we take $p(0) = q(0) = 0$. We also set $p(1) = p_1$, $q(1) = q_1$ where p_1 and q_1 are arbitrary positive constants, to describe the final, anisotropic three-geometry. With these boundary conditions, the equation of motion for p has no real solution. There is, however, a pair of complex conjugate solutions,

$$p_\pm(t) = \pm iNt(t-1) + p_1 t^2. \tag{6.10}$$

for which the corresponding classical action (6.9) is given by

$$S_{\pm}(N)/(2\pi^2) = -\frac{p_1 q_1}{N} \pm i q_1 + N\left(4 - \frac{\Lambda}{3}p_1\right) \mp i\frac{\Lambda}{3}N^2\,. \tag{6.11}$$

We claim that the original, Lorentzian path integral over positive real values of N is convergent. According to (6.11), after integrating out $p(t)$ and $q(t)$, for the two possible classical solutions (6.10), the semiclassical exponent $iS_{\pm}(N) \sim \pm\frac{\Lambda}{3}N^2$ at large N. In order for the N integral to converge, we must take $S_{-}(N)$, corresponding to the solution $p_{-}(t)$. This choice is actually in conflict with the "momentum constraint" imposed in [117] (in fact it corresponds to the opposite "momentum constraint"), but it is mandatory if one starts from the Lorentzian path integral. Nevertheless, we will also later consider their choice of solution, $p_{+}(t)$ with action S_{+}, in order to highlight some aspects of this choice. But we emphasize that it is incompatible with the Lorentzian path integral.

The corresponding propagators simplify to an oscillatory integral over the lapse, *i.e.*,

$$G_{\pm}[q_1, p_1; 0, 0] \propto \int \frac{\mathrm{d}N}{N} e^{iS_{\pm}[q_1,p_1;0,0;N]/\hbar}\,. \tag{6.12}$$

6.3.2 Picard-Lefschetz theory

Having reduced the path integral to an ordinary integral over the lapse function N, we are now in a position to evaluate it in the saddle point approximation. Figures 6.1 and 6.2 show the locations of the saddle points and steepest ascent/descent lines emanating from them for the two choices of the action given in (6.11). It is straightforward to see that the integral over real Lorentzian metrics, with semiclassical action S_{-}, can be deformed into the steepest descent contour $\mathcal{J}_1$ passing through saddle point 1. The location of this saddle point for various values of p_1 and fixed q_1 is shown in Fig. 6.3. For large anisotropies it moves closer and closer to the real N line, without however ever reaching it. The induced weighting is shown by the blue curve of the left panel in Fig. 6.4, where it can be seen that the isotropic boundary conditions (here $p_1 = q_1 = 10000$) receive the lowest weighting. In other words, the model is out of control, as more anisotropic geometries are favoured.

An even more dramatic failure of the model is seen by sending q_1 to large values. The second term in S_{-}, given in (6.11), contributes a semiclassical exponent $+q_1$ which clearly leads to a non-normalizable wavefunction. Writing the volume of the final three-geometry as $V \sim p_1 q_1^{\frac{1}{2}}$ (see (6.2)), and the anisotropy as $\alpha = q_1/p_1$, the terms which determine the saddle point value of N in $S_{-}(N)$ at large N (see (6.11)) are the first and the last.

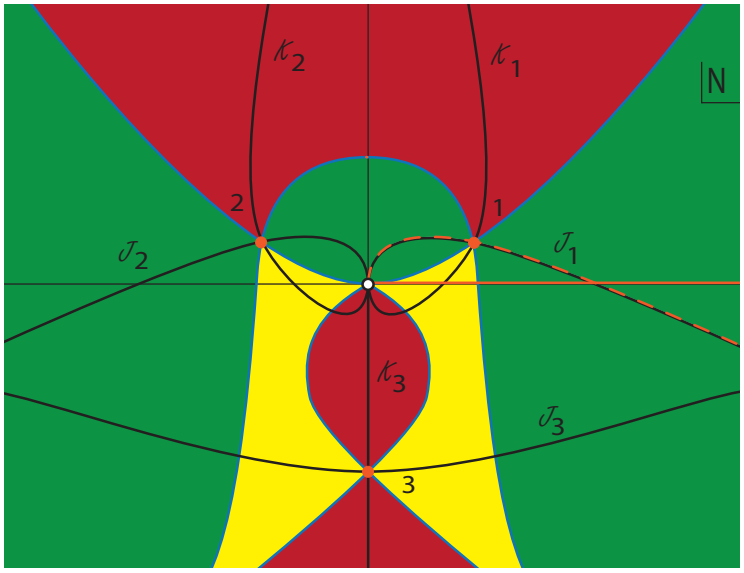


Figure 6.1: The location of the saddle points and flow lines for the action we advocate, S_- in (6.11), for which the Lorentzian integral is convergent. The saddle points are indicated by the orange dots. Green regions have a lower magnitude of the integrand than at the adjacent saddle point, red regions have a higher magnitude and yellow regions have a magnitude in between two saddle point values. If N approaches the singular point at infinity or the essential singularity at $N = 0$ along a contour in a green region, we obtain a convergent integral. Conversely, if N approaches these points along a contour in a red region, the integral diverges.

Evaluating the action at this saddle, one finds the second term dominates for anisotropy $\alpha > V^{\frac{1}{2}}\Lambda^{\frac{3}{4}}$. Defining the de Sitter radius $R_H \equiv \Lambda^{-\frac{1}{2}}$, one sees that the semiclassical exponent iS_- is dominated by the $+q_1$ term as long as the length in the 3 direction (see (6.2)), $q_1^{\frac{1}{2}} = \alpha^{\frac{1}{3}}V^{\frac{1}{3}}$ exceeds $V^{\frac{1}{2}}R_H^{-\frac{1}{2}}$ while the length in the other two spatial directions $p_1^{\frac{1}{2}} = V^{\frac{1}{3}}/\alpha^{\frac{1}{6}}$ is smaller than $V^{\frac{1}{4}}R_H^{\frac{1}{4}}$. Clearly, both lengths can grow without bound as V is increased while remaining in this regime. So the semiclassical exponent can become arbitrarily large, whilst all curvature scales remain greater than the Planck length so that semi-classical Einstein gravity remains valid.

These findings, that imposing "no-boundary" boundary conditions in the Lorentzian path integral lead to an unacceptable amplitude, favouring large deviations from isotropy, confirm those of our earlier analysis of inhomogeneous perturbations around the back-

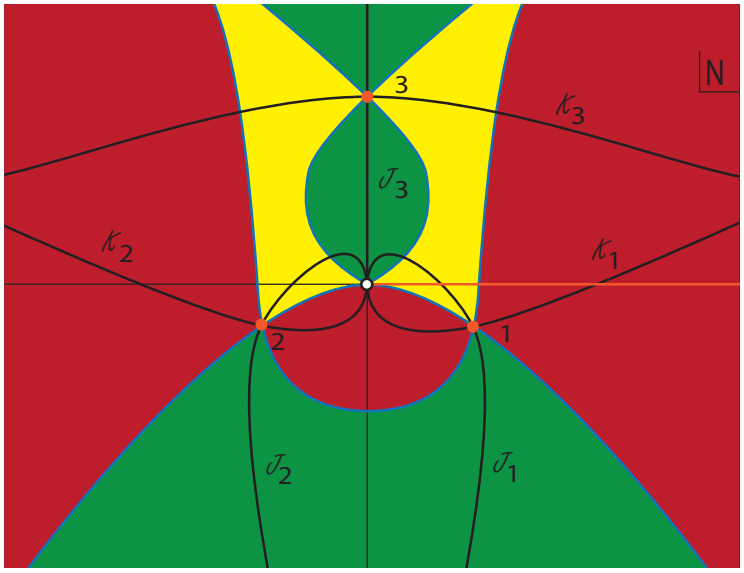


Figure 6.2: The location of the saddle points and flow lines for the action S_+, for which the Lorentzian path integral diverges, but which is chosen by Diaz Dorronsoro *et al.* [117]. For a description of the colour scheme, see the caption of Fig. 6.1. Note that, as this Figure shows, it would also have been possible to define a purely Euclidean contour along the positive imaginary axis for this choice of action, and this would have led to only saddle point 3 contributing. This latter saddle point leads to a purely Euclidean geometry, without any classical Lorentzian evolution.

ground FLRW cosmology (see chapters 4 and 5). The advantage of the present discussion is that it is fully non-perturbative, albeit still semi-classical. Hence, our analysis removes any hope that a treatment going beyond cosmological perturbation theory might yet rescue the no-boundary proposal. As explained previously, we adhere strictly to the Lorentzian formulation of the gravitational path integral which, we have argued, is the only one with a chance of making mathematical and physical sense.

As shown by Diaz Dorronsoro *et al.*, if one takes the integral over N along a circular contour around $N = 0$ for the action S_+ in (6.11), the contour can be deformed to a sum over the two steepest descent paths $\mathcal{J}_1$ and $\mathcal{J}_2$ in Fig. 6.2. These saddle points lie respectively at the complex conjugate and negative values of the Lorentzian saddle point 1 in Fig. 6.1. The weighting of these saddle points is just the inverse of the weighting of the Lorentzian saddle point, and is shown by the orange curve in the left panel of Fig. 6.4. For these the isotropic configuration $p_1 = q_1$ is indeed the configuration with the highest

weighting. However, having a maximum is not enough to ensure normalizability. Indeed, just as for the Lorentzian saddle point, the weighting of these saddle points tends to a constant at large values of p_1 (the inverse of a constant being another constant), so that again an integral of the weighting $e^{-2Im(S_+)/\hbar}$ over p_1 is unbounded, and the corresponding wavefunction is non-normalizable. Thus if normalizability is regarded as a crucial criterion, the new circular contour must also be discarded on these grounds.

For reasons that are not clear to us, the authors of [117], even though they also noticed the unboundedness of the integral, simply chose to truncate it by hand. The stated reason was that the approximations involved in the calculation break down. However, this statement is puzzling, as the axial Bianchi IX model allows one to calculate the action exactly and, moreover, the saddle point approximation becomes better and better at large p_1 (see again the right panel of Fig. 6.4, which also applies to the saddle points in question). Thus the implied non-normalizability seems robust, to the extent that normalizability is understood at all.

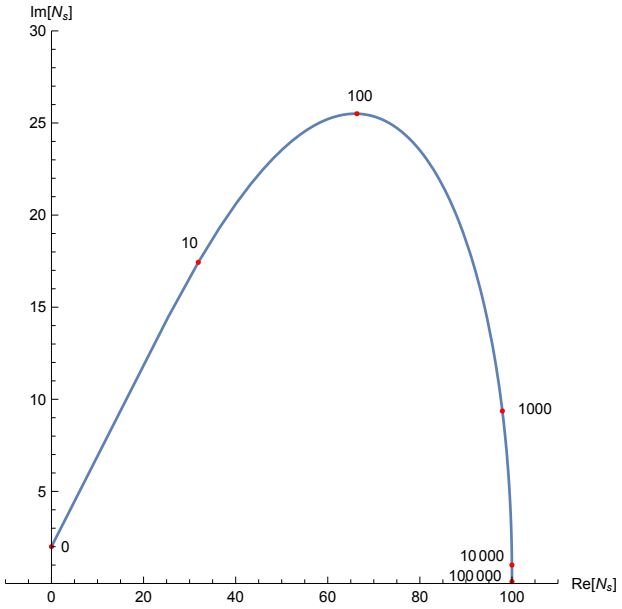


Figure 6.3: The location of the relevant saddle point, for fixed $q_1 = 10000$ and as a function of $0 < p_1 < 100000$. Some indicative values of p_1 are shown next to the curve. At large values of p_1 the saddle point remains complex but moves very close to the real N line.

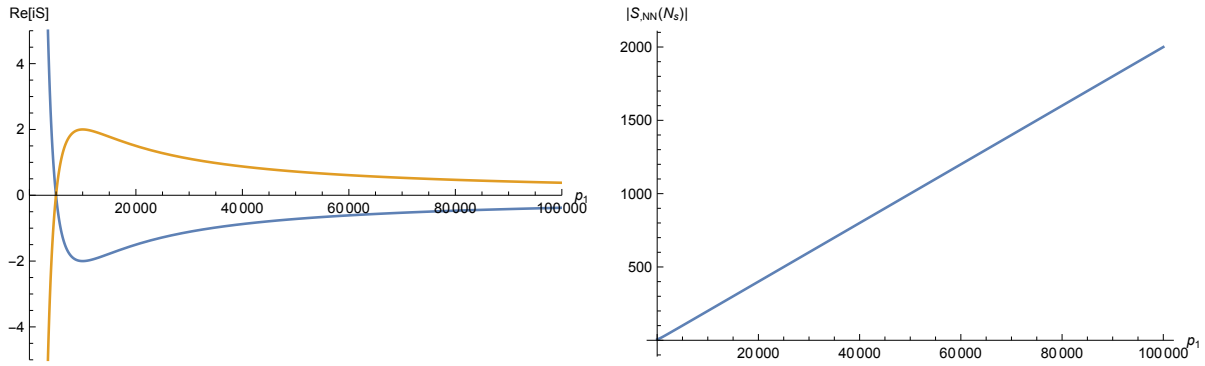


Figure 6.4: *Left:* The h-function $-Im(S)$ for fixed $q_1 = 10000$ and as a function of $0 < p_1 < 100000$ both for the relevant Lorentzian saddle point, with action $S_-(N)$, (blue) and for one of the saddle points advocated by Diaz Dorronsoro *et al.* (orange), with action S_+. *Right:* The absolute value of the second derivative at the same saddle points for fixed $q_1 = 10000$ (with $\Lambda = 3$) and as a function of $0 < p_1 < 100000$. For large p_1 the saddle point approximation becomes better and better.

6.4 Mathematical and physical consistency

We now come to what we regard as the biggest flaw in the proposal of Diaz Dorronsoro *et al.*, namely that it seems to us to fail some simple tests of physical and mathematical consistency. When we take the limit in which the final three-geometry is isotropic, it seems reasonable to expect that we should recover the result of the truncated, isotropic theory, at least in the semi-classical limit where quantum backreaction is negligible. Likewise, if we add an additional metric perturbation mode to the final three-geometry, for example one of an inaccessibly small wavelength, this should not immediately lead to inconsistent results. We will discuss these two tests of their proposal, in turn.

First, consider the isotropic limit, where $p_1 = q_1$. Here we would expect the axial Bianchi IX model to reproduce the results of an isotropic FLRW minisuperspace model, defined using the same integration contour for the lapse function. Certainly, for Lorentzian integrals, this is the case and the action S_- in Eq. (6.11) indeed reproduces our earlier isotropic results of chapter 3. When $p_1 = q_1$, the relevant saddle point of the action S_- is located at

$$N_{s1}^{iso} = \sqrt{\frac{3}{\Lambda}}\sqrt{q_1 - \frac{3}{\Lambda}} + i\frac{3}{\Lambda}\,, \tag{6.13}$$

i.e., it resides at the same value of N as for the isotropic model, where the action is given

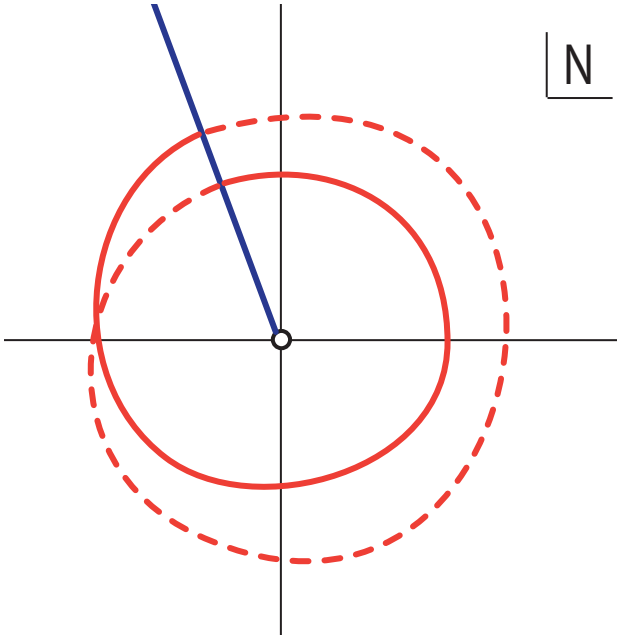


Figure 6.5: An example of a circular contour in the presence of a branch cut (in blue), for a two-sheeted integrand. The contour must complete two loops before it can close.

by a different function of N, namely [186] and chapter 3.

$$S^{iso}(N)/2\pi^2 = \left[N^3\frac{\Lambda^2}{36} + N\left(3 - \frac{\Lambda}{2}q_1\right) - \frac{3q_1^2}{4N}\right] . \tag{6.14}$$

Moreover, at the isotropic saddle point (6.13), the values of the axial Bianchi IX action S_- and the isotropic action (6.14) agree,

$$S_{conv}(N_{s1}^{iso}) = S^{iso}(N_{s1}^{iso}) = 2\pi^2\left(-\frac{2\sqrt{3}}{\Lambda}(\Lambda q_1 - 3)^{3/2} + i\frac{6}{\Lambda}\right) . \tag{6.15}$$

Thus we find a well-behaved isotropic limit, as we believe we should, since in the isotropic limit we are describing the same physical situation.

However, when we take the circular contour advocated by Diaz Dorronsoro *et al.*, a problem arises. In the isotropic case, the path integral reduces to an ordinary integral over

the lapse function of the form [186] and chapter 3.

$$G[q_1, 0] = \sqrt{\frac{3\pi i}{2\hbar}} \int \frac{\mathrm{d}N}{\sqrt{N}} e^{iS^{iso}(N)/\hbar} \,. \tag{6.16}$$

The prefactor, which arises from the integral over the isotropic scale factor, contains a factor of $1/\sqrt{N}$, so that there is a branch cut in the integrand, emanating from the origin. This branch cut requires that a circular contour must complete two loops around the origin before it can close – see Fig. 6.5. However, on the second loop the factor $1/\sqrt{N}$ will acquire a minus sign relative to its value on the first loop, so that the contributions from the second loop exactly cancel those of the first loop. The result is that, for isotropic boundary conditions, a closed circular contour yields precisely zero! Hence there is blatant disagreement with the isotropic limit of the Bianchi IX model, although the physical situation being described is identical. (One may easily verify that the saddle points contributing to the path integral with final boundary $p_1 = q_1$ also have $p(t) = q(t)$ throughout the entire geometry $0 \leq t \leq 1$). Hence this choice of contour fails to satisfy our consistency check.

The second inconsistency manifests itself in the following, closely related manner. In minisuperspace models, when we include n deformations of the metric in addition to the lapse, the prefactor generally takes the form $1/N^{n/2}$ [186]. For n odd, the integrand will thus be taken around a branch point at $N = 0$ and a closed contour about the origin will again yield a vanishing result. But the results of our calculations should not depend on how many possible deformations we include as long as the same physical situation is described. One should be able to add a possible deformation and then consider boundary conditions in which this additional deformation is zero – and the results should, at this leading semi-classical level, be unchanged. A straightforward example is to use the full Bianchi IX metric and then restrict to boundary conditions corresponding to the axial Bianchi IX truncation studied in this chapter. Once again this does not lead to consistent results, as the Bianchi IX metric contains one additional deformation, so that a closed contour enclosing the origin again leads to a vanishing wavefunction.

6.5 Discussion

When constructing theories of the very early universe, the difficulty of making direct observations means that mathematical and physical consistency requirements must necessarily play a critical, guiding role. In our view, the new path integral for semi-classical gravity advocated by Diaz Dorronsoro *et al.* [117], involving a closed integral for the complexified lapse function, seems inadequate in this regard: it has no geometrical interpretation as it

involves metrics with complex proper times. Likewise, it abandons any notion of causality from the outset. Furthermore, when describing the same physical situation using different truncations of the degrees of freedom in the spacetime metric, it yields vastly different results. The clearest example is provided by truncating the model to an isotropic, one-dimensional minisuperspace, for which a closed contour about the origin yields a vanishing "wavefunction." More generally, such a closed contour fails to yield a meaningful wavefunction for any odd-dimensional truncation of minisuperspace – violating the physically reasonable requirement that including one additional mode, for example one with an inaccessibly tiny wavelength, should not change a physical result. A general formal argument against ad hoc contours of the type Diaz Dorronsoro *et al.* consider was given at the end of Section II. Unless a closed contour starts and ends at the point at infinity in the complex N-plane, it cannot give a nonzero result unless the path integral depends on the order in which the integrals are taken. Such a theory is clearly ambiguous at best.

In previous chapters, we have shown that attempts to define a "smooth beginning" for inflation based on either the no-boundary proposal [188] or the tunneling proposal [325, 323, 329] are either mathematically inconsistent or they lead to the physically unacceptable results. The semiclassical "no-boundary" path integral taken over real Euclidean metrics fails in the first manner, whereas the path integral taken over real Lorentzian metrics, as posited in the "tunneling proposal," fails in the second since it favors wildly fluctuating geometries. (In appendix 6.B we show that the recent rescue of the tunneling proposal proposed by Vilenkin and Yamada [330]) unfortunately fails due to the breakdown of perturbation theory, *i.e.*, a strong coupling problem.) At the root of this disaster are two key assumptions; i) that it makes sense to compute an amplitude for an "out" state when there is no "in" state (or when the "in" state is replaced by a "three-geometry of zero size") and ii) that the universe started out dominated by some kind of inflationary energy, which behaves in effect like a large, temporary cosmological constant. It is not yet clear to us whether abandoning one of these assumptions would lead to a more acceptable result. Conceivably, one should abandon both (see, *e.g.*, [163, 164]).

As well as these negative conclusions, our investigations have also opened up a very interesting avenue to pursue. We have found many indications that the Lorentzian path integral for gravity, tackled consistently using Picard-Lefschetz theory and with sensible "in" and "out" states, has a remarkable physical and mathematical uniqueness and consistency. In future work, we shall outline what we consider to be a far less arbitrary and more promising approach to the problem of the initial conditions for the universe, based on a precise treatment of this gravitational path integral.

6.A Conditionally convergent integrals

Highly oscillatory integrals, such as the Fresnel integrals, play an important role in physics. They are of particular interest in Lorentzian quantum physics, as they arise naturally in the form of real-time path integrals. In minisuperspace models of Lorentzian quantum cosmology, the integral over the lapse function is quite generically of this type. In this appendix we will spell out a simple proof that such integrals, defined over real values of the lapse, are convergent, using only real field values and without requiring a complexification as is used in Picard-Lefschetz theory. Our proof will simply demonstrate the convergence without however showing what value the integral converges to. Then, as used earlier, one may conveniently use Picard-Lefschetz theory to obtain a saddle point approximation to the integral.

The integral of the function g over an infinite domain is defined as

$$\int_0^\infty \mathrm{d}x\ g(x) = \lim_{R\to\infty} \int_0^R \mathrm{d}x\ g(x)\,. \tag{6.17}$$

Assuming that the integral is convergent, the integral is called absolutely convergent when the integral over the magnitude of the integral converges

$$\int_0^\infty \mathrm{d}x\ |g(x)| < \infty\,. \tag{6.18}$$

The integral is called conditionally convergent when the integral over the magnitude diverges

$$\int_0^\infty \mathrm{d}x\ |g(x)| = \infty\,. \tag{6.19}$$

Conditionally convergent integrals, converge due to cancellations from violent oscillations for large x. It is important to note that for conditionally convergent integrals, as for conditionally convergent series, the result depends on the order of summation. For integrals this is naturally prescribed by definition (6.17). By changing the order of summation one can engineer the integral to converge to any number. However, for such a deformation of the theory one can no longer use complex continuations and deformations of the integration contour in the complex plane to evaluate these integrals. This would in particular invalidate the commonly used $i\epsilon$ and Wick rotation methods in quantum physics. For this reason we stick to the natural definition and argue that conditionally convergent integrals are well-defined in real time quantum physics.

For many oscillatory integrals, convergence can be demonstrated with the Leibniz convergence test for alternating series. A real alternating series is defined as

$$a = \pm \sum_{i=0}^{\infty} (-1)^i a_i \,, \tag{6.20}$$

with a_i positive real numbers. The Leibniz convergence test states that the series is convergent when the arguments decrease monotonically, *i.e.*, $a_{i+1} \leq a_i$ for sufficiently large i, and the argument goes to zero in the limit of large i, *i.e.*, $\lim_{i\to\infty} a_i = 0$. To see the relation to oscillatory integrals, consider the integral

$$I = \int_0^\infty \mathrm{d}x \; e^{if(x)} \,, \tag{6.21}$$

for a real valued polynomial f (for the Fresnel integrals $f(x) = x^2$). The real and imaginary parts of I are given by

$$\mathrm{Re}[I] = \int_0^\infty \mathrm{d}x \;\; \cos(f(x)) \,, \tag{6.22}$$

$$\mathrm{Im}[I] = \int_0^\infty \mathrm{d}x \;\; \sin(f(x)) \,. \tag{6.23}$$

For simplicity we concentrate on the real part. Let us assume that the leading term of f goes like x^n in the limit $x \to \infty$ for $n \in \mathbb{N}$. A change of coordinates $u = x^n$ gives the integral

$$\mathrm{Re}[I] = \int_0^\infty \frac{\mathrm{d}u}{n u^{1-1/n}} \;\; \cos(f(\sqrt[n]{u})) \,, \tag{6.24}$$

and ensures that $f(\sqrt[n]{u}) \sim u$ for large u. Now let the zero crossings of the argument be given by z_i for $i \in \mathbb{N}$. The real part of the integral can be written as an alternating series

$$\mathrm{Re}[I] = \left[\int_0^{z_0} + \sum_{i=0}^{\infty} \int_{z_i}^{z_{i+1}} \right] \frac{\mathrm{d}u}{n u^{1-1/n}} \;\; \cos(f(u^{1/n})) \,, \tag{6.25}$$

$$= c \pm \sum_{i=0}^{\infty} (-1)^i \left| \int_{z_i}^{z_{i+1}} \frac{\mathrm{d}u}{n u^{1-1/n}} \;\; \cos(f(u^{1/n})) \right| \,, \tag{6.26}$$

$$= c \pm \sum_{i=0}^{\infty} (-1)^i a_i \,, \tag{6.27}$$

without changing the order of summation, with c the integral over the interval $(0, z_0)$, either the positive or the negative sign for $\pm$ depending on the details of f, and the positive real numbers

$$a_i = \left| \int_{z_i}^{z_{i+1}} \frac{\mathrm{d}u}{n u^{1-1/n}} \cos(f(u^{1/n})) \right| . \tag{6.28}$$

The argument of the alternating series can be dominated with a simple approximation

$$a_i = \left| \int_{z_i}^{z_{i+1}} \frac{\mathrm{d}u}{n u^{1-1/n}} \cos(f(u^{1/n})) \right| \tag{6.29}$$

$$< \int_{z_i}^{z_{i+1}} \frac{\mathrm{d}u}{n u^{1-1/n}} = \sqrt[n]{z_{i+1}} - \sqrt[n]{z_i} = b_i \, . \tag{6.30}$$

In the limit of large u the function $f(\sqrt[n]{u})$ asymptotes to a function proportional to u. For this reason, in the limit of large i, the zero crossings z_i in u will asymptote to a regular spacing, leading to the conclusion that for $n > 1$ and for sufficiently large i, the coefficients b_i satisfy the conditions of the Leibniz convergence test. Since $a_i < b_i$ for all i we conclude that $\mathrm{Re}[I]$ converges when $n > 1$. A similar argument can be given for the imaginary part of I, making I conditionally convergent.

The discussion above applies to a more general class of integrals. When the integral function $f(x)$ diverges as x^{-n} in the limit $x \to 0$ with $n \in \mathbb{N}$, the change of coordinates $u = x^{-n}$ leads to convergence for $n > 1$. More generally, when f is not a polynomial but dominates some polynomial x^n with $n > 1$, the oscillatory integral can be shown to converge due to cancellations from oscillations at large x. Note that generally one should include the prefactor in the analysis.

The propagator for axial Bianchi IX consists of an oscillatory integral over the lapse N

$$G[q_1, p_1; q_0, p_0] \propto \int_0^\infty \frac{\mathrm{d}N}{N} e^{iS[q_1,p_1;q_0,p_0;N]/\hbar} \, , \tag{6.31}$$

where the classical action $S[q_1, p_1; q_0, p_0; N] \sim N^2$ as $N \to \infty$ and $\sim N^{-1}$ for $N \to 0$. The discussion above directly shows that the integral converges at large N. The behavior of the integral for small N is more subtle, since the discussion of the Leibniz convergence test is agnostic about polynomials for which $n = \pm 1$. By including the prefactor in the analysis, we now show that convergence is guaranteed. Consider the real integral

$$I = \int_0^1 \frac{e^{i/x}}{x} \mathrm{d}x \, . \tag{6.32}$$

The integral does not converge absolutely. However, by a change of variables $u = -\ln x$, we can write the integral as

$$I = \int_0^\infty e^{ie^u} \mathrm{d}u \,. \tag{6.33}$$

We treat the real and imaginary part of the integral separately and for simplicity concentrate on the real part. The real part can be written as an alternating series

$$\mathrm{Re}[I] = \int_0^\infty \cos(e^u)\mathrm{d}u = \left[\int_0^{z_0} + \sum_{i=0}^{\infty} \int_{z_i}^{z_{i+1}}\right] \cos(e^u)\mathrm{d}u \tag{6.34}$$

$$= c - \sum_{i=0}^{\infty} (-1)^i \left| \int_{z_i}^{z_{i+1}} \cos(e^u)\mathrm{d}u \right| , \tag{6.35}$$

where $z_i = \ln\left(\left(i + \frac{1}{2}\right)\pi\right)$ are the roots of the integrand. The arguments of the sum satisfies the Leibnitz condition since

$$\left| \int_{z_i}^{z_{i+1}} \cos(e^u)\mathrm{d}u \right| \leq z_{i+1} - z_i = \ln\left(\frac{2i+3}{2i+1}\right) . \tag{6.36}$$

Since the summands satisfy Leibniz condition for alternating series, the real part of the integral exists. Convergence of the imaginary part follows analogously. We thus conclude that the Lorentzian propagator for axial Bianchi IX is well defined as a conditionally convergent integral.

6.B Remark on the new tunneling proposal

The "tunneling proposal" of Vilenkin [325] is closely related to the no-boundary proposal: it proposes to view the origin of the universe as a smooth tunneling event from "nothing," defined as a three-geometry of "zero size." If one imposes that in the limit where the initial three-geometry is taken to zero size, the resulting semiclassical saddle point solution should be a regular complex solution of the Einstein equations, then there is no associated boundary term and the tunneling proposal leads to exactly the same path integral as that we consider for the no-boundary proposal (as explained in chapter 3). Consequently the tunneling proposal suffers from exactly the same instability. However, there is an ambiguity in the tunneling proposal's prescription of taking the initial three-geometry to have zero size: one could allow it to have large fluctuations in its local curvature or "shape." In an attempt to rescue the tunneling proposal, Vilenkin and Yamada have attempted to

exploit this ambiguity by adding a boundary term for the fluctuations on the initial three-geometry, which they precisely tune in order to control the distribution for the fluctuations on the final three-geometry [330].

In their paper, they work in linear perturbation theory, where the fluctuations are regarded as small perturbations around the smooth, complex, classical background saddle point solution. The equation of motion for a Fourier mode of the perturbation ϕ_k (with wavenumber k) is of second order and it admits two solutions, which can be thought of as the positive and negative frequency modes, as usual. For instance, at the de Sitter saddle point, the general solution is given by a linear combination of two modes,

$$\phi_k(\tau) = c_1 \frac{(1-\cos(H\tau))^{(k-1)/2}(\cos(H\tau)+k)}{(1+\cos(H\tau))^{(k+1)/2}} + c_2 \frac{(1+\cos(H\tau))^{(k-1)/2}(\cos(H\tau)-k)}{(1-\cos(H\tau))^{(k+1)/2}}\,, \tag{6.37}$$

which represent the analogue of the Bunch-Davies solutions for the closed slicing of de Sitter space. Here we have expressed the solution in terms of Euclidean time τ, so that the scale factor is given by $a = \frac{1}{H}\sin(H\tau)$. Near the origin $a = 0$ (which also corresponds to the origin of τ, with $a \approx \tau$ for small τ) these two modes behave very differently: the mode proportional to c_1 tends to zero as τ^{k-1}, while the mode proportional to c_2 tends to infinity as τ^{-k-1}. Since the action of the second mode diverges due to contributions near $a = 0$, we have discarded this mode in chapter 5. The mode proportional to c_1 leads to a finite action, but unfortunately it also leads to an inverse Gaussian distribution for the final perturbations. Vilenkin and Yamada propose to add a boundary term to the action, which has precisely the effect of cancelling the divergence of the action of the c_2 mode, while rendering that of the c_1 mode infinite. With their new proposal, one would then be led to pick out the stable c_2 mode.

There is a good reason, however, for why this new proposal fails in its current form. The treatment of Vilenkin and Yamada rests on the applicability of a perturbative expansion around the classical background. In the presence of a perturbation mode such as the one above, the equation of motion for the background is corrected at quadratic order by terms involving the linear perturbations, namely (see *e.g.* [102])

$$-2\frac{a_{,\tau\tau}}{a} - \frac{a_{,\tau}^2}{a^2} + \frac{1}{a^2} = \Lambda + \frac{1}{2}\phi_{,\tau}^2 + \frac{(k^2-1)}{6a^2}\phi^2 \tag{6.38}$$

In order for perturbation theory to be valid, the perturbative terms must be small compared to the background terms, in particular they must be small compared to the cosmological constant Λ. But for the c_2 mode the perturbative terms scale as τ^{-2k-4} near $a = 0$. Thus at

the "bottom" of the instanton the solutions become entirely untrustworthy. Put differently, in going beyond the leading term in perturbation theory, one would encounter an infinity of additional terms, which would all blow up at small values of the scale factor, leading to a strong coupling problem where the theory is out of control. We thus conclude that, unfortunately, Vilenkin and Yamada's proposed rescue of the tunneling proposal also fails.

Chapter 7

Spacetime amplitudes: The Schwinger effect revisited

> Already in the beginning I had said that I'll deal with single electrons, and I was going to describe this idea about a positron being an electron going backward in time, and Dirac asked, "Is it unitary?" I said, "Let me try to explain how it works, and you can tell me whether it is unitary or not!" I didn't even know then what "unitary" meant. So I proceeded further a bit, and Dirac repeated his question: "Is it unitary?" So I finally said: "Is what unitary?" Dirac said: "The matrix which carries you from the present to the future position." I said, "I haven't got any matrix which carries me from the present to the future position. I go forwards and backwards in time, so I do not know what the answer to your question is."
>
> Richard P. Feynman, Poconos conference of 1948

7.1 Introduction

Quantum field theory is an extremely successful relativistic generalization of quantum mechanics. It has managed to describe, among many phenomena, ensembles of quantum particles propagating at relativistic speeds, the interactions of particles with photons, and particle-antiparticle pair-creation events. The history of quantum field theory can be traced back to Paul Dirac's early attempt to quantize the electromagnetic field. In 1927, Dirac proposed a model for the emission and absorption of a photon by an electron [107]. In the

subsequent year, Dirac derived a relativistic wave equation – similar to the Klein-Gordon equation for scalar fields proposed by Oskar Klein and Walter Gordon in 1926 [224, 169] – for spin-$\frac{1}{2}$ particles such as the electron [109, 108]. This equation is now known as the Dirac equation. Even though the Dirac equation did successfully predict the spin of the electron, he soon realized that the equation has solutions with both positive and negative kinetic energy. Dirac noticed that these negative energy solutions correspond to a particle with the mass of the electron but with the opposite charge. In 1930, he speculated that the solutions might represent the positively charged proton [110, 111]. However, soon after, Robert Oppenheimer, argued that the proton cannot be the anti-electron as it renders the hydrogen atom unstable [264]. In 1931, Dirac revised his views and postulated the existence of an as-yet-unobserved new particle, the anti-electron, which was later named the positron [112]. Amazingly, the new particle was in 1932 observed by Carl Anderson in a particle accelerator experiment [10]. In 1936 he received the Nobel prize for the detection.

Despite the early success, quantum field theory has from the onset been plagued by several infinities. Basic physical quantities such as the self-energy of the electron, and the corresponding energy shift of the electron states due to the presence of an electromagnetic field lead to divergent contributions. The ultraviolet 'divergence problem' in quantum electrodynamics was solved in the late 1940s by the analysis of renormalization. Quantum field theory was given a mathematical framework distinct from quantum mechanics, with quantum fields and Feynman diagrams used to compute amplitudes in the S-matrix. Quantum electrodynamics became an extremely successful theory of the interaction between electrically charged particles and photons, predicting interactions up to a 10^{-8} accuracy level. The field theory framework has subsequently been successfully extended to quantum chromodynamics and the standard model of particle physics. However, its application to general relativity has so far been unsuccessful. In particular the ultraviolet divergences have not yet been tamed [148].

The relativistic extension of quantum mechanics has led to a distinct framework with its own interpretation. While quantum mechanics describes the evolution of a quantum mechanical system by means of an evolving wave function, whose magnitude squared gives the probability for an event to be observed, quantum field theory describes nature in terms of the excitations of quantum fields. The quantum fields themself are not directly measurable. It should be remarked that the today prevalent tools in quantum field theory originate from an algebraic and a geometric approach to the problem. The canonical quantization, or second quantization, is algebraic in nature and describes ensembles of particles in the Hamiltonian formulation in terms of operators. The path integral formulation, spearheaded by Richard Feynman, is geometric in nature and deals with the Feynman propagator of a single particle in the Lagrangian formalism. Nowadays, however, Feynman

diagrams, which originated from the path integral quantization, are often presented as a computational device to evaluate perturbative calculations of the second quantized theory or the path integral over quantum fields.

Notwithstanding the great success of quantum field theory, the current formulation in terms of the Fock space has its limitations. In S-matrix theory, calculations can be performed by scattering particles described by wavepackets extending to the infinite past and future. These scattering amplitudes can, with the Lehmann-Symanzik-Zimmermann reduction formula [272], be expressed in terms of time-ordered correlation functions of the theory. The algebraic nature of the resulting theory make S-matrix theory an extremely powerfull tool in quantum physics. However, the quantum fields are not directly observable and the S-matrix has difficulty describing the localized evolution of a particle in spacetime such as for example, a cosmic rays entering a cloud chamber or the relativistic particles coming out of particle accelerator experimentst. Every measurement is performed using equipment which is localized in space and time. It is moreover not clear why these momentum states should be preferred when the spacetime is curved and the translation symmetry of the spacetime is broken. In this Chapter we propose a localized formulation of relativistic quantum mechanics which allows us to study the evolution of a single relativistic particle. In this formalism, we adopt a geometric path integral approach to relativistic quantum mechanics, and give a probabilistic interpretation, embodied by Richard Feynman's quote at the start of this chapter, along with a description of the dynamics based solely on the Feynman propagator. The Feynman propagator is normally described as a computational device to evaluate Feynman diagrams. We will here instead consider a more literal interpretation of the Feynman propagator as the amplitude for a particle to move between two spacetime points. The resulting formulation is not in conflict with the conventional formulation of quantum field theory, but rather studies the theory from a broader perspective.

In addition, we extend the framework of weak measurements – first developed by Yakir Aharonov and collaborators [5] – to relativistic theories. These techniques can be used to weakly probe the propagation of a system between an initial and a final state, without affecting the quantum process. We propose the concept of a weak trajectory and evaluate the weak number and charge densities. We argue that the former is a generalization of the (complex) classical trajectories interpolating between two classically connected or disconnected states in the quantum regime. The latter can be used to study the back-reaction of pair-creation events on the pair creating fields. While the weak value analysis of quantum field theory might not lead to revolutionary new insights in particle physics, we anticipate that the framework will be of value in the formulation of quantum geometrodynamics and quantum cosmology.

We illustrate the new formulation of relativistic quantum mechanics and the relativistic weak value analysis, by applying it to scalar quantum electrodynamics in $1+1$-dimensional Minkowski spacetime within a fixed background electromagnetic field. We in particular focus on the Schwinger effect, where an electron turns into an positron by turning around with respect to coordinate time. The results complement the weak value analysis of the Schwinger effect in canonical quantum field theory by Brout et al. [59]. The details of our investigation into the Schwinger effect will be presented in an upcoming publication [135].

7.2 Spacetime amplitudes

There exist several distinct formulations of quantum field theory. We here briefly summarize the canonical quantization and describe the Feynman propagator of a relativistic particle in Minkowski spacetime. We subsequently present the spacetime amplitude formulation of relativistic quantum mechanics. We show that it captures the localized evolution of a relativistic particle by analysing electron-positron pair-creation in the Schwinger effect.

7.2.1 Canonical quantum field theory

Canonical quantum field theory of scalar particles in an electromagnetic field is formulated in terms of a quantum field ϕ satisfying the Klein-Gordon equation

$$\hat{H}\phi = 0\,. \tag{7.1}$$

The Klein-Gordon operator is the second order differential operator

$$\hat{H} = \hbar^2 D_\mu D^\mu + m^2, \tag{7.2}$$

with the covariant derivative

$$D_\mu = \partial_\mu - ieA_\mu\,, \tag{7.3}$$

the mass of the particle m, the electric charge e, and the vector potential A_μ.

Unlike the Schrödinger equation, the Klein-Gordon equation does not have a positive definite norm which is preserved in coordinate time. The current

$$j_\mu = -\frac{ie\hbar}{2m}\phi^* \overleftrightarrow{D_\mu}\, \phi \tag{7.4}$$

is conserved,

$$\partial_\mu j^\mu = -\frac{ie\hbar}{2m}\partial_\mu(\phi^* D^\mu\phi) + \frac{ie\hbar}{2}\partial_\mu(\phi\bar{D}^\mu\phi^*) \tag{7.5}$$

$$= \frac{e\hbar}{m}\,\mathrm{Im}\left[\partial_\mu\phi^* D^\mu\phi + \phi^*\partial_\mu D^\mu\phi\right] \tag{7.6}$$

$$= \frac{e\hbar}{m}\,\mathrm{Im}\left[\bar{D}_\mu\phi^* D^\mu\phi + \phi^* D_\mu D^\mu\phi\right] \tag{7.7}$$

$$= \frac{e\hbar}{m}\,\mathrm{Im}\left[\bar{D}_\mu\phi^* D^\mu\phi - \hbar^{-2}m^2\phi\phi^*\right] \tag{7.8}$$

$$= 0\,, \tag{7.9}$$

with $\bar{D}_\mu$ the complex conjugate of the covariant derivative D_μ. As a consequence, the Klein-Gordon inner product

$$(f,g)_{KG} = -\frac{i}{2}\int_{\Sigma_t} f^*(x^\mu) n^\mu \overleftrightarrow{D_\mu} g(x^\mu)\mathrm{d}\sigma \tag{7.10}$$

integrated over the time slice (3-manifold)

$$\Sigma_{t'} = \{(t,x)|t=t'\}\,, \tag{7.11}$$

with the induced 3-volume element $\mathrm{d}\sigma$, and normal four-vector $n^\mu = (1,0,0,0)$, is preserved in time, *i.e.*,

$$\frac{\partial}{\partial t}\,(\phi,\phi)_{KG} = 0\,, \tag{7.12}$$

where ϕ is a complex solution of equation (7.1), provided we ingnore terms at space-like infinity. However, the integrand is generally not positive definite and can thus not be interpreted as a probability distribution. This fact prevented Schrödinger from building a relativistic quantum theory. Instead, it is interpreted as the charge density

$$\rho = -\frac{ie\hbar}{2m}\phi^*(x^\mu) n^\mu \overleftrightarrow{D_\mu}\phi(x^\mu)\,, \tag{7.13}$$

while equation (7.12) represents the conservation of the total charge [1].

The Klein-Gordon equation is a relativistic wave equation. The free theory of a complex scalar ϕ with $A_\mu = 0$, can be solved in terms of normal modes, *i.e.*,

$$\phi(x^\mu) = \int \frac{\mathrm{d}^3 p}{(2\pi)^3}\frac{1}{\sqrt{2\omega_{\mathbf{p}}}}\left(a_{\mathbf{p}} e^{-i\omega_{\mathbf{p}}t + i\mathbf{p}\cdot\mathbf{x}} + b^*_{\mathbf{p}} e^{i\omega_{\mathbf{p}}t - i\mathbf{p}\cdot\mathbf{x}}\right) \tag{7.17}$$

[1] When the solution of the Klein-Gordon equation is well approximated by the WKB approximation,

with the frequency of the normal mode

$$\omega_{\mathbf{p}} = \sqrt{|\mathbf{p}|^2 + \left(\frac{m}{\hbar}\right)^2}\,. \tag{7.18}$$

Each mode, corresponding to a single momentum $\mathbf{p}$, can be interpreted as a harmonic oscillator with the frequency $\omega_{\mathbf{p}}$. The canonical quantization of the theory is achieved by replacing the modes $a_{\mathbf{p}}, b_{\mathbf{p}}$ and their conjugates $a^*_{\mathbf{p}}, b^*_{\mathbf{p}}$ with creation and annihilation operators $\hat{a}_{\mathbf{p}}, \hat{a}^\dagger_{\mathbf{p}}, \hat{b}_{\mathbf{p}}, \hat{b}^\dagger_{\mathbf{p}}$, *i.e.*,

$$\hat{\phi}(x^\mu) = \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \frac{1}{\sqrt{2\omega_{\mathbf{p}}}} \left(\hat{a}_{\mathbf{p}} e^{-i\omega_{\mathbf{p}} t + i\mathbf{p}\cdot\mathbf{x}} + \hat{b}^\dagger_{\mathbf{p}} e^{i\omega_{\mathbf{p}} t - i\mathbf{p}\cdot\mathbf{x}}\right)\,, \tag{7.19}$$

$$\hat{\phi}^\dagger(x^\mu) = \int \frac{\mathrm{d}^3 p}{(2\pi)^3} \frac{1}{\sqrt{2\omega_{\mathbf{p}}}} \left(\hat{b}_{\mathbf{p}} e^{-i\omega_{\mathbf{p}} t + i\mathbf{p}\cdot\mathbf{x}} + \hat{a}^\dagger_{\mathbf{p}} e^{i\omega_{\mathbf{p}} t - i\mathbf{p}\cdot\mathbf{x}}\right)\,, \tag{7.20}$$

with the commutation relations

$$[\hat{a}_{\mathbf{p}}, \hat{a}^\dagger_{\mathbf{q}}] = (2\pi)^3 \delta(\mathbf{p}-\mathbf{q}), \quad [\hat{b}_{\mathbf{p}}, \hat{b}^\dagger_{\mathbf{q}}] = (2\pi)^3 \delta(\mathbf{p}-\mathbf{q})\,, \tag{7.21}$$

$$[\hat{a}_{\mathbf{p}}, \hat{a}_{\mathbf{q}}] = [\hat{b}_{\mathbf{b}}, \hat{b}_{\mathbf{q}}] = [\hat{a}_{\mathbf{p}}, \hat{b}_{\mathbf{q}}] = [\hat{a}_{\mathbf{p}}, \hat{b}^\dagger_{\mathbf{q}}] = 0\,. \tag{7.22}$$

Particle states enter the theory as excitations of a vacuum state $|0\rangle$, defined as the state which is annihilated by the annihilation operators, *i.e.*,

$$\hat{a}_{\mathbf{p}}|0\rangle = 0\,, \quad \hat{b}_{\mathbf{p}}|0\rangle = 0\,, \tag{7.23}$$

for all momenta $\mathbf{p}$. The on-shell state consisting of a single particle with momentum $\mathbf{p}$ is defined as

$$|\mathbf{p}\rangle = \hat{a}^\dagger_{\mathbf{p}}|0\rangle\,. \tag{7.24}$$

---

we can write the quantum field as

$$\phi(x^\mu) = A(x^\mu) e^{iS(x^\mu)}\,, \tag{7.14}$$

where $S(x^\mu)$ is assumed to be a real function varying quickly with respect to the magnitude $A(x^\mu)$. When the WKB approximation applies the current takes the form

$$j_\mu = \frac{e\hbar}{m}|A|^2\left(\partial_\mu S - eA_\mu\right) \tag{7.15}$$

$$= \frac{e\hbar}{m}|A|^2 P_\mu\,, \tag{7.16}$$

with the kinetic momentum P_μ, for which the charge density $\rho = ej_0$ is either positive or negative when the WKB approximation applies, $P_0 \neq 0$, for a suitable gauge choice (*e.g.* $A_0 = 0$).

The on-shell antiparticle state with momentum $\mathbf{p}$ is defined as

$$|\bar{\mathbf{p}}\rangle = \hat{b}^\dagger_{\mathbf{p}}|0\rangle\,. \tag{7.25}$$

Iterative application of creation operators generates the Fock space of the theory, consisting of the many-particle antiparticle states. In quantum electrodynamics the vector potential A_μ is introduced perturbatively. Properties of canonical quantum field theory are evaluated by keeping track of the algebra of the relevant creation and annihilation operators. This formulation of relativistic quantum mechanics focusses on multi-particle states and is manifestly written in terms of a fixed foliation of spacetime.

These on-shell momentum states $|\mathbf{p}\rangle$ have been successfully used in the calculation of scattering amplitudes in Minkowski spacetime, *i.e.*, to build an S-matrix. However, it should be noted that these states are, by complementarity, completely de-localized in spacetime. They give a sharp description of the four-momentum p^μ of the particle and, correspondingly, no knowledge at all of its position in spacetime. It is for this reason useful to consider different solutions to the Klein-Gordon equation, consisting of wave packets extending from the infinite past to the infinite future. Using the Lehmann-Symanzik-Zimmermann reduction formula, the scattering amplitudes in terms of these wavepackets can be expressed in terms of the time-ordered correlation functions of these de-localized states. In the next section we will go beyond these wave packets. We believe that this extension will in particular be useful in the study the localized evolution of relativistic quantum particles on curved spacetime, where the geometry of spacetime can evolve and generally breaks spacetime translation symmetry. Moreover, we think that this extension is useful in the development of Wheeler and DeWitt's quantum geometrodynamics.

7.2.2 The Feynman propagator

Richard Feynman's path integral formulation of relativistic quantum mechanics focusses on the propagator (see Chapters 2 and 8), *i.e.*,

$$G[x^\mu_1; x^\mu_0] = \int_{0^+}^{\infty} \mathrm{d}s\, \langle x^\mu_1 | e^{-is\hat{H}/\hbar} | x^\mu_0 \rangle \tag{7.26}$$

$$= \int_{0^+}^{\infty} \mathrm{d}s\, G[x^\mu_1; x^\mu_0; s]\,, \tag{7.27}$$

describing the propagation of a particle from spacetime event x^μ_0 to event x^μ_1, with the Schwinger time $s > 0$ measuring the proper internal time of the particle. The amplitude

to propagate between spacetime events x_0^μ and x_1^μ in Schwinger time s is given by the Lorentzian functional integral

$$G[x_1^\mu; x_0^\mu; s] = \int_{x^\mu(0)=x_0^\mu}^{x^\mu(s)=x_1^\mu} \mathcal{D}x^\mu \mathcal{D}p_\mu \; e^{iS[x^\mu;p_\mu;s]/\hbar} \,, \tag{7.28}$$

with the action

$$S[x^\mu, p_\mu; s] = \int_0^1 \left[p_\mu \dot{x}^\mu - s\tilde{H}[x^\mu, p_\mu] \right] \mathrm{d}t \,, \tag{7.29}$$

where $\tilde{H}$ is the Hamiltonian. In a background electromagnetic field, with vector potential A_μ, we have the Hamiltonian

$$\tilde{H}[x^\mu, p_\mu] = \frac{1}{2m} \left[(p_\mu - eA_\mu)\eta^{\mu\nu}(p_\nu - eA_\nu) + m^2 \right] \,, \tag{7.30}$$

where e and m are the charge and the mass of the particle. The variable

$$P_\mu = p_\mu - eA_\mu \tag{7.31}$$

is the kinetic momentum of the particle, which is related to the four-velocity by the relation $\dot{x}^\mu = P^\mu/m$. We will in this Chapter include the factor $\frac{1}{2m}$ in the Hamiltonian, as it makes the connection to non-relativistic quantum mechanics more straightforward. This factor can however be removed by redefining the Schwinger time. The functional integral ranges over all possible trajectories interpolating between the spacetime events x_0^μ and x_1^μ making the path integral quantization geometrical in nature. The Feynman propagator $G[x_1^\mu; x_0^\mu]$ is a superposition of amplitudes for different paths and all allowed Schwinger times, s: since the boundary conditions for the propagator $G[x_1^\mu; x_0^\mu]$ do not specify the proper time, we integrate over it in accordance with the rules of quantum amplitudes (see [150]). The integral over s is taken over the half line $(0^+, \infty)$ since we wish to include only paths for which the parametrization is monotonic, *i.e.*, one-to-one, and for which the metric on the paths corresponding to the line element, $s^2\mathrm{d}t^2$, is invertible, hence $s > 0$. This corresponds to the causation between the two states, *i.e.*, we propagate from the initial state $|x_0\rangle$ to the final state $|x_1\rangle$ along all possible properly parametrized worldlines of the particle. Mathematically this formula follows directly from the quantization of constrained systems, as discussed in detail in Chapter 2.

Note that, by construction, the amplitude $G[x_1^\mu; x_0^\mu; s]$ is a Green's function of the Schrödinger equation

$$\left[i\hbar \frac{\partial}{\partial s} - \hat{H} \right] G[x_1^\mu; x_0^\mu; s] = i\hbar\delta(x_0^\mu - x_1^\mu)\delta(s) \,, \tag{7.32}$$

with the coincidence limit

$$\lim_{s\to 0} G[x_1^\mu; x_0^\mu; s] = \delta(x_0^\mu - x_1^\mu)\,. \tag{7.33}$$

By analogy with the propagator of non-relativistic quantum mechanics, it follows that the propagator $G[x_1^\mu; x_0^\mu; s]$ satisfies a composition rule with respect to the Schrödinger inner product, *i.e.*,

$$G[x_2^\mu; x_0^\mu; s_1 + s_2] = \int \mathrm{d}x_1^\mu G[x_2^\mu; x_1^\mu; s_2] G[x_1^\mu; x_0^\mu; s_1]\,. \tag{7.34}$$

The Feynman propagator $G[x_1^\mu; x_0^\mu]$ is a Green's function of the Klein-Gordon equation

$$\hat{H} G[x_1^\mu; x_0^\mu] = -i\hbar\delta(x_0^\mu - x_1^\mu)\,. \tag{7.35}$$

Note that in a spacetime with a time-like Killing vector field, it satisfies a composition rule with respect to the Klein-Gordon inner product

$$G[x_2^\mu, x_0^\mu] = -\int_\Sigma G[x_2^\mu, x_1^\mu]\, n^\mu \overleftrightarrow{D}_\mu G[x_1^\mu, x_0^\mu]\, \mathrm{d}\sigma\,, \tag{7.36}$$

with Σ a spacelike slice, n^μ the normal four-vector and $\mathrm{d}\sigma$ the induced measure on Σ [183].

7.2.3 Seeds, trees and relative probabilities

Instead of describing the relativistic particles algebraically, in terms of creation and annihilation operators acting on the vacuum state, we here use the Feynman propagator to construct a geometric formulation of the theory. We interpret the Feynman propagator $G[x_1^\mu; x_0^\mu]$ as the amplitude to propagate between the spacetime points x_0^μ and x_1^μ. When the particle propagates from a more general state $|\psi_0\rangle$ to the position state $|x_1^\mu\rangle$, we can write the spacetime amplitude as a convolution of the initial state with the Feynman propagator

$$\varphi(x_1^\mu) = \int_{0^+}^\infty \mathrm{d}s\ \langle x_1^\mu | e^{-is\hat{H}/\hbar} | \psi_0 \rangle \tag{7.37}$$

$$= \int \mathrm{d}x_0^\mu\, G[x_1^\mu; x_0^\mu] \psi_0(x_0^\mu)\,, \tag{7.38}$$

where in the position representation

$$\psi_0(x_0^\mu) = \langle x_0^\mu | \psi_0 \rangle\,. \tag{7.39}$$

In our analysis we call the initial state, $|\psi_0\rangle$, the *seed*, and the spacetime amplitude, φ, the *tree*. In coordinate independent notation, we describe the tree as the state $|\varphi\rangle$.

Alternatively we can write the tree $\varphi(x_1^\mu)$ as the s-integral of the *wavefunction* $\psi(x_1^\mu; s)$, *i.e.*,

$$\varphi(x_1^\mu) = \int_{0^+}^{\infty} \mathrm{d}s\, \psi(x_1^\mu; s)\,, \tag{7.40}$$

with

$$\psi(x_0^\mu; s) = \int \mathrm{d}x_0^\mu\, G[x_1^\mu; x_0^\mu; s]\psi_0(x_0^\mu)\,. \tag{7.41}$$

The wavefunction $\psi_0(x_1^\mu; s)$ satisfies the Schrödinger-like equation

$$i\hbar\frac{\partial\psi(x^\mu; s)}{\partial s} = \hat{H}\psi(x^\mu; s)\,, \tag{7.42}$$

for positive s, and the boundary condition

$$\psi(x^\mu; 0) = \psi_0(x^\mu)\,. \tag{7.43}$$

In coordinate independent notation we write this wavefunction as $|\psi_s\rangle$, *i.e.*,

$$\psi(x^\mu; s) = \langle x^\mu|\psi_s\rangle\,. \tag{7.44}$$

The seed $|\psi_0\rangle$, the tree $|\varphi\rangle$, and the wavefunction $|\psi_s\rangle$ have a number of algebraic properties which are significant in the development of their physical interpretation:

- In our formulation, the seed $|\psi_0\rangle$ is an arbitrary superposition of the position eigenstates $|x^\mu\rangle$. It describes the initial state of the particle and does generally not solve the Klein-Gordon equation. Remark that the state $|x^\mu\rangle$ is, though mathematically suitable, not a physically motivated state, as it does not overlap with the Hamiltonian constraint and does not specify whether the particle is initially moving forward or backward in time. Nor is it a solution of the constraint equation, *i.e.*, $\hat{H}|x^\mu\rangle \neq 0$. Note that when we require the seed to be a solution of the Klein-Gordon equation,

$$\hat{H}|\psi_0\rangle = 0\,, \tag{7.45}$$

the tree will diverge

$$\varphi(x_1^\mu) = \int_{0^+}^{\infty} \mathrm{d}s \ \langle x_1^\mu | e^{-is\hat{H}/\hbar} | \psi_0 \rangle \tag{7.46}$$

$$= \psi_0(x_1^\mu) \int_{0^+}^{\infty} \mathrm{d}s \tag{7.47}$$

and can only be made sense of if it is carefully regularized. Note that solutions of the constraint equation do not propagate and are generally not normalizable.

- For a spacetime with a time-like Killing vector field, the composition rule for the Feynman propagator (7.36) implies an interesting gluing operation for the tree. Given the tree, φ, and its derivatives, $n^\mu \partial_\mu \varphi$, restricted to a space-like surface Σ, the tree can be recovered using the Klein-Gordon inner product, *i.e.*,

$$-\int_\Sigma G[x_1^\mu; x_0^\mu] n^\mu \overleftrightarrow{D_\mu} \varphi|_\Sigma(x_0^\mu) \mathrm{d}\sigma = -\int \mathrm{d}x^\mu \left[\int_\Sigma G[x_1^\mu; x_0^\mu] n^\mu \overleftrightarrow{D_\mu} G[x_0^\mu; x^\mu] \mathrm{d}\sigma \right] \psi_0(x^\mu) \tag{7.48}$$

$$= \int \mathrm{d}x^\mu \, G[x_1^\mu; x^\mu] \psi_0(x^\mu) \tag{7.49}$$

$$= \varphi(x_1^\mu) \, . \tag{7.50}$$

That is to say, the tree can be sourced by either the initial seed $|\psi_0\rangle$ and the tree and its first order derivative restricted to an arbitrary space-like surface $\varphi|_\Sigma$ by convolving them with the Feynman propagator with respect to either the Schrödinger

$$(f, g)_S = \int f^*(x^\mu) g(x^\mu) \mathrm{d}x^\mu \tag{7.51}$$

or Klein-Gordon inner product

$$(f, g)_{KG} = -\frac{i}{2} \int_{\Sigma_t} f^*(x^\mu) n^\mu \overleftrightarrow{D_\mu} g(x^\mu) \mathrm{d}\sigma \, . \tag{7.52}$$

- Since the tree, φ, is a linear combination of the Feynman propagator, it does *not* satisfy the Klein-Gordon equation on the support of the seed,

$$\hat{H}\varphi(x_1^\mu) = \int \mathrm{d}x_0^\mu \hat{H} G[x^\mu; x_0^\mu] \psi_0(x_0) \tag{7.53}$$

$$= -i\hbar \int \mathrm{d}x_0^\mu \delta(x_0^\mu - x^\mu) \psi_0(x_0) \tag{7.54}$$

$$= -i\hbar \psi_0(x^\mu) \, . \tag{7.55}$$

Below, we discuss how the tree is related to solutions of the Klein-Gordon equation and canonical quantum field theory. It will at that point become clear how this violation of the Klein-Gordon equation should be interpreted.

- The wavefunction $\psi(x^\mu; s)$ does not satisfy the Schrödinger equation, since

$$\left[i\hbar\frac{\partial}{\partial s} - \hat{H}\right]\psi(x^\mu; s) = \int \mathrm{d}x_0^\mu \left[i\hbar\frac{\partial}{\partial s} - \hat{H}\right] G[x_1^\mu; x_0^\mu; s]\psi_0(x_0^\mu) \tag{7.56}$$

$$= i\hbar \int \mathrm{d}x_0^\mu \delta(x_0^\mu - x_1^\mu)\delta(s)\psi_0(x_0^\mu) \tag{7.57}$$

$$= i\hbar\psi_0(x^\mu)\delta(s)\,. \tag{7.58}$$

Note that the wavefunction can be extended to negative Schwinger time, s, if one wishes to study the behaviour of the wavefunction for negative Schwinger time.

- For a solution to the Klein-Gordon equation ϕ, the electric current

$$j^\mu = -\frac{ie\hbar}{2m}\phi^*(x^\mu)\overleftrightarrow{D^\mu}\phi(x^\mu) \tag{7.59}$$

is locally conserved as $\partial_\mu j^\mu = 0$. The analogous current of the tree

$$j^\mu = -\frac{ie\hbar}{2m}\varphi^*(x^\mu)\overleftrightarrow{D^\mu}\varphi(x^\mu) \tag{7.60}$$

is conserved away from the seed $|\psi_0\rangle$ since

$$\partial_\mu j^\mu = -\frac{ie\hbar}{2m}\partial_\mu(\varphi^* D^\mu\varphi) + \frac{ie\hbar}{2m}\partial_\mu(\varphi\bar{D}^\mu\varphi^*) \tag{7.61}$$

$$= \frac{e\hbar}{m}\mathrm{Im}\left[D_\mu\varphi\bar{D}^\mu\varphi^* + \varphi^* D_\mu D^\mu\varphi\right] \tag{7.62}$$

$$= \frac{e\hbar}{m}\mathrm{Im}\left[D_\mu\varphi\bar{D}^\mu\varphi^* - \hbar^{-2}m^2\varphi\varphi^* - i\hbar^{-2}\psi_0\varphi^*\right] \tag{7.63}$$

$$= -\frac{e}{m\hbar}\mathrm{Re}[\psi_0\varphi^*]\,. \tag{7.64}$$

Within the support of the seed, electric charge is not conserved in order to accommodate the introduction of the charged particle. Away from the support of the seed, the charge current is locally conserved.

The spacetime amplitudes of a particle in an electric field

We develop the interpretation of the seed $|\psi_0\rangle$, the wavefunction $|\psi_s\rangle$, and the tree $|\varphi\rangle$, by evaluating them for a relativistic particle in an electric field. We consider the wavefunction and the tree originating from the Gaussian seed

$$\psi_0(x^\mu) = \frac{1}{\sqrt{2\pi\sigma_t\sigma_x}} e^{-\frac{(t-\mu_t)^2}{4\sigma_t^2} - \frac{(x-\mu_x)^2}{4\sigma_x^2} - i(p_t - eA_t)(t-\mu_t) + i(p_x - eA_x)(x-\mu_x)}\,, \tag{7.65}$$

specified by the mean initial position (μ_t, μ_x) with the initial spread (σ_t, σ_x) and the mean momentum (p_t, p_x). In this section we will only consider initial states with a significant overlap with the Hamiltonian constraint, for which

$$\tilde{H}((\mu_t, \mu_x), (p_t, p_x)) = 0\,. \tag{7.66}$$

We will assume the initial spread to be symmetric in space and time, *i.e.*, $\sigma_t = \sigma_x$.

These initial states might seem counter intuitive, as they do not solve the Klein-Gordon equation. However note that these states not only implement the well-known Heisenberg uncertainty relation for space and momentum

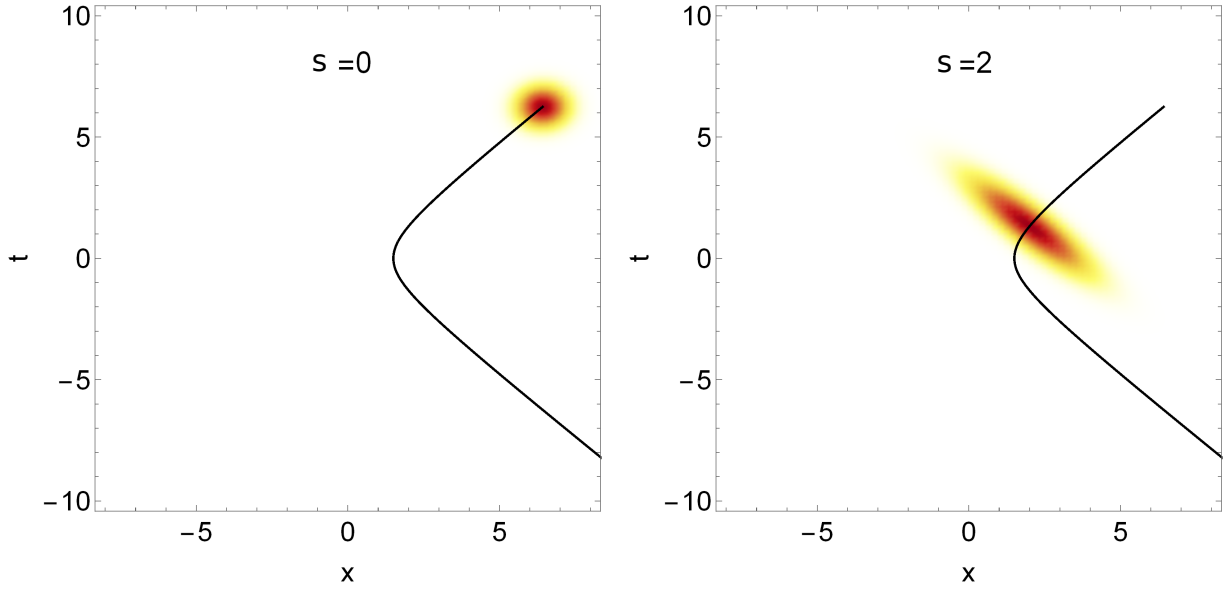
$$\Delta x \Delta p \geq \frac{\hbar}{2}\,, \tag{7.67}$$

but also exhibit the well-known time energy uncertainty relation

$$\Delta t \Delta E \geq \frac{\hbar}{2}\,, \tag{7.68}$$

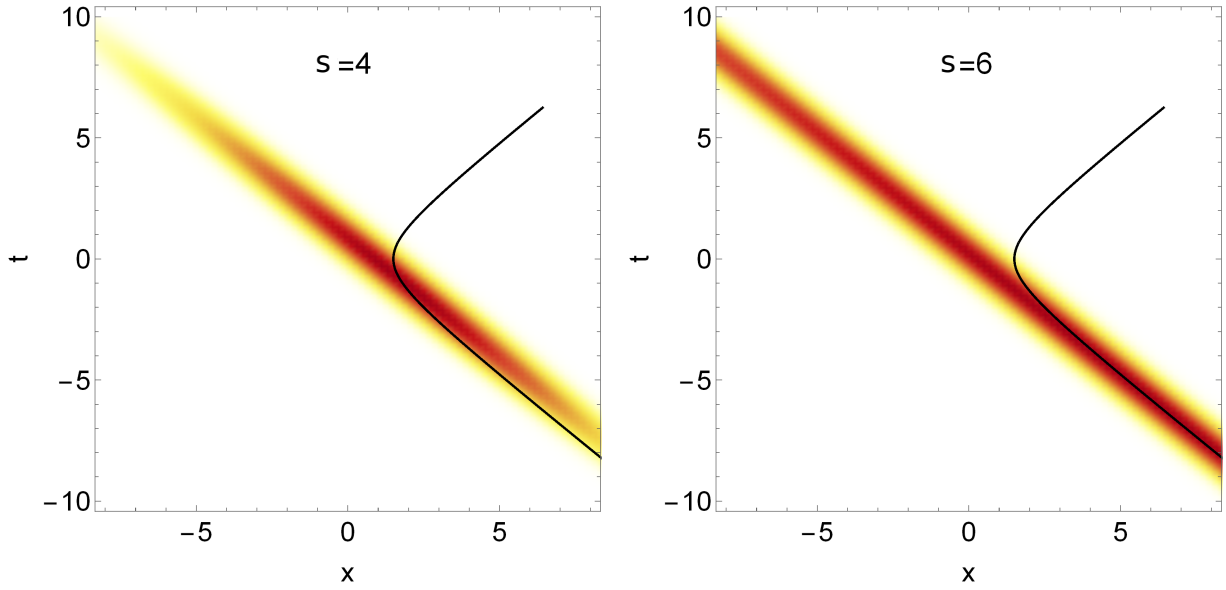
in a natural manner. Note that derivations of (7.68) coming from the non-relativistic Schrödinger equation are generally somewhat subtle [5].

First consider the absolute square of the wavefunction $|\psi(x^\mu; s)|^2$ of a particle in a mild constant electric field (see Figure 7.1). See appendix 7.A for a derivation of the Feynman propagator, the wavefunction and the tree of a relativistic particle in a constant electric field. In the lab frame, the seed represents a right moving anti-particle. We here follow Feynman's interpretation and consider the particle to be moving initially backwards in time and to the left in space. As the particle propagates forwards in Schwinger time s, it decelerates and turns around. During this initial part of its evolution, the wavefunction follows the classical trajectory starting from (μ_t, μ_x) with the kinetic momentum (P_t, P_x) (the black curve). While it starts to move to the right, the wavefunction stretches to the upper left and lower right corner of the plot. This is an indication that there is pair creation



(a) The wavefunction $\psi(x^\mu, s = 0)$

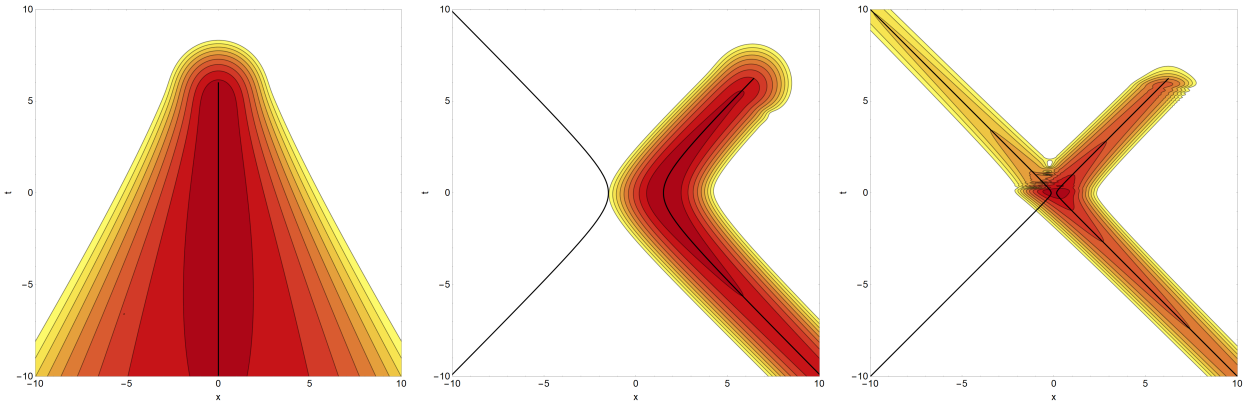
(b) The wavefunction $\psi(x^\mu, s = 2)$



(c) The wavefunction $\psi(x^\mu, s = 4)$

(d) The wavefunction $\psi(x^\mu, s = 6)$

Figure 7.1: The evolution of the absolute square of the wavefunction, $|\psi(x^\mu; s)|^2$, of the relativistic particle in a constant electric field. The black curve is the corresponding classical trajectory.



(a) $eE = 0$ (b) $eE = 4$ (c) $eE = 40$

Figure 7.2: The tree, $|\varphi(x^\mu)|^2$, of the relativistic particle in an electric field for various field strengths $eE = 0, 4, 40$. The black curve is the classical worldline corresponding to the initial conditions of the seed $|\psi_0\rangle$. We have chosen the distance between the two classical trajectories at $t = 0$ to be the tunnelling length of the Schwinger effect $L = \frac{2m}{|eE|}$. This length follows from the analysis in appendix 7.A.

in an electric field: one finds that when the particle turns round in the spatial direction, $(P_t, P_x) = (-1, 0)$, and the WKB approximation for the wavefunction $\psi(x; s)$ fails, and an amplitude is generated for turning around in time too.

In Figure 7.2, we plot the absolute value squared of the tree, $|\varphi(x_1^\mu)|^2$, for the relativistic particle at various electric field strengths Ee. The contours are plotted on a logarithmic scale. In the left Figure we see that the tree of the free antiparticle with the initial kinetic momentum $(P_t, P_x) = (-m, 0)$ has support around the classical world line displayed by the black curve. The tree can be interpreted as a fuzzy worldline. The central Figure corresponds to a mild electric field. This is the same configuration as the one discussed in Figure 7.1. The particle initially moves in the lower left direction, decelerates, turns around and is subsequently accelerated by the electric field. Note that even though the wavefunction for this configuration has support in the upper left corner of the diagram, it is exponentially suppressed and hence particle production is insignificant. Away from the classical world line, destructive interference suppresses the amplitude, φ (equation (7.2)). Finally, in the right Figure we repeat the calculation for an electric field eE exceeding the threshold of the Schwinger effect, *i.e.*, the electric field is strong enough to create abundant particle-antiparticle pairs. The main part of the tree, $|\varphi|^2$, still follows the classical trajectory on the right determined by the configuration of the seed. The classical

trajectory on the left is a mirror image displaced by the Schwinger length

$$L_c = \frac{2m}{|eE|}\,. \tag{7.69}$$

However, the tree now has significant support in the upper left corner along the left classical trajectory. The larger the electric field, the more likely it is for the particle to turn around in coordinate time, hence pair creation becomes more probable.

We interpret the absolute value squared of the tree $|\varphi(x_1^\mu)|^2$ as the relative probability to find the particle in spacetime location x_1^μ given that it started in an initial state described by the seed $\psi_0(x_0)$. In other words, if the seed $|\psi_0\rangle$ contains an on-shell component, the tree, φ, is the quantum representation of the worldline of the particle. Note that the tree is generally not normalizable over spacetime as classical worldlines on Minkowski spacetime normally extend to the infinite future or the infinite past.

The case of a constant electric field is easiest to analyse, as the relativistic path integral is Gaussian (see appendix 7.A). However, this is physically not the best motivated configuration, as an electric field generally only persists for a finite amount of time. In Figure 7.3, we plot the numerically evaluated tree in the case where the electric field is only turned on for the interval of time between the dashed lines, *i.e.*,

$$A^0 = -Exf(t)\,, \quad A^1 = 0\,, \tag{7.70}$$

with

$$f(t) = Si(A(t_a - t)) - Si(A(t_b - t))\,, \tag{7.71}$$

where the sigmoid function Si is defined as

$$Si(x) = \frac{1}{1 + e^{-x}}\,. \tag{7.72}$$

For a description of the Suzuki-Trotter method used to create this Figure see appendix 7.B. The particle travels as a free particle in the absence of the electric field. When the electric field is turned on, the particle decelerates and turns around in both the space and time direction.

Virtual particles

It is instructive to study the behaviour of the tree, φ, for a seed, ψ_0, which does not have a significant overlap with the relativistic constraint $\tilde{H} = 0$ of the system. Consider the tree,

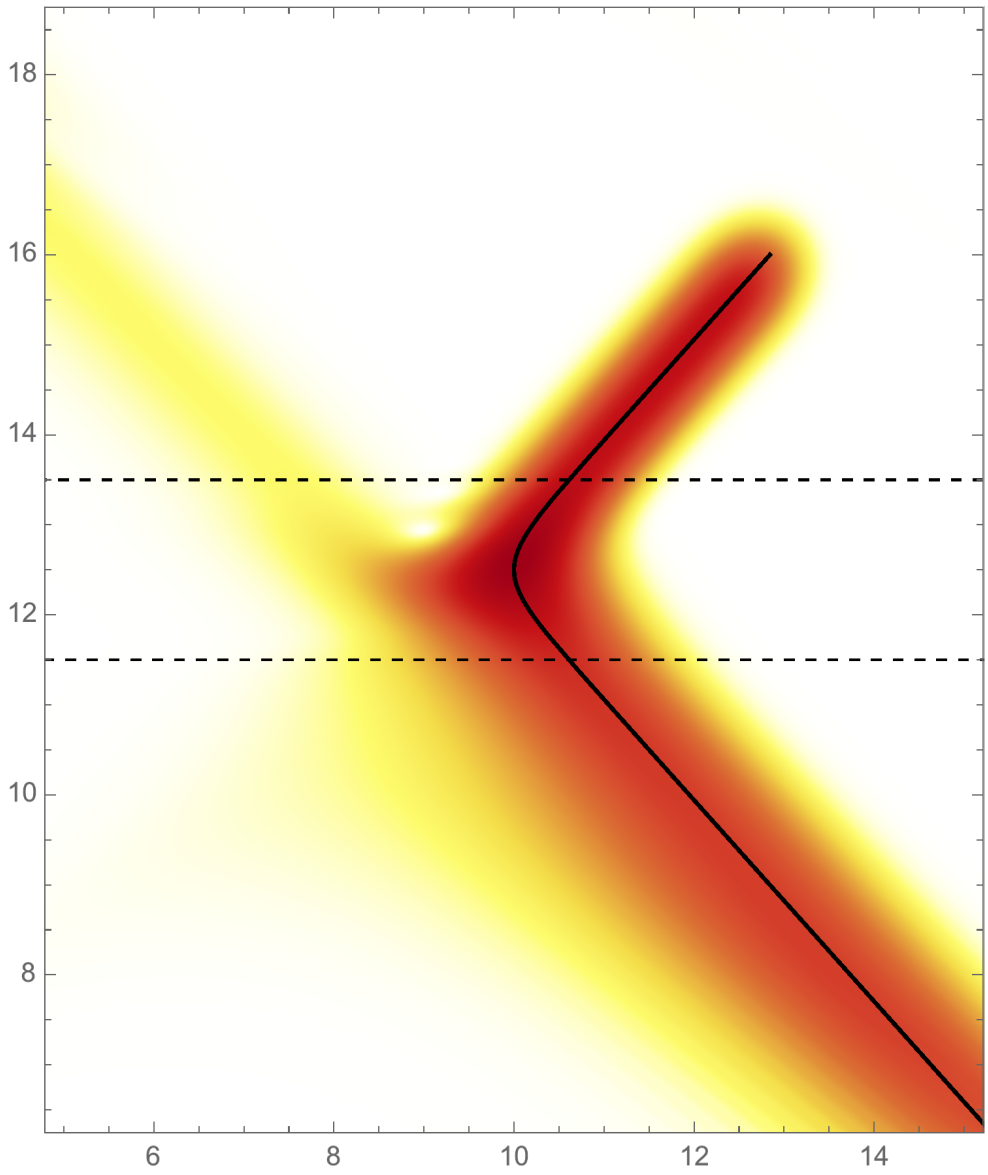
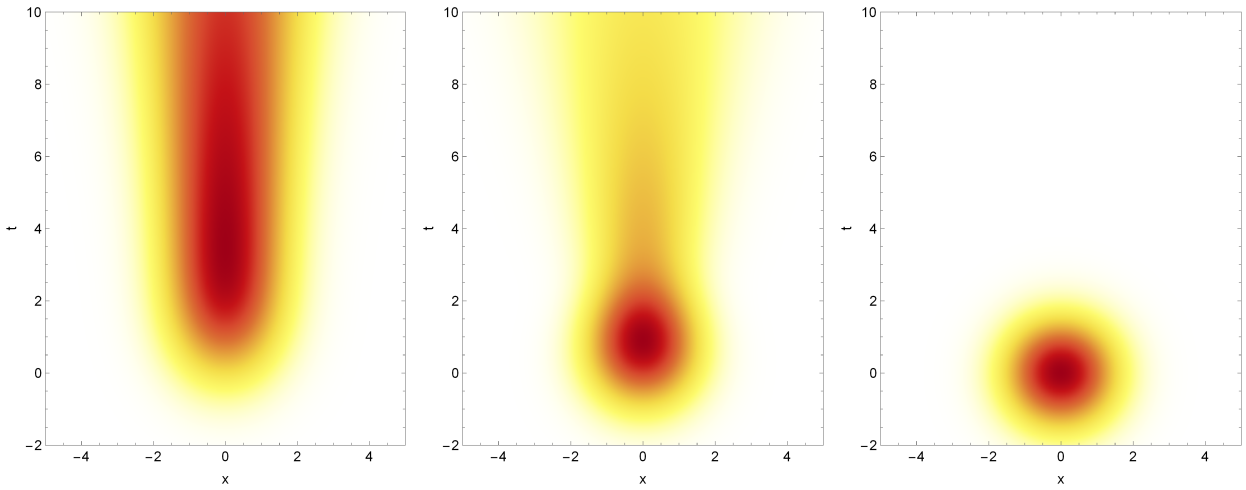


Figure 7.3: Numerical evaluation (horizontal space and vertical time) of the tree in an electric field which is switched on and off according to equation (7.70), with $A = 6$, $t_a = 11.5$, and $t_b = 13.5$. The color scale represents $\log |\varphi|^2$. The black curve is the classical trajectory of the particle. The dashed lines indicate the boundaries of the spacetime region for which the electric field is on.



(a) On-shell Gaussian localized state, $p^2 = -m^2$.

(b) Off-shell Gaussian localized state, $p^2 < -m^2$.

(c) An off-shell Gaussian localized state, $p^2 \ll -m^2$.

Figure 7.4: The absolute square of the spacetime amplitude $|\varphi|^2$ of a localized free particle initial state evolved with the Feynman propagator. The spacetime amplitude is progressively reduced as the initial state is taken more and more off-shell. In Figure a, the seed is on-shell. The tree propagates to infinity. In Figure b, the seed is slightly off-shell and only propagates for a short distance. In Figure c, the seed is completely off-shell and does not propagate.

$|\varphi(x^\mu)|^2$, corresponding to a free particle with vanishing initial spatial momentum $p_x = 0$ for various values of the temporal momentum p_t (see Figure 7.4). When the initial state is on-shell, *i.e.*, $p_t^2 = m^2$, the tree follows the classical worldline. The oscillatory integral over the wavefunction (7.40) exhibits constructive interference extending to future infinity. When the temporal momentum, p_t, is either decreased or increased we observe that the interference becomes destructive. The integral over the Schwinger time projects the tree onto the on-shell states. That is to say, a virtual off-shell particle does not propagate very far. For an analytic derivation of this phenomenon we refer to the upcoming publication [135].

The relation to quantum field theory

As we observed above, the tree, φ, is not a solution of the Klein-Gordon equation, *i.e.*,

$$\hat{H}\varphi(x_1^\mu) = -i\hbar\psi(x_1^\mu)\,. \tag{7.73}$$

The tree deviates from a solution on the support of the seed. This is not surprising as the tree is constructed from the Feynman propagator which also does not satisfy the Hamiltonian constraint, even in non-relativistic quantum mechanics: see Chapter 2 of this thesis and Feynman and Hibbs [150] (equation 4.29).

These spacetime amplitudes and localized states might at first glance seem like a big departure from conventional quantum field theory. However, it is possible to relate the tree to a solution of the Klein-Gordon equation and extend the interpretation to quantum fields. First consider a tree, $|\varphi\rangle$, corresponding to a seed, $|\psi_0\rangle$, with compact support consisting of only one connected component. Consider a time slice Σ chosen away from the support of the seed, which intersects the support of the tree. The time slice, Σ, will generally partition the spacetime into a region for which the seed has support (region I) to the future or past of Σ, and a region for which does not have support (region II), the complement of region I. The tree and its derivatives restricted to the time slice, $\varphi|_\Sigma$, $\partial_\mu\varphi|_\Sigma$, satisfy the Klein-Gordon equation. If we now solve the Klein-Gordon equation with this initial data, we obtain an on-shell amplitude ϕ, which coincides with the tree, φ, in region II. In region I, the on-shell amplitude, ϕ, agrees with the tree, φ, till it hits the seed beyond which ϕ extends the amplitude. This extension represents the evolution of the particle beyond the initial state $s = 0$ for negative internal Schwinger time, *i.e.*, $s < 0$. Using this observation we can, besides constructing the amplitude $|\phi\rangle$ from the tree restricted to the time slice $\varphi|_\Sigma$, build the on-shell spacetime amplitude $|\phi\rangle$ from the initial seed $|\psi_0\rangle$ by extending the wavefunction $|\psi_s\rangle$ to negative Schwinger time and integrating over the real line,

$$\phi(x_1^\mu) = \int_{-\infty}^{\infty} \mathrm{d}s\, \langle x_1^\mu | e^{-is\hat{H}/\hbar} | \psi_0 \rangle \tag{7.74}$$

$$= \int \mathrm{d}x_0^\mu\, G_H[x_1^\mu; x_0^\mu]\psi_0(x_0^\mu)\,, \tag{7.75}$$

where the function $G_H[x_1^\mu; x_0^\mu]$ is defined as

$$G_H[x_1^\mu; x_0^\mu] = \int_{-\infty}^{\infty} \mathrm{d}s\, \langle x_1^\mu | e^{-i\hat{H}/\hbar} | x_0^\mu \rangle\,. \tag{7.76}$$

In Minkowski spacetime, this function is known as the Hadamard function, which can be expressed in terms of the real part of the Feynman propagator, *i.e.*,

$$G_H[x_1^\mu; x_0^\mu] = 2\,\mathrm{Re}\left[G[x_1^\mu; x_0^\mu]\right]\,. \tag{7.77}$$

Note that the operator projects the state onto an on-shell spacetime amplitude, since formally

$$\int_{-\infty}^{\infty} \mathrm{d}s\, e^{-s\hat{H}/\hbar} = 2\pi\delta(\hat{H}) \tag{7.78}$$

or in more explicit terms in the setting of Minkowski spacetime

$$\hat{H}G_H[x_1^\mu; x_0^\mu] = 2\,\mathrm{Re}\left[\hat{H}G[x_1^\mu; x_0^\mu]\right] \tag{7.79}$$

$$= 2\,\mathrm{Re}\left[-i\hbar\delta(x_0^\mu - x_1^\mu)\right] \tag{7.80}$$

$$= 0\,. \tag{7.81}$$

See Figure 7.5 for a comparison between the tree and the on-shell state constructed in this way, for a massive particle in a mild electric field (for the derivation see appendix 7.A). This extension is similar to the extension of the wavefunction prior to the initial conditions in quantum mechanics. Since the on-shell amplitude ϕ coincides with the tree in region I, it is natural to extend the probabilistic interpretation of the tree to the quantum field. This link between the tree φ and the spacetime amplitude ϕ extends to more complicated seeds with more complicated supports. As far as the quantum worldline interpretation is concerned, we can interpret the tree as a computational device to construct a solution of the Klein-Gordon equation which is localised around a particular world line of interest.

Given the Feynman propagator for a single particle, we remark that it is possible to describe the multi-particle states in the Fock space of quantum field theory, by imposing unitarity with respect to the embedding time of Minkowski space. This was first worked out by Richard Feynman in quantum electrodynamics [149]. In an upcoming paper we will explicitly illustrate how the rate of pair creation is derived from the single particle propagator [135].

Finally, note that the electric current $j_\mu = -\frac{ie\hbar}{2m}\varphi^*(x^\mu)\overleftrightarrow{D_\mu}\varphi(x^\mu)$ is conserved away from the support of the seed $|\psi_0\rangle$ *i.e.*,

$$\partial_\mu j^\mu = -\hbar^{-1}\,\mathrm{Re}[\psi_0\varphi^*]\,. \tag{7.82}$$

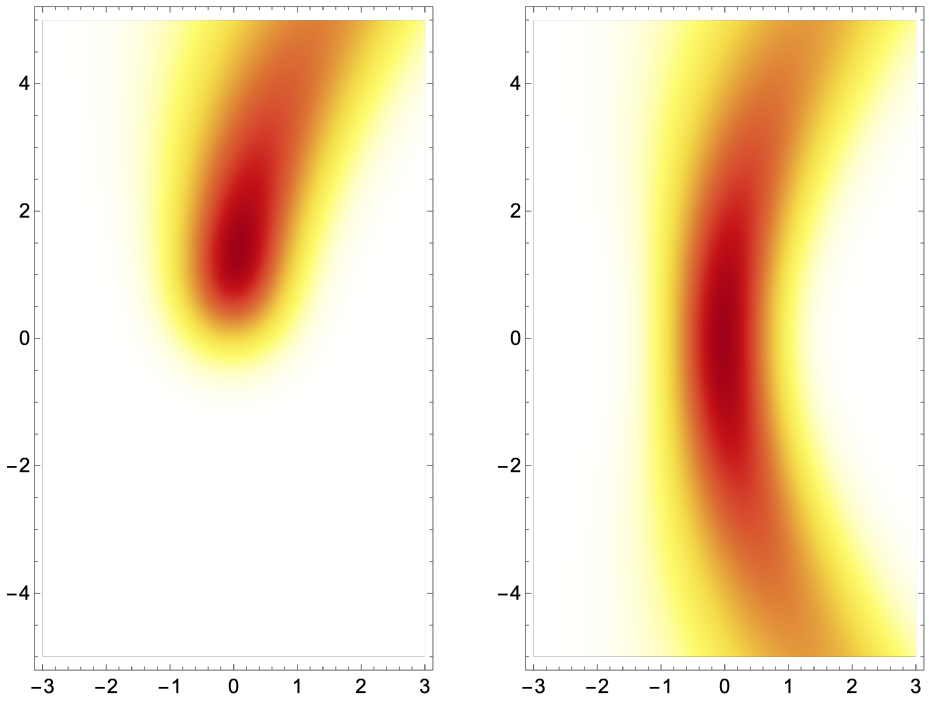
The corresponding charge density

$$\rho(t,x) = -\frac{ie\hbar}{2m}\varphi^*(t,x)\overleftrightarrow{D_0}\varphi(t,x) \tag{7.83}$$

is conserved along the time slice $\Sigma_{t'} = \{(t,x)|t = t'\}$, *i.e.*,

$$\frac{\partial}{\partial t}\left[\int_{\Sigma_t}\rho(t,x)\mathrm{d}\sigma\right] = 0\,, \tag{7.84}$$

with the induced volume element $\mathrm{d}\sigma$. In the example of Figure 7.3 the charge density in the right half of the diagram is negative since it represents an antiparticle. The charge density



(a) $|\varphi(x^\mu)|^2$ (b) $|\phi(x^\mu)|^2$

Figure 7.5: Left: the absolute square of the tree $\varphi(x_1^\mu) = \int G[x_1^\mu; x_0^\mu]\psi_0(x_0^\mu)\mathrm{d}x_0^\mu$, right: the absolute value squared of the solution of the Hamiltonian constraint $\phi(x_1^\mu) = \int G_H[x_1^\mu; x_0^\mu]\psi_0(x_0^\mu)\mathrm{d}x_0^\mu$ of a massive scalar particle, $m = 4.5$ in a mild electric field, $eE = 0.5$, generated using the Feynman propagator and the Hadamard function and at Gaussian initial state centred at $(\mu_t, \mu_x) = (0, 0)$ with initial momentum $(p_t, p_x) = (m, 0)$ and initial spread $\sigma_t = \sigma_x = 0.5$.

is positive in the left half where the antiparticle has turned into a particle. Note that the sign of the charge density is locally preserved when the particle does not turn around in the time direction, while the WKB approximation is valid. Under these conditions, we can take the part of the density which is either positive or negative and interpret this density as a relative probability. More concretely, we can construct the relative probability

$$p_t(x) = |\Pi_\pm \rho(t, x)| \tag{7.85}$$

where the projection operator $\Pi_\pm$ is defined as

$$\Pi_\pm(f) = \begin{cases} f, & \text{when sign}(f) = \pm 1; \\ 0 & \text{otherwise.} \end{cases} \tag{7.86}$$

A similar probabilistic interpretation of the current has been proposed by Alexander Vilenkin as a candidate for a probability distribution in quantum geometrodynamics [328]. Note, however, that the charge density fails as a probability distribution near the seed and more importantly near a pair-creation event, where the WKB approximation is no longer valid and the particle exhibits its quantum mechanical behaviour. The proposal only works whenever the particle behaves classically and does not tunnel.

7.3 Relativistic weak value theory

In the previous section we considered the amplitude corresponding to the propagation from the state $|\psi_0\rangle$ to the state $|x_1^\mu\rangle$, *i.e.*,

$$\varphi(x_1^\mu) = \int_{0^+}^{\infty} \mathrm{d}s \ \langle x_1^\mu | e^{-is\hat{H}/\hbar} | \psi_0 \rangle \,. \tag{7.87}$$

We subsequently argued that the absolute square of the amplitude, $|\varphi(x_1^\mu)|^2$, can be interpreted as the relative probability for the spacetime point x_1^μ to be part of the worldline of the particle starting in state $|\psi_0\rangle$. It is natural to extend this interpretation to arbitrary final states, *i.e.*, the amplitude to propagate from the initial state $|\psi_0\rangle$ to the final state $|\psi_1\rangle$ is given by

$$\int_{0^+}^{\infty} \mathrm{d}s \, \langle \psi_1 | e^{-is\hat{H}/\hbar} | \psi_0 \rangle = \int \mathrm{d}x_0^\mu \mathrm{d}x_1^\mu \, \psi_1^*(x_1^\mu) G[x_1^\mu; x_0^\mu] \psi_0(x_0^\mu) \tag{7.88}$$

$$= \int \mathrm{d}x_1^\mu \, \psi_1^*(x_1^\mu) \varphi(x_1^\mu) \,, \tag{7.89}$$

with the final state $|\psi_1\rangle$ in the position representation

$$\psi_1(x^\mu) = \langle x^\mu | \psi_1 \rangle \,. \tag{7.90}$$

As for the tree, we propose to interpret the absolute value squared,

$$P[\psi_1 | \psi_0] = \left| \int_{0^+}^{\infty} \mathrm{d}s \, \langle \psi_1 | e^{-is\hat{H}/\hbar} | \psi_0 \rangle \right|^2 , \tag{7.91}$$

as a relative probability in the space of post-selections. This places the initial and final state of the propagator on the same footing.

When the particle managed to propagate from $|\psi_0\rangle$ to $|\psi_1\rangle$ with a reasonable relative probability with respect to other configurations, it is natural to ask how it managed to do so. Generally, this question is not well-defined in quantum physics, as the classical measurement – and corresponding collapse of the wavefunction – of an intermediate state influences the outcome of the experiment. That is to say, in the double slit experiment, we can generally not measure through which slit the particle went while preserving the interference pattern. Moreover, if we attempt to continuously measure the jump of an electron between two Bohr orbits, we will not observe the transition. The electron will simply remain in its initial orbit. This is known as the quantum Zeno's paradox [5].

However, not withstanding these limitations, there exists a way in which we can probe the average evolution between the states $|\psi_0\rangle$ and $|\psi_1\rangle$, while preserving the outcome of the experiment. We couple the experiment to a quantum mechanical measuring device and study the limit of weak coupling. This is known as weak value theory, as developed by Yakir Aharonov and collaborators for non-relativistic quantum mechanics. For a detailed exposition see [5]. We here extend weak value theory to relativistic quantum mechanics and study its implications in Schwinger pair creation. For a detailed analysis of weak measurements of relativistic systems, by coupling the system to a von Neumann pointer, we refer the reader to appendix 7.C.

There exist a range of weak measurements, which allow us to probe a quantum process without interfering with the experiment. These measurements are all conditional with respect to an initial pre-selection $|\psi_0\rangle$ and a final post-selection $|\psi_1\rangle$. Note that the states $|\psi_0\rangle$ and $|\psi_1\rangle$ are initial and final states as described from the perspective of the particle, *i.e.* the pre-selection of an antiparticle is typically in the future of the post-selection on the spacetime manifold as the particle moves backwards in embedding time. The two states are bridged by the Schrödinger evolution.

7.3.1 The relativistic Aharonov-Bergmann-Lebowitz formula

The Aharonov-Bergmann-Lebowitz (ABL) formula is a time symmetric interpretation of quantum mechanics which, assuming a pre-selection ψ_0 at time $t = 0$ and a post-selection ψ_1 at time $t = T$, attempts to capture the probability distribution of the particle at an

intermediate time $t \in [0,T]$ [7], *i.e.*,

$$P_{ABL}(x_m,t) = \frac{\left|\langle\psi_1(t=T)|e^{-i\hat{H}(T-t)/\hbar}|x_m\rangle\langle x_m|e^{-i\hat{H}t/\hbar}|\psi_0(t=0)\rangle\right|^2}{\int\left|\langle\psi_1(t=T)|e^{-i\hat{H}(T-t)/\hbar}|x_m\rangle\langle x_m|e^{-i\hat{H}t/\hbar}|\psi_0(t=0)\rangle\right|^2 \mathrm{d}x_m}\,. \tag{7.92}$$

The initial state ψ_0 is propagated forward by a time t and the final state is propagated backwards by a time $T-t$. The probability distribution is then normalized to unity. When the pre- and post-selection have a reasonable overlap, $\left|\langle\psi_1|e^{-i\hat{H}T/\hbar}|\psi_0\rangle\right|^2$, and the WKB approximation holds, the ABL-formula will follow the classical or semi-classical evolution of the particle.

For the setting of a relativistic particle we propose a relativistic ABL-formula for the probability distribution at a fraction of its evolution $r \in [0,1]$ given that the particle started (at $r=0$) in state $|\psi_0\rangle$ and was later (at $r=1$) observed in state $|\psi_1\rangle$,

$$P_{ABL}(x^\mu_m,r) = \frac{\left|\int_{0^+}^{\infty}\langle\psi_1|e^{-i\hat{H}s(1-r)/\hbar}|x^\mu_m\rangle\langle x^\mu_m|e^{-i\hat{H}sr/\hbar}|\psi_0\rangle\mathrm{d}s\right|^2}{\int\left|\int_{0^+}^{\infty}\langle\psi_1|e^{-i\hat{H}s(1-r)/\hbar}|x^\mu_m\rangle\langle x^\mu_m|e^{-i\hat{H}sr/\hbar}|\psi_0\rangle\mathrm{d}s\right|^2 \mathrm{d}x^\mu_m}\,. \tag{7.93}$$

We propagate the initial state $|\psi_0\rangle$ forward by a Schwinger time sr and evolve the final state $|\psi_1\rangle$ backwards by a Schwinger time $s(1-r)$. The proposed relativistic ABL-formula differs from the conventional non-relativistic ABL-formula by the integral over the Schwinger time s. In terms of the path integral formalism, the relativistic ABL-formula is given by

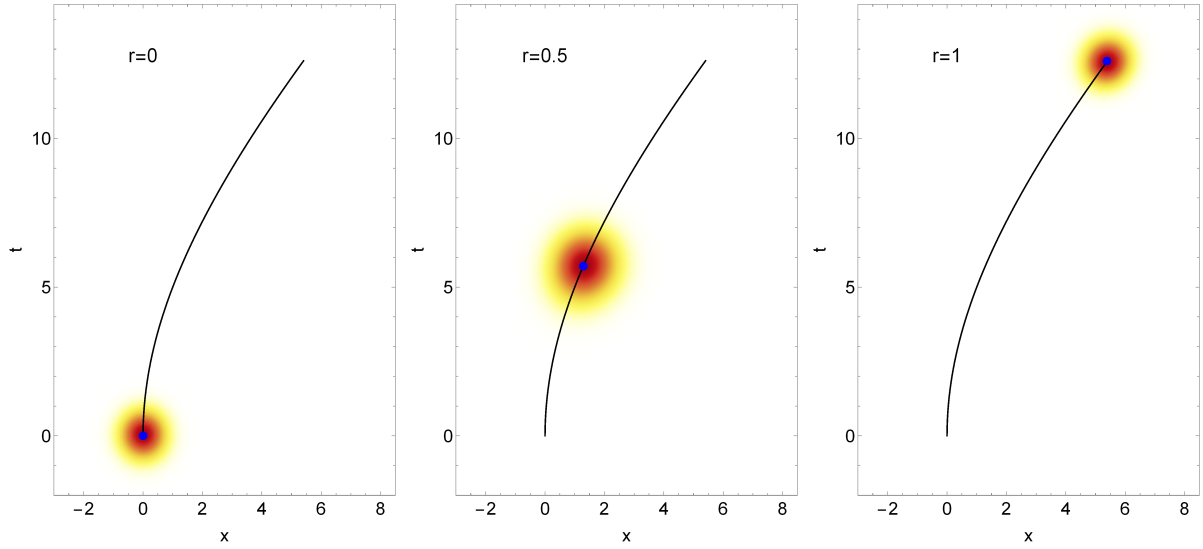
$$P_{ABL}(x^\mu_m,r) = \frac{\left|\int_{0^+}^{\infty}\mathrm{d}s\int_{\psi_0}^{\psi_1}\mathcal{D}x^\mu\, e^{iS[x^\mu]/\hbar}\delta(x^\mu_m - x^\mu(rs))\right|^2}{\int\left|\int_{0^+}^{\infty}\mathrm{d}s\int_{\psi_0}^{\psi_1}\mathcal{D}x^\mu\, e^{iS[x^\mu]/\hbar}\delta(x^\mu_m - x^\mu(rs))\right|^2 \mathrm{d}x^\mu_m}\,. \tag{7.94}$$

Note that, for the path integral, we use the shorthand

$$\int_{\psi_0}^{\psi_1}\mathcal{D}x\, Q[x] = \int \mathrm{d}x^\mu_0\mathrm{d}x^\mu_1\,\psi^*_1(x^\mu_1)\left[\int_{x^\mu_0(0)=x^\mu_0}^{x^\mu_1(s)=x^\mu_1}\mathcal{D}x\, Q[x]\right]\psi_0(x^\mu_0)\,, \tag{7.95}$$

for arbitrary functionals $Q[x]$.

In Figure 7.6, we evaluate the relativistic ABL-formula for a charged quantum particle in a weak electric field, conditioned with a Gaussian pre- and post-selection which are classically connected (for the derivation of the wavefunction see appendix 7.A). We observe



(a) $r = 0$ (b) $r = 1/2$ (c) $r = 1$

Figure 7.6: The relativistic ABL-formula for time-like separated boundary conditions, for a scalar particle in a background electric field with $E = 0.5$. The black line is the classical trajectory, the blue dot is the classical position of the particle at the corresponding fraction of proper time, while the density plot is the probability distribution of the position according to the relativistic ABL-formula (7.94).

that the relativistic ABL-formula nicely follows the classical trajectory of the particle. Note that the ABL-formula will only give reasonable results when we post-select with a state $|\psi_1\rangle$ which has a high relative probability with respect to the initial state $|\psi_0\rangle$. We will not consider exponentially suppressed combinations of initial and final states, as is often analysed in the literature on weak value theory [6].

In Figure 7.7, we evaluate the ABL-formula for an relativistic particle in a relatively strong electric field. We symmetrically select the Gaussian initial and final states $|\psi_0\rangle, |\psi_1\rangle$ to be close to on-shell and to have a relatively large overlap when one is propagated to the other,

$$\left| \int \mathrm{d}x_0^\mu \, \mathrm{d}x_1^\mu \, \psi_1^*(x_1^\mu) G[x_1^\mu; x_0^\mu] \psi_0(x_0^\mu) \right|^2 . \tag{7.96}$$

The particle initially ($r = 0$) resembles an antiparticle moving to the right. The particle decelerates (at $r = 0.2$) and tunnels into a particle while turning around in the spatial direction ($r = 0.5$). The particle is subsequently accelerated to the right ($r = 0.8$) to arrive

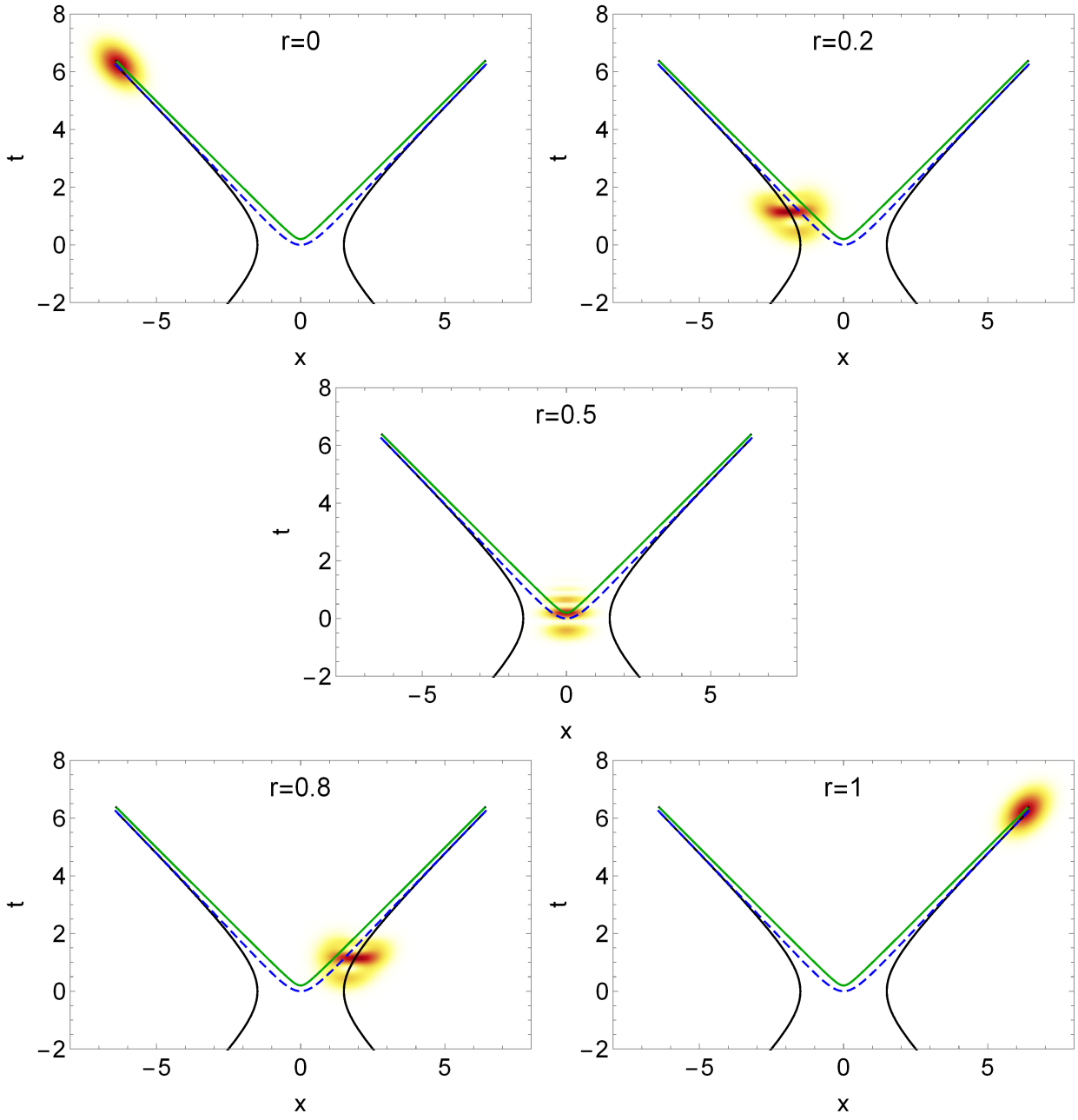


Figure 7.7: The ABL-formula for space-like separated boundary conditions – the Schwinger process. The blue dashed line is the complex classical trajectory, the green line is the weak trajectory, while the density plot is the probability distribution of the position according to the ABL-formula. The two classical wordlines are separated at $t = 0$ by the Schwinger tunnelling length $L_c = \frac{2m}{|eE|}$.

at the post-selection ($r = 1$). Note that relativistic ABL-formula at $r = 0.5$ consists of a number of fringes while the particle is tunnelling in the pair-creation event. Remark that the ABL-formula closely matches the classical (black) trajectory for the antiparticle or particle in the initial and final part of its evolution.

7.3.2 The weak trajectory

Additional information on the evolution of the particle can be obtained by weakly measuring the trajectory of the particle. We define the relativistic weak measurement of an observable Q of a relativistic particle as the expectation value

$$Q_w(r) = \frac{\int_{0^+}^{\infty} \mathrm{d}s \, \langle \psi_1 | e^{-i\hat{H}s(1-r)/\hbar} Q[\hat{x}^\mu, \hat{p}^\mu] e^{-i\hat{H}sr/\hbar} | \psi_0 \rangle}{\int_{0^+}^{\infty} \mathrm{d}s \, \langle \psi_1 | e^{-i\hat{H}s/\hbar} | \psi_0 \rangle} \,. \tag{7.97}$$

As for the ABL-formula, we evolve the initial state, $|\psi_0\rangle$, a Schwinger time sr forward, and evolve the final state $|\psi_1\rangle$ a Schwinger time $s(1-r)$ backwards. We subsequently evaluate the expectation value of the position operator at this time and integrate over the total Schwinger time. The relativistic weak value differs from the non-relativistic weak value by the integral over the Schwinger time. See appendix 7.C for the derivation of the relativistic weak value formula. In the path integral representation, the relativistic weak value takes the form

$$Q_w(r) = \frac{\int_{0^+}^{\infty} \mathrm{d}s \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu \, e^{iS[x^\mu]/\hbar} Q[x^\mu]}{\int_{0^+}^{\infty} \mathrm{d}s \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu \, e^{iS[x^\mu]/\hbar}} \,. \tag{7.98}$$

When the initial and the final states coincide, *i.e.*, $|\psi_0\rangle = |\psi_1\rangle$, the weak value is a conventional expectation value. The weak value of a Hermitian observable with coinciding initial and final states is real-valued. In quantum field theory it is common to insert Hermitian observables between two vacuum states. For more general initial and final states, the weak value of a Hermitian operator is complex-valued. In order to obtain a valid interpretation for these complex numbers, it is important to carefully model the measurement by coupling the observable to an idealized von Neumann pointer. The real part of the weak value generally results in a shift in the mean position of the von Neumann pointer,

$$\langle X_1 \rangle = \langle X_0 \rangle + gT \, \mathrm{Re}[Q_w] \,, \tag{7.99}$$

while the imaginary part is generally realized as a shift in the mean momentum,

$$\langle P_1 \rangle = \langle P_0 \rangle + \frac{Tg\hbar}{2\Delta_i^2} \, \mathrm{Im}[Q_w] \,, \tag{7.100}$$

with the initial (X_0, P_0) and the final position and momentum (X_1, P_1) of the pointer, the initial standard deviation of the pointer Δ_i, the duration of the measurement T and the coupling strength g. For a detailed derivation see appendix 7.C.

We study the trajectory of the relativistic particle propagating between the states $|\psi_0\rangle$ and $|\psi_1\rangle$ by measuring the spacetime position of the particle. We define the *weak trajectory* as the expectation value of the trajectory at a fraction $r \in [0, 1]$ of its evolution, *i.e.*,

$$x^\mu_w(r) = \frac{\int_{0^+}^{\infty} \mathrm{d}s \, \langle\psi_1|e^{-i\hat{H}s(1-r)/\hbar}\hat{x}^\mu(rs)e^{-i\hat{H}sr/\hbar}|\psi_0\rangle}{\int_{0^+}^{\infty} \mathrm{d}s \, \langle\psi_1|e^{-i\hat{H}s/\hbar}|\psi_0\rangle}\,, \tag{7.101}$$

or in terms of the path integral formulation,

$$x^\mu_w(r) = \frac{\int_{0^+}^{\infty} \mathrm{d}s \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu \, e^{iS[x^\mu]/\hbar} x^\mu(rs)}{\int_{0^+}^{\infty} \mathrm{d}s \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu \, e^{iS[x^\mu]/\hbar}}\,. \tag{7.102}$$

We can define an analogous weak momentum trajectory

$$p^\mu_w(r) = \frac{\int_{0^+}^{\infty} \mathrm{d}s \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu \, e^{iS[x^\mu]/\hbar} p^\mu(rs)}{\int_{0^+}^{\infty} \mathrm{d}s \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu \, e^{iS[x^\mu]/\hbar}}\,. \tag{7.103}$$

The weak trajectory is by construction localized near the support of the ABL-formula. It thus follows from the analysis above that for a classically reachable final state $|\psi_1\rangle$ given an initial state $|\psi_0\rangle$ the weak trajectory will closely follow the classical trajectory (see Figure 7.6). For these initial and final states, the path integral is generally dominated by a single classical trajectory. It follows that the weak trajectory coincides with this classical trajectory (provided it exists) in the classical limit $\hbar \to 0$.

For initial and final states, which are not connected by classical evolution the situation is more involved (see Figure 7.7). The weak trajectory is again typically described by a single solution to the equations of motion. This solution will however generally be a complex trajectory. In the semi-classical limit $\hbar \to 0$, the weak trajectory will again approach this complex classical trajectory, but the amplitude for the process will tend to zero. In Figure 7.7, the real part of the weak trajectory is plotted in green, while the real part of the corresponding complex classical trajectory is plotted in blue.

In general, the complex classical trajectory of a relativistic particle satisfies both the equation of motion

$$\frac{\partial S}{\partial x^\mu} = 0\,, \tag{7.104}$$

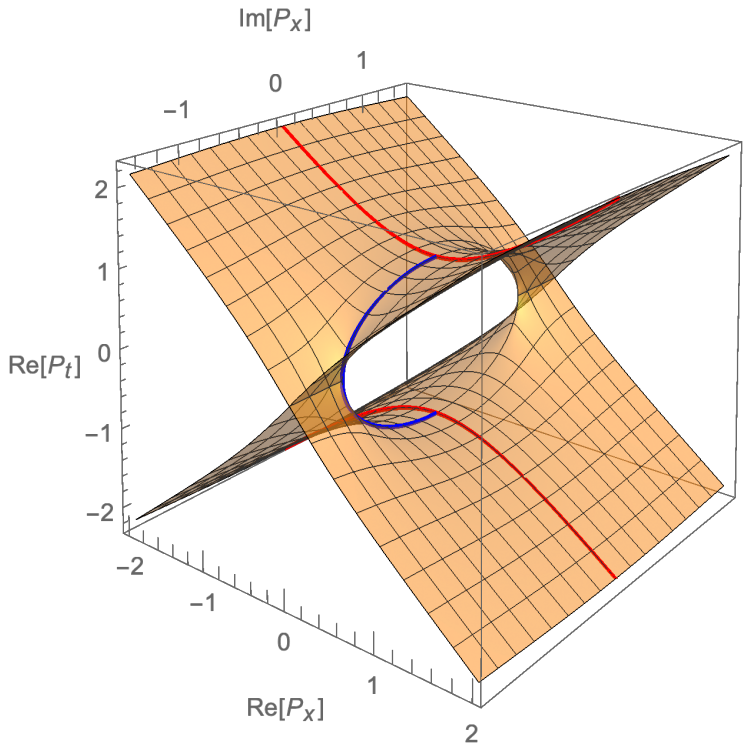


Figure 7.8: The constraint surface $\tilde{H}[x^\mu; p^\mu] = 0$ in the complex space of the kinetic momentum P^μ (in yellow). The red hyperbolas are the intersection of the constraint with real momentum space. The upper hyperbola represents particle states while the lower hyperbola represents antiparticle states. The blue curve is an example of a complex classical trajectory interpolating between the two real branches.

and the constraint

$$\tilde{H}[x^\mu, p^\mu] = 0\,, \tag{7.105}$$

and is supplied with the boundary conditions $x^\mu(0) = x_0^\mu$, $x^\mu(1) = x_1^\mu$. In the setting of a relativistic particle in an electric field, the Hamiltonian constraint is a condition

$$\tilde{H}[x^\mu; p^\mu] = \frac{1}{2m}\left[-P_t^2 + P_x^2 + m^2\right] = 0\,. \tag{7.106}$$

with the kinetic momentum $P^\mu = p^\mu - eA^\mu$. The solution of this equation in the real (P_t, P_x)-plane for non-zero mass consists of two disconnected hyperboloids. On the first, the particle has positive momentum in the time direction, *i.e.*, $P_t > 0$ corresponding to

particle, and on the second, it has negative momentum in the time direction, *i.e.*, $P_t < 0$ corresponding to an antiparticle. The complex classical trajectory interpolating between a particle and an antiparticle is a complex solution lying within the constraint surface, $\tilde{H} = 0$, within complexified $(P_t, P_x) \in \mathbb{C}^2$ space. See Figure 7.8 for an illustration of a projection of the constraint surface in the complex (P_t, P_x) space. The two red lines are the two hyperboloids representing the particle and antiparticle states. The yellow surface is the constraint surface $\tilde{H} = 0$. The blue line is an example of a complex classical solution interpolating between the two real branches. We have found that the weak momentum trajectory $p_w(r)$ corresponding to the pair-creation example plotted in Figure 7.7 generally follows a trajectory close to the constraint surface for finite Planck constant $\hbar$. For more details see [135].

Complex classical trajectories are commonly used in the semi-classical analysis of path integrals while studying tunnelling phenomena. They are the saddle points of the integrand of the path integral. The relevant saddle points can be used to approximate the integral with the saddle point method. We propose the weak trajectory as a generalization of the complex classical trajectory with a proper probabilistic interpretation, obtained with the idealized von Neumann pointers. Remark that the complex classical trajectory is often difficult to obtain numerically, as it generally requires a shooting analysis with complex boundary conditions. Moreover, it is often not clear whether the derived complex classical trajectory is relevant to the path integral, as functional Picard-Lefschetz analysis is in general difficult to execute. The weak trajectory on the other hand can be evaluated with straightforward numerical methods and is moreover, guaranteed to be relevant to the path integral, assuming that the path integral is dominated by a single saddle point.

7.3.3 The weak density

We conclude our analysis with the weak number density and weak charge density as measured in the lab frame. The weak number density in spacetime is the relativistic weak expectation value of the observable

$$Q[x(sr)] = \delta(x_m^\mu - x^\mu(sr))\,. \tag{7.107}$$

The von Neumann pointer is positioned at the point x_m^μ and registers the probability for the particle to pass through x_m^μ, *i.e.*,

$$n_w(x_m^\mu) = \frac{\int_0^1 \mathrm{d}r \int_{0^+}^\infty \mathrm{d}s \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu \, e^{iS[x^\mu]/\hbar} \delta(x_m^\mu - x^\mu(sr))}{\int_{0^+}^\infty \mathrm{d}s \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu \, e^{iS[x^\mu]/\hbar}}\,. \tag{7.108}$$

Note that we integrate over the variable r, as it is irrelevant when the particle crosses x^μ_m during its evolution.

The weak number density for the Schwinger effect is plotted in Figure 7.9. The number density along the initial and final leg of its evolution, where the behaviour is approximately classical, is nearly real-valued. Near the pair creation region, the particle tunnels and produces an interference pattern. The observed fringes, in the relativistic weak number density, correspond to the fringes observed in the ABL-formula (see Figure 7.7). Note that weak number density decreases along the legs in the classical regime. This is the result of the acceleration of the particle in the electric field. When the particle moves with a higher velocity it spends less time in a given location.

The spacetime charge density of the particle, with respect to the time foliation

$$\Sigma_{t'} = \{(t, x) | t = t'\}, \tag{7.109}$$

is given by the observable

$$Q[x(sr)] = e\dot{x}^0 \delta(x^\mu_m - x^\mu(sr)) \tag{7.110}$$

$$= \frac{eP_0}{m} \delta(x^\mu_m - x^\mu(sr)) \,. \tag{7.111}$$

The weak charge density ρ_w is given by the weak measurement

$$\rho_w(x^\mu_m) = \frac{e}{m} \frac{\int_0^1 \mathrm{d}r \int_{0^+}^\infty \mathrm{d}s \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu \, e^{iS[x^\mu]/\hbar} P_0 \delta(x^\mu_m - x^\mu(sr))}{\int_{0^+}^\infty \mathrm{d}s \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu \, e^{iS[x^\mu]/\hbar}} \,. \tag{7.112}$$

In the canonical representation, the kinetic momentum is represented by the directional derivative $P_0 = -\frac{i}{2}\overleftrightarrow{D_0}$. We need to use the directional derivative, to ensure $-\frac{i}{2}\phi^* \overleftrightarrow{D_0} \phi$ to be real-valued. The weak charge density for the Schwinger effect is plotted in Figure 7.10. The spacetime charge density is qualitatively similar to the relativistic weak number density. The weak charge density is largely real valued in the classical regime and shows a number of fringes in the pair-creation regime. The main difference is that the charge density does not decay while the particle is being accelerated. This is a result of the weighting with the kinetic momentum P_0 (which is proportional to the Lorentz γ factor) in the integral. The particle on the left is an antiparticle represented by the blue region. The particle on the right is a particle represented by the red region. Remark that the weak charge density is locally conserved with respect to embedding time away, from the support of the initial and final states, *i.e.*, the weak charge current j^μ_w can be shown to be conserved, $\partial_\mu j^\mu_w$ away from the support of $|\psi_0\rangle, |\psi_1\rangle$.

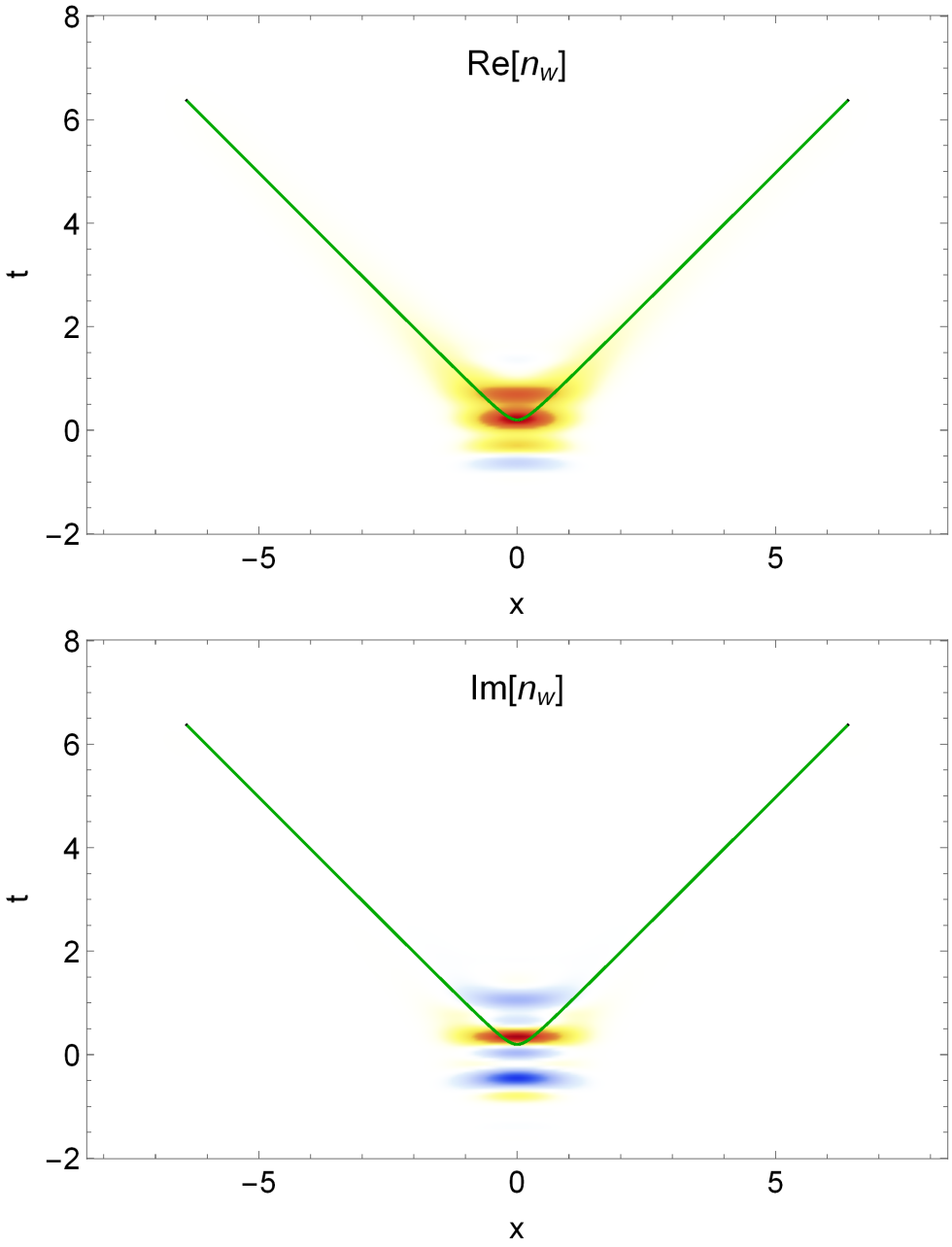


Figure 7.9: The real and imaginary part of the weak number density n_w in Schwinger pair creation together with the weak trajectory in green. Red represents positive function values, whereas blue represents negative function values.

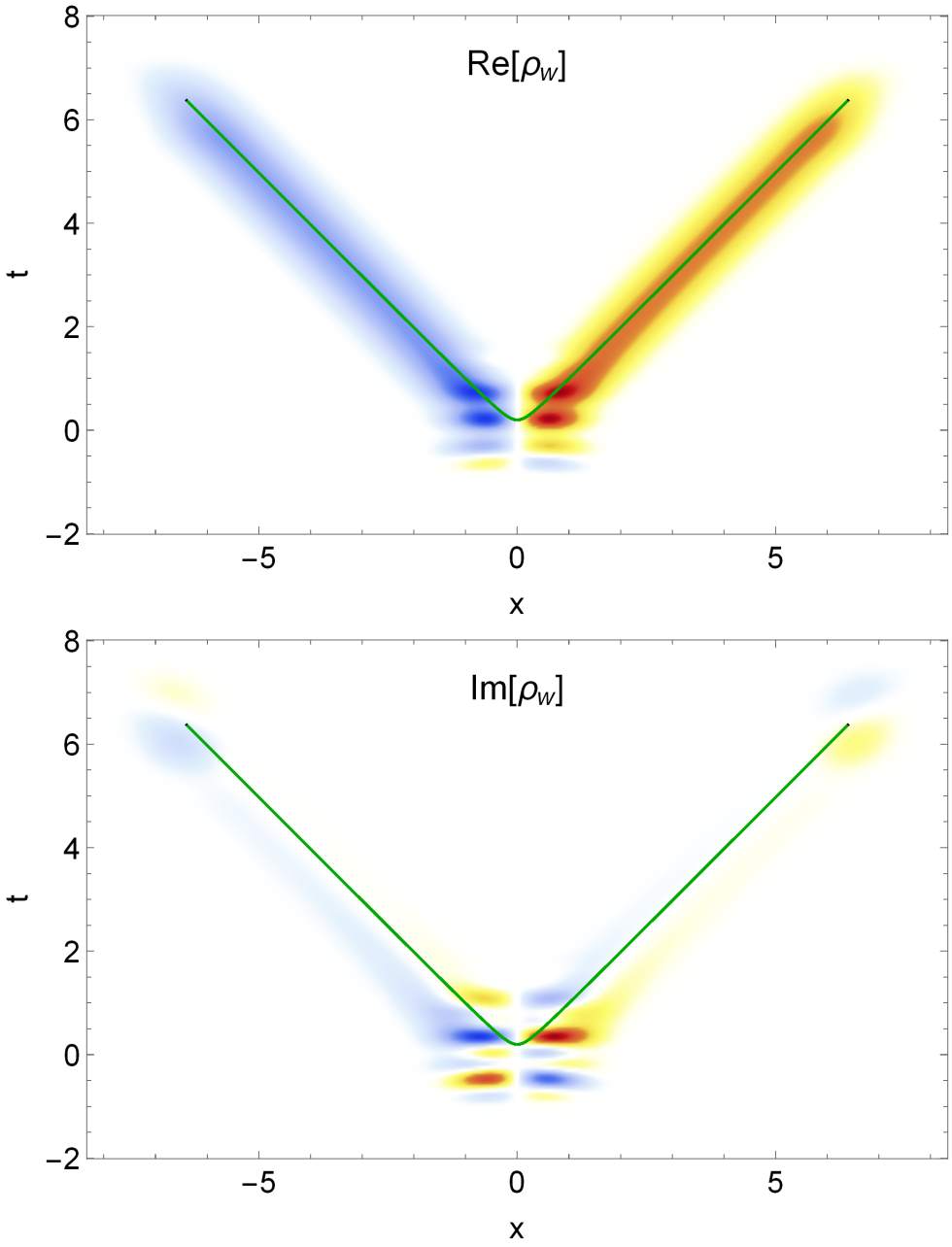


Figure 7.10: The real and imaginary part of the weak charge density ρ_w in Schwinger pair creation; positive charge in red, negative in blue. The weak trajectory is shown in green. Red represents positive function values, whereas blue represents negative function values.

Finally, we use the weak charge density to estimate the back-reaction of the pair-creation event on the electric field, to first order. In a $1+1$-dimensional spacetime, the electric field is topological, and does not have propagating degrees of freedom. Using the Maxwell equations and the weak charge density, we evaluate the corrected electric field

$$E_w(x,t) = E + \int_{-\infty}^{x} \mathrm{d}x' \, \rho_w(x',t) \,. \tag{7.113}$$

The weak electric field for the Schwinger effect is plotted in Figure 7.11. The white background represents the original, unperturbed field. In the pair-creation zone, the electric field oscillates and is complex valued. In the classical evolution of the particle, the net electric field decreases in between the particle- antiparticle pair. After the pre- and post-selection events, the electric field increases to its original value, so that the total energy is conserved. These results are consistent with and considerably generalize the weak value analysis of Brout et al. [58] obtained using quantum field theory methods. In addition, they seem to us conceptually simpler.

7.4 Conclusion

Quantum field theory is an amazingly successful theory of relativistic quantum effects. It has led to spectacular results in quantum electrodynamics and has likewise been very fruitful in quantum chromodynamics and the standard model of particle physics. However, its generalization to a theory of quantum gravity has so far failed. It is moreover difficult to model the localized evolution of a single relativistic particle moving, for example, in a cloud chamber. The on-shell momentum states $|\mathbf{p}\rangle$ have a sharply defined four-momentum p^μ, but are completely de-localized in spacetime. This is, therefore, not a realistic model of the particles we observe in nature and is moreover a serious limitation in the development of quantum field theory in a general curved spacetime and in quantum gravity.

In this Chapter we propose a new spacetime amplitude interpretation for relativistic quantum mechanics in which we focus on the localized evolution of a single particle. Rather than using the Feynman propagator as a computational device for the evaluation of Feynman diagrams, we pursue a more literal interpretation in which the Feynman propagator is the amplitude to propagate between two spacetime points. The Feynman propagator is used to propagate a seed $|\psi_0\rangle$, namely a localized superposition of spacetime position states $|x^\mu\rangle$, into a tree $|\varphi\rangle$ representing the seeds evolution. We have shown that the tree can be interpreted as the quantum mechanical worldline of a particle. Using the function G_H, in Minkowski spacetime coinciding with the Hadamard function (the real part of the

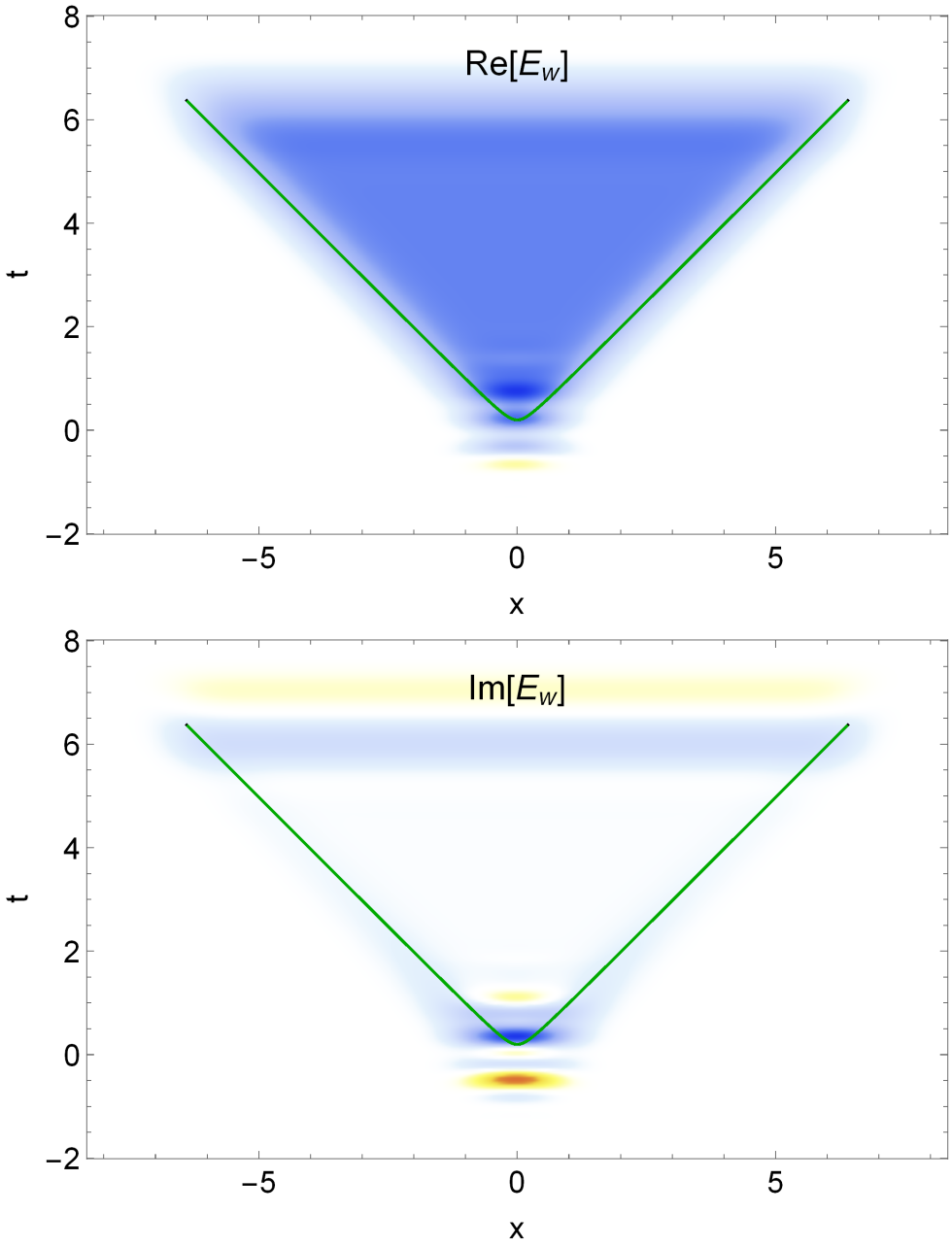


Figure 7.11: The real and imaginary part of the weak electric field E_w in Schwinger pair creation, and the weak trajectory in green. The background white here represents $E(x,t) = E$.

Feynman propagator), can be used to extend the tree to a on-shell spacetime amplitude, which solves the Klein-Gordon equation.

We moreover show how weak value theory can be extended to relativistic systems, enabling us to study the evolution of a quantum mechanical particle propagating between an initial and a final state. We construct the relativistic weak trajectory, which is a fully quantum mechanical generalization of the complex classical trajectories (instantons) often used in the study of tunnelling phenomena. We furthermore define the relativistic weak number density and the relativistic weak charge density, and use the latter to study the backreaction of Schwinger pair creation on the electric field. A detailed analysis of the Schwinger effect will appear in an accompanying paper [135]. The relativistic weak number and the relativistic weak charge density of the relativistic particle in the Schwinger effect, obtained using relativistic quantum mechanics, nicely complement the results obtained by Brout et al. [58] in quantum field theory.

The here proposed techniques work well for a scalar particle in Minkowski spacetime. In an upcoming series of papers we hope to apply these techniques to a number of different problems in particle physics. We are currently applying the weak trajectory to the study of tunnelling phenomena in quantum mechanics. The techniques are directly applicable to the study of vacuum decay. We also study the evolution of spin-$\frac{1}{2}$ particles described by the Dirac equation, in order to re-examine the Klein paradox. We are furthermore analysing the propagation of relativistic particles in de Sitter spacetime.

We are also applying these techniques to quantum geometrodynamics. The spacetime amplitude picture of the relativistic particle presented in this Chapter is conceptually very close to John Wheeler's original vision of quantum gravity in terms of interference patterns (see Chapter 2). The weak trajectory, in this context, is a representation of a quantum spacetime interpolating between initial and final three-geometries. In a recently posted paper, we use the seed and tree analysis to study of the Bunch-Davies vacuum as an initial state for inflation [101]. In an upcoming publication we plan to use the spacetime amplitudes and relativistic weak value formalism to study the foundations of quantum geometrodynamics, paying special attention to the problem of time and emergence of spacetime [137].

7.A The Feynman propagator of a particle in a constant electromagnetic field

The action for a charged spin-0 particle in $1+1$-dimensional spacetime is given by

$$S[\dot{x}^\mu] = \int_0^1 \left[\frac{m}{2s}\dot{x}_\mu\dot{x}^\mu - \frac{m}{2}s + e\,\dot{x}_\mu A^\mu\right]\mathrm{d}\tau\,, \tag{7.114}$$

with the mass m, the electric charge e, the electromagnetic vector potential A_μ and the Schwinger time s. We will here consider the Feynman propagator of a particle in a constant electric field E described by the vector potential $A^\mu = (-Ex, 0)$.

Since the corresponding Feynman propagator for a fixed Schwinger time,

$$G[x_1^\mu; x_0^\mu; s] = \int_{x^\mu(0)=x_0^\mu}^{x^\mu(1)=x_1^\mu} \mathcal{D}x^\mu\, e^{iS[x^\mu]/\hbar}\,, \tag{7.115}$$

is a Gaussian integral, we can evaluate it by the saddle point method [174]. The equation of motion

$$\ddot{t} = \frac{seE}{m}\dot{x}\,, \quad \ddot{x} = \frac{seE}{m}\dot{t}, \tag{7.116}$$

with the boundary conditions $x^\mu(0) = x_0^\mu = (t_0, x_0)$, and $x^\mu(0) = x_1^\mu = (t_0, x_0)$, is solved by the worldline $\bar{x}^\mu(\tau) = (\bar{t}(\tau), \bar{x}(\tau))$ with

$$\bar{t}(\tau) = \operatorname{csch}\left[\frac{seE}{2m}\right]\sinh\left[\frac{seE}{2m}\tau\right]\left[\Delta t\cosh\left[\frac{seE}{2m}(\tau-1)\right] + \Delta x\sinh\left[\frac{seE}{2m}(\tau-1)\right]\right] + t_0\,, \tag{7.117}$$

$$\bar{x}(\tau) = \operatorname{csch}\left[\frac{seE}{2m}\right]\sinh\left[\frac{seE}{2m}\tau\right]\left[\Delta x\cosh\left[\frac{seE}{2m}(\tau-1)\right] + \Delta t\sinh\left[\frac{seE}{2m}(\tau-1)\right]\right] + x_0\,, \tag{7.118}$$

where the differences $\Delta t = t_1 - t_0$, $\Delta x = x_1 - x_0$ and $\tau \in [0, 1]$. The classical action is given by

$$S[\bar{x}^\mu(\tau)] = \frac{eE}{4}\left[2\Delta t\Delta x + (\Delta t^2 - \Delta x^2)\coth\left(\frac{seE}{2m}\right)\right] - \frac{ms}{2}\,. \tag{7.119}$$

Using the classical action, we evaluate the path integral [174, 119], *i.e.*,

$$G[x_1^\mu; x_0^\mu; s] = \frac{1}{2\pi i\hbar}\sqrt{\det\left(-\frac{\partial^2 S[\bar{x}^\mu]}{\partial x_0^\mu \partial x_1^\nu}\right)}\, e^{iS[\bar{x}^\mu]/\hbar}\,, \tag{7.120}$$

with the determinant in the prefactor

$$\sqrt{\det\left(-\frac{\delta^2 \bar{S}}{\delta x_0^\mu \delta x_1^\nu}\right)} = i\frac{|eE|}{2}\sqrt{1+\coth^2\left(\frac{eEs}{2m}\right)}\,. \tag{7.121}$$

The Feynman propagator between the spacetime points x_0^μ and x_1^μ can thus be expressed as a one-dimensional integral over the Schwinger time

$$G[x_1^\mu; x_0^\mu] = \int_{0^+}^{\infty} \mathrm{d}s\, G[x_1^\mu; x_0^\mu; s] \tag{7.122}$$

$$= \frac{|eE|}{4\pi\hbar}\int_{0^+}^{\infty}\mathrm{d}s\,\sqrt{1+\coth^2\left(\frac{eEs}{2m}\right)}\, e^{\frac{i}{\hbar}\left[\frac{eE}{4}\left(2\Delta t\Delta x+(\Delta t^2-\Delta x^2)\coth\left(\frac{seE}{2m}\right)\right)-\frac{ms}{2}\right]}\,. \tag{7.123}$$

The corresponding Hadamard function is obtained by integrating over the real line

$$G_H[x_1^\mu; x_0^\mu] = \frac{|eE|}{4\pi\hbar}\int_{-\infty}^{\infty}\mathrm{d}s\,\sqrt{1+\coth^2\left(\frac{eEs}{2m}\right)} \tag{7.124}$$

$$\times\cos\left[\frac{i}{\hbar}\left[\frac{eE}{4}\left(2\Delta t\Delta x+(\Delta t^2-\Delta x^2)\coth\left(\frac{seE}{2m}\right)\right)-\frac{ms}{2}\right]\right]\,. \tag{7.125}$$

Since classical action $S[\bar{x}^\mu]$ is quadratic in the boundary conditions x_0^μ, x_1^μ we can solve for the wavefunction $\psi(x_1^\mu; s)$ originating from the Gaussian initial state

$$\psi_0(x^\mu) = \frac{1}{\sqrt{2\pi\sigma_t\sigma_x}} e^{-\frac{(t-\mu_t)^2}{4\sigma_t^2}-\frac{(x-\mu_x)^2}{4\sigma_x^2}-i(p_t-eA_t)(t-\mu_t)+i(p_x-eA_x)(x-\mu_x)}\,, \tag{7.126}$$

by evaluating the Gaussian integral

$$\psi(x_1^\mu; s) = \int \mathrm{d}x_0^\mu\, G[x_1^\mu; x_0^\mu; s]\psi_0(x_0^\mu)\,. \tag{7.127}$$

The tree of a relativistic particle in a constant electric field originating from a Gaussian seed can thus be written as a one-dimensional integral

$$\varphi(x_1^\mu) = \int_0^\infty \mathrm{d}s\,\psi(x_1^\mu; s)\,, \tag{7.128}$$

which we evaluate numerically.

7.B Numerical evaluation of the spacetime amplitude

In the previous section we analysed the Feynman propagator and the tree of a relativistic charged particle in a constant electric field. The path integral is Gaussian and can be evaluated with analytic techniques. Unfortunately, the class of electromagnetic fields for which we can evaluate the path integral exactly is limited. For these more complicated configurations we can still evaluate the wavefunction and tree of the particle using numerical techniques. There exist several techniques which can be used to numerically approximate the solutions of a Schrödinger equation. In this Chapter, for Figure 7.3, we used the Suzuki-Trotter method as described in [89, 90] and references therein.

7.B.1 The Suzuki-Trotter integrator

The Schrödinger-like equation

$$i\hbar\frac{\partial\psi}{\partial s} = \hat{H}\psi\,, \tag{7.129}$$

with the boundary condition

$$\psi(x^\mu; s=0) = \psi_0(x^\mu)\,, \tag{7.130}$$

can, in the case of a (Schwinger) time independent Hamiltonian, be solved by the exponentiation of the Hamiltonian operator

$$\psi(x^\mu; s) = e^{-is\hat{H}/\hbar}\psi_0(x)\,. \tag{7.131}$$

Solving the Schrödinger equation is thus reduced to exponentiating the Hamiltonian operator.

In the Suzuki-Trotter method we put the wave function on a lattice $\psi(x) \to \{\Psi_i\}_{i=1,\ldots,d}$ and approximate the Hamiltonian operator as a $d\times d$ Hermitian matrix. We subsequently decompose the Hamiltonian in block diagonal form $\hat{H} = K_1 + K_2 + \cdots + K_m$ and use the Lie-Trotter formula $e^{-is\hat{H}/\hbar} = \lim_{n\to\infty}(e^{-\frac{is}{n\hbar}K_1}\ldots e^{-\frac{is}{n\hbar}K_m})^n$ to approximate the solution as

$$\Psi = e^{-is\hat{H}/\hbar}\Psi_0 \approx \left[e^{-i\Delta s K_1/\hbar}\ldots e^{-i\Delta s K_m/\hbar}\right]^n \Psi_0\,, \tag{7.132}$$

with $\Delta s = s/n$. The action of the exponential of these block diagonal matrices on the wave function can be efficiently evaluated, as the exponentiation of a block diagonal matrix is equal to the exponentiation of its blocks.

In the analysis presented here, we partition the Hermitian matrix $\hat{H}$ into 2×2 blocks, following

$$\begin{pmatrix} \alpha & \beta & 0 & 0 & 0 \\ \beta^* & \alpha & \beta & 0 & 0 \\ 0 & \beta^* & \alpha & \beta & 0 \\ 0 & 0 & \beta^* & \alpha & \beta \\ 0 & 0 & 0 & \beta^* & \alpha \end{pmatrix} = \begin{pmatrix} \alpha & & & & \\ & \alpha & & & \\ & & \alpha & & \\ & & & \alpha & \\ & & & & \alpha \end{pmatrix} + \begin{pmatrix} 0 & \beta & & & \\ \beta^* & 0 & & & \\ & & 0 & \beta & \\ & & \beta^* & 0 & \\ & & & & 0 \end{pmatrix} + \begin{pmatrix} 0 & & & & \\ & 0 & \beta & & \\ & \beta^* & 0 & & \\ & & & 0 & \beta \\ & & & \beta^* & 0 \end{pmatrix} . \tag{7.133}$$

This method requires generally one Hermitian block diagonal matrix K_i for every diagonal band in the matrix $\hat{H}$. It thus suffices to evaluate the exponentiation of the diagonal 2×2 matrix

$$\mathrm{Exp}\left[i\Delta s \begin{pmatrix} \alpha & 0 \\ 0 & \alpha \end{pmatrix} \right] = \begin{pmatrix} e^{i\alpha\Delta s} & 0 \\ 0 & e^{i\alpha\Delta s} \end{pmatrix}, \tag{7.134}$$

and the Hermitian anti-diagonal 2×2 matrix

$$\mathrm{Exp}\left[i\Delta s \begin{pmatrix} 0 & \beta \\ \beta^* & 0 \end{pmatrix} \right] = \begin{pmatrix} \cos(|\beta|\Delta s) & \frac{i\beta}{|\beta|}\sin(|\beta|\Delta s) \\ \frac{i\beta^*}{|\beta|}\sin(|\beta|\Delta s) & \cos(|\beta|\Delta s) \end{pmatrix} . \tag{7.135}$$

The exponential of the diagonal matrix acts on the wavefunction by changing the phases of the wavefunction, *i.e.*,

$$\Psi_i \mapsto e^{i\alpha\Delta s}\Psi_i \, . \tag{7.136}$$

The exponential of the block diagonal matrix, consisting of anti-diagonal Hermitian 2×2 blocks, acting on the wavefunction, leads to to rotation of pairs of elements, *i.e.*,

$$\Psi_i \mapsto \cos(|\beta|\Delta s)\Psi_i + \frac{i\beta}{|\beta|}\sin(|\beta|\Delta s)\Psi_j \, , \tag{7.137}$$

$$\Psi_j \mapsto \frac{i\beta^*}{|\beta|}\sin(|\beta|\Delta s)\Psi_i + \cos(|\beta|\Delta s)\Psi_j \, . \tag{7.138}$$

Remark that both operation are unitary and thus preserves the norm of the wavefunction $\{\Psi_i\}$. Furthermore note that this method does not require the matrices $K_1, \ldots, K_m$ to be stored in memory.

The Suzuki-Trotter method can be made more accurate by symmetrizing the terms in the Lie-Trotter formula. The approximation of the Lie-Trotter formula described above is first order in Δs, *i.e.*,

$$\| e^{i\Delta s(A+B)} - e^{i\Delta s A} e^{i\Delta s B} \| \leq \frac{\Delta s^2}{2} \| [A, B] \| \, . \tag{7.139}$$

The corresponding numerical method,

$$\Psi \approx U_1(\Delta s)^n \Psi_0, \tag{7.140}$$

with $U_1(\Delta s) = e^{-i\Delta s K_1/\hbar} \dots e^{-i\Delta s K_m/\hbar}$, is known as Trotter-Suzuki-1. By symmetrization of the block diagonal matrices, the Lie-Trotter formula becomes quadratic in Δs, *i.e.*,

$$\|e^{i\Delta s(A+B)} - e^{i\Delta s A/2} e^{i\Delta s B} e^{i\Delta s A/2}\| \leq c\Delta s^3\,, \tag{7.141}$$

for some positive constant c. The corresponding numerical method

$$\Psi \approx U_2(\Delta s)^n \Psi_0, \tag{7.142}$$

with $U_2(\Delta s) = U_1^T(\Delta s/2) U_1(\Delta s/2)$ is known as Trotter-Suzuki-2. Using Trotter-Suzuki-2 we construct Trotter-Suzuki-4 as

$$\Psi \approx U_4(\Delta s)^n \Psi_0, \tag{7.143}$$

with

$$U_4(\Delta s) = U_2(p\Delta s) U_2(p\Delta s) U_2((1-4p)\Delta s) U_2(p\Delta s) U_2(p\Delta s), \tag{7.144}$$

where $p = 1/(4 - 4^{1/3})$. This method is fourth order in Δs since

$$\|e^{i\Delta s H} - U_4(\Delta s)\| \leq c\Delta s^5\,, \tag{7.145}$$

for a positive constant c.

In this Chapter we use the fourth order scheme in proper time to solve the Schrödinger equation. The tree is obtained by summing over the evolution of the wavefunction

$$\varphi_i = \sum_{j=0}^{n} \Psi_i(\Delta s j)\,, \tag{7.146}$$

where j runs from the initial condition $j = 0$ at Schwinger time 0 to the final condition $j = n$ at Schwinger time s. This is a good approximation of the tree in the region where the wavefunction vanishes for Schwinger time larger than s.

7.B.2 The relativistic particle in an electromagnetic field

The Hamiltonian of a relativistic particle in an electro-magnetic field, acting on the wave function Ψ, can be written as

$$\hat{H}\Psi = \frac{1}{2m}\left[(i\partial_\mu + eA_\mu)^2 + m^2\right]\Psi \tag{7.147}$$

$$= \frac{1}{2m}\left[\partial_t^2 - \partial_x^2 + e^2(-(A^t)^2 + (A^x)^2) + m^2\right]\Psi \tag{7.148}$$

$$+ \frac{ie}{2m}\left[\partial_t(A^t\Psi) + A^t\partial_t\Psi + \partial_x(A^x\Psi) + A^x\partial_x\Psi\right]. \tag{7.149}$$

Using fourth order finite difference equation for the first and second derivative

$$f_i' \mapsto \frac{-f_{i-2} + 8f_{i-1} - 8f_{i+1} + f_{i+2}}{12\delta}, \tag{7.150}$$

$$f_i'' \mapsto \frac{-f_{i-2} + 16f_{i-1} - 30f_i + 16f_{i+1} - f_{i+2}}{12\delta^2}, \tag{7.151}$$

we can write the Hamiltonian operator acting on Ψ as

$$\begin{aligned}
\hat{H}\Psi = \sum_{i,j}\Bigg[& \frac{1}{24m\delta^2}\Bigg[- \left[1 + ie\delta(A^t_{i,j} + A^t_{i-2,j})\right]\Psi_{i-2,j} \\
& - \left[1 - ie\delta(A^t_{i,j} + A^t_{i+2,j})\right]\Psi_{i+2,j} \\
& + \left[1 - ie\delta(A^x_{i,j} + A^x_{i,j-2})\right]\Psi_{i,j-2} \\
& + \left[1 + ie\delta(A^x_{i,j} + A^x_{i,j+2})\right]\Psi_{i,j+2}\Bigg] \\
& + \frac{1}{3m\delta^2}\Bigg[+ \left[2 + ie\delta(A^t_{i,j} + A^t_{i-1,j})\right]\Psi_{i-1,j} \\
& + \left[2 - ie\delta(A^t_{i,j} + A^t_{i+1,j})\right]\Psi_{i+1,j} \\
& - \left[2 - ie\delta(A^x_{i,j} + A^x_{i,j-1})\right]\Psi_{i,j-1} \\
& - \left[2 + ie\delta(A^x_{i,j} + A^x_{i,j+1})\right]\Psi_{i,j+1}\Bigg] \\
& + \frac{1}{2m}\left[e^2[-(A^t_{i,j})^2 + (A^x_{i,j})^2] + m^2\right]\Psi_{i,j}\Bigg].
\end{aligned} \tag{7.152}$$

The Hamiltonian operator can be written in terms of 9 block diagonal matrices

$$\hat{H} = K_1 + \cdots + K_9\,, \tag{7.153}$$

where K_1, K_2 correspond to the $\Psi_{i\pm2,j}$ components, K_3, K_4 correspond to the $\Psi_{i\pm1,j}$ components, K_5, K_6 correspond to the $\Psi_{i,j\pm2}$ components, K_7, K_8 correspond to the $\Psi_{i,j\pm1}$ components, and A_9 correspond to the diagonal $\Psi_{i,j}$ components, as expressed in equation (7.152).

7.C The von Neumann pointer

We can model the measurement of a general observable $Q[x^\mu]$ by coupling the observable to the momentum of a von Neumann pointer with the canonical variables (X, P), evolving over the interval $[0, T]$, described by the pointer action

$$S_p[X, P] = \int_0^T dt \, \left[-X\dot{P} - \mathcal{H}_p \right] , \tag{7.154}$$

with the pointer Hamiltonian

$$\mathcal{H}_p = \frac{P^2}{2M} + g\, P\, Q[x^\mu] , \tag{7.155}$$

where the observable is coupled to the pointer with coupling strength g.

We will assume the pointer to be initialized in a Gaussian state centred at the initial position $\mu_i = 0$, with initial standard deviation Δ_i and vanishing initial momentum $P_i = 0$, *i.e.*,

$$\psi_{0,p}(X_0) = \frac{1}{\sqrt[4]{2\pi\Delta_i^2}}\, e^{-\frac{X_0^2}{4\Delta_i^2}} . \tag{7.156}$$

In the momentum representation, the wavefunction takes the form

$$\psi_0(P_0) = \sqrt[4]{\frac{2\Delta_i^2}{\pi\hbar^2}}\, e^{-\frac{P_0^2\Delta_i^2}{\hbar^2}} . \tag{7.157}$$

The evolution of the pointer, assuming the particle to evolve from the state $|\psi_0\rangle$ to the state $|\psi_1\rangle$, can be described by the path integral

$$\psi_{1,p}(P_1) = \int_{0^+}^{\infty} \mathrm{d}s \int_{\psi_{0,p}}^{P(T)=P_1} \mathcal{D}X\, \mathcal{D}P \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu \, e^{\frac{i}{\hbar}(S[x^\mu]+S_p[x^\mu,X,P])} . \tag{7.158}$$

We are interested in the asymptotics of the weak limit, for which the coupling $g \to 0$. We expand the integrand in g and concentrate on the linear contribution

$$\psi_{1,p}(P_1) = \int_{\psi_{0,p}}^{P(T)=P_1} \mathcal{D}X\,\mathcal{D}P\, e^{\frac{i}{\hbar}\int_0^T \mathrm{d}t\left[-X\dot{P}-\frac{P^2}{2M}\right]} \int_{0^+}^{\infty} \mathrm{d}s \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu\, e^{\frac{i}{\hbar}S[x^\mu]-\frac{i}{\hbar}g\int_0^T \mathrm{d}tPQ[x^\mu]} \tag{7.159}$$

$$= \int_{\psi_{0,p}}^{P(T)=P_1} \mathcal{D}X\,\mathcal{D}P\, e^{\frac{i}{\hbar}\int_0^T \mathrm{d}t\left[-X\dot{P}-\frac{P^2}{2M}\right]} \tag{7.160}$$

$$\times \int_{0^+}^{\infty} \mathrm{d}s \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu\, e^{\frac{i}{\hbar}S[x^\mu]} \left[1 - \frac{i}{\hbar}g\int_0^T \mathrm{d}tPQ[x^\mu]\right] + \mathcal{O}(g^2)\,. \tag{7.161}$$

We can identify the path integral over the linear term with the weak value Q_w defined as

$$Q_w = \frac{\int_{0^+}^{\infty}\mathrm{d}s\int_{\psi_0}^{\psi_1}\mathcal{D}x^\mu e^{iS[x^\mu]/\hbar}Q[x^\mu]}{\int_{0^+}^{\infty}\mathrm{d}s\int_{\psi_0}^{\psi_1}\mathcal{D}x^\mu e^{iS[x^\mu]/\hbar}} \tag{7.162}$$

and write the wave function for the pointer up to linear order in g as

$$\psi_{1,p}(P_1) = \int_{\psi_{0,p}}^{P(T)=P_1} \mathcal{D}X\,\mathcal{D}P\, e^{\frac{i}{\hbar}\int_0^T \mathrm{d}t\left[-X\dot{P}-\frac{P^2}{2M}\right]} \left[1 - \frac{i}{\hbar}g\int_0^T \mathrm{d}tPQ_w\right] \tag{7.163}$$

$$\times \int_{0^+}^{\infty} \mathrm{d}s \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu\, e^{\frac{i}{\hbar}S[x^\mu]} + \mathcal{O}(g^2) \tag{7.164}$$

$$= \int_{\psi_{0,p}}^{P(T)=P_1} \mathcal{D}X\,\mathcal{D}P\, e^{\frac{i}{\hbar}\int_0^T \mathrm{d}t\left[-X\dot{P}-\frac{P^2}{2M}-gPQ_w\right]} \int_{0^+}^{\infty} \mathrm{d}s \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu\, e^{\frac{i}{\hbar}S[x^\mu]} + \mathcal{O}(g^2)\,. \tag{7.165}$$

Note that this is the reason we define the weak value by equation (7.162). Observe that in the weak limit, $g \to 0$, the pointer decouples from the particle. The influence of the von Neumann pointer on the particle is quadratic in coupling strength g.

We can evaluate the Gaussian path integral over X and P,

$$\int \mathrm{d}P_0 \int_{P(0)=P_0}^{P(T)=P_1} \mathcal{D}X\,\mathcal{D}P\, e^{\frac{i}{\hbar}\int_0^T \mathrm{d}t\left[-X\dot{P}-\frac{P^2}{2M}-gPQ_w\right]}\psi_{0,p}(P_0) \tag{7.166}$$

by solving the equations of motion with the appropriate boundary condition imposed by the initial wave function

$$\dot{P} = 0\,, \qquad P(1) = P_1\,, \tag{7.167}$$

$$\dot{X} = \frac{P}{M} + g\,Q_w\,, \qquad iX_0 - 2P_0\frac{\Delta_i^2}{\hbar} = 0\,. \tag{7.168}$$

The conditions on the momentum imply the solution

$$P(t) = P_1\,. \tag{7.169}$$

$$X(t) = X_0 + \left[\frac{P_1}{M} + g\,Q_w\right] t\,. \tag{7.170}$$

Substituting the solution in the action, we obtain the wavefunction of the pointer, to leading order in g,

$$\psi_{1,p}(P_1) = \sqrt[4]{\frac{2\Delta_i^2}{\pi\hbar^2}}\, e^{-\frac{P_1^2\Delta_i^2}{\hbar^2} - \frac{iT}{\hbar}\left[\frac{P_1^2}{2M} + gP_1Q_w\right]} \int_{0^+}^{\infty} \mathrm{d}s \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu\, e^{\frac{i}{\hbar}S[x^\mu]} + \mathcal{O}(g^2)\,. \tag{7.171}$$

The pointer is Gaussian distributed with the mean and standard deviation

$$\langle P_1\rangle = \frac{Tg\hbar}{2\Delta_i^2}\,\mathrm{Im}[Q_w], \tag{7.172}$$

$$\sqrt{\langle P_1^2\rangle - \langle P_1\rangle^2} = \frac{\hbar}{2\Delta_i}. \tag{7.173}$$

Observe that the imaginary part of the weak value, Q_w, induces a shift in the mean of the momentum of the pointer.

In the position representation, the pointer is distributed like the Gaussian

$$\psi_1(X_1) = \sqrt[4]{\frac{2M^2\Delta_i^2}{\pi(T^2\hbar^2 + 4M^2\Delta_i^4)}}\, e^{-\frac{(X_1 - gTQ_w)^2}{4\Delta_i^2 + i\,2\hbar T/M}} \int_{0^+}^{\infty} \mathrm{d}s \int_{\psi_0}^{\psi_1} \mathcal{D}x^\mu\, e^{\frac{i}{\hbar}S[x^\mu]} + \mathcal{O}(g^2)\,, \tag{7.174}$$

with the mean and standard deviation

$$\langle X_1\rangle = gT\,\mathrm{Re}[Q_w] + \frac{gT^2\hbar}{2M\Delta_i^2}\,\mathrm{Im}[Q_w], \tag{7.175}$$

$$\sqrt{\langle X_1^2\rangle - \langle X_1\rangle^2} = \sqrt{\Delta_i^2 + \frac{T^2\hbar^2}{4M^2\Delta_i^2}}\,. \tag{7.176}$$

We thus observe that, up to linear order in the coupling g, the mean momentum of the von Neumann pointer is shifted by a term proportional to the imaginary part of the weak measurement

$$\langle P_1\rangle = \langle P_0\rangle + \frac{Tg\hbar}{2\Delta_i^2}\,\mathrm{Im}[Q_w]\,, \tag{7.177}$$

while the position of the von Neumann pointer shifted by a term involving both the real and imaginary parts of the weak value,

$$\langle X_1 \rangle = \langle X_0 \rangle + gT \,\mathrm{Re}[Q_w] + \frac{gT^2\hbar}{2M\Delta_i^2}\,\mathrm{Im}[Q_w]\,. \tag{7.178}$$

We thus observe that we can construct an idealized measurement for the real part of the weak value by considering a massive von Neumann pointer, $M \to \infty$, with a vanishing spread $\Delta_i \to 0$, for which $M\Delta_i^2 \to \infty$. The imaginary part of the weak value can be measured with an idealized von Neumann pointer with a momentum eigenstate, in the limit $\Delta_i \to \infty$.

Chapter 8

Lorentzian path integrals

> Euclidean techniques will be avoided. Historically, most problems in quantum mechanics that can be illuminated by analytic continuation were first solved by other more straightforward methods. The advantage of proceeding in a direct manner is that the physics remains in the foreground and our understanding is enhanced.
>
> Arlen Anderson and Bryce DeWitt

From about the beginning of the twentieth century, experimental physics unveiled an impressive array of phenomena – such as the photo-electric effect and the double slit experiment – which cannot be explained with classical physics. Theoretical investigations on the structure of these phenomena led to the insight that light and particles can exhibit both particle- and wave-like behavior. This apparent inconsistency was resolved by Werner Heisenberg (1925) [197] and Erwin Schrödinger (1926) [293]. Werner Heisenberg developed a picture of quantum mechanics in which the dynamics of the system is expressed in terms of the evolution of observables. In the subsequent year, Erwin Schrödinger presented an independent picture in which the dynamics of the system is expressed in terms of an evolving wavefunction. In the years following Heisenberg's and Schrödinger's proposals both formulations were demonstrated to accurately match experiments. Even though the two formations were developed from different philosophical points of view, they were soon proven to be mathematically equivalent (Marshall Stone and John von Neumann (1931) [304, 260, 305, 318]).

In 1948, following the work of Paul Dirac, Richard Feynman developed yet another radically different formulation in terms of the Feynman path integral [146]. Instead of

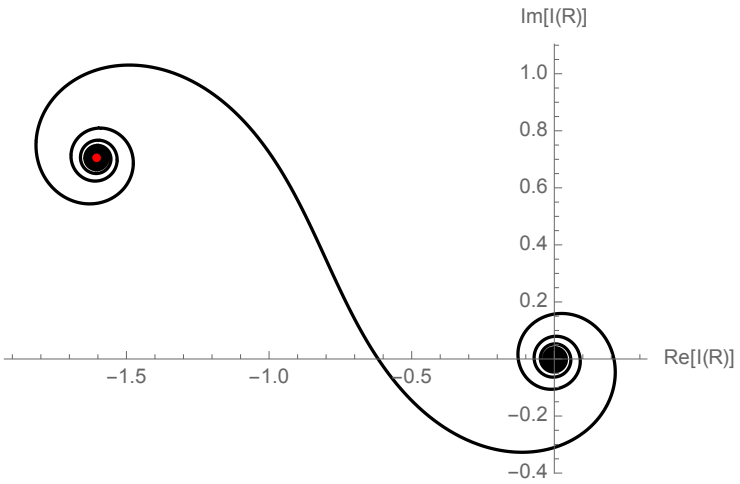


Figure 8.1: The real and imaginary parts of the integral $I(R)$ as a function of R (black curve starting at the origin $I(0) = 0$). The red dot is the limit $\lim_{R\to\infty} I(R) = (-1)^{1/4} e^{2i}\sqrt{\pi}$.

describing the amplitudes to evolve from state A to B in terms of operators and partial differential equations, Feynman proposed to evaluate the amplitude as the sum over the possible 'histories' between the states A and B. Even though this path integral formulation is nowadays frequently used in many branches of theoretical physics, a mathematically rigorous formulation remains elusive. It is often claimed that the measure only exists in imaginary time, known as the Wick rotation. This is based on the observation that after the Wick rotation, the resulting integral can be related to the theory of Brownian motion, known as the Feynman-Kac formula. However, it is not always clear whether and when the Wick rotation is a legal procedure. Moreover, there exist physical systems for which the Wick-rotated integral diverges.

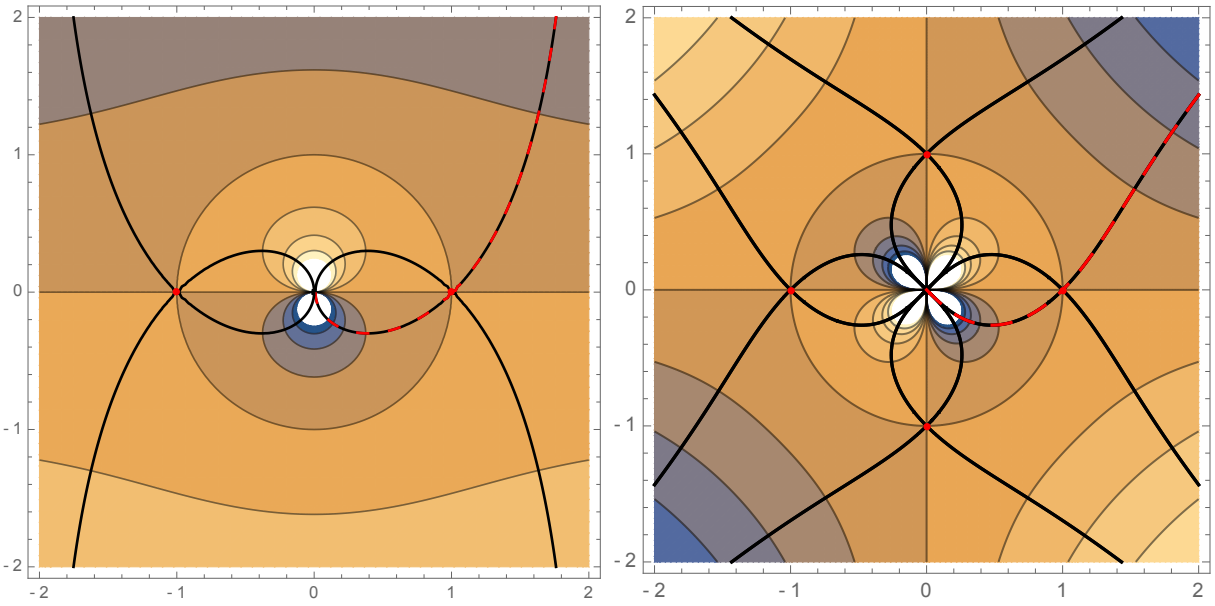
To see this, consider the conditionally convergent integral I, defined in the limit

$$I = \lim_{R\to\infty} I(R)\,, \tag{8.1}$$

of the partial integral

$$I(R) = \int_0^R e^{i\left(x^{-1}+x\right)} \frac{\mathrm{d}x}{\sqrt{x}}\,. \tag{8.2}$$

The ordinary integral appears in both the path integral formulation of the free relativistic



(a) Picard-Lefschetz diagram for $e^{i(x^{-1}+x)}$. (b) Picard-Lefschetz diagram for $e^{i(z^{-2}+z^2)}$.

Figure 8.2: The Lefschetz thimble (red dashed line) in complex x/z-plane. The saddle points (in red) with the corresponding steepest ascent and descent contours (in black). The background is the real part of the exponent.

particle and quantum de Sitter spacetime[1]. The limit,

$$\lim_{R\to\infty} I(R) = (-1)^{1/4} e^{2i} \sqrt{\pi}\,, \tag{8.3}$$

is well-defined (see Figure 8.1 for the limit in the complex plane). However, the Wick-rotated integral, $x \to \pm ix$, diverges for both signs (see Figure 8.2a). The integral along the contour $(0, i\infty)$ diverges near $x = 0$ whereas the integral along the contour $(0, -i\infty)$ diverges near $x = -i\infty$. In general relativity and condensed matter physics this phenomenon is respectively known as the conformal factor [160, 196] and the sign problem [313].

Picard-Lefschetz theory enables us to transform the oscillatory integral into a convex one (see chapter 3). Along the Lefschetz thimble of the exponent[2], given by

[1]For the relativistic particle the variable x represents the Schwinger time. In quantum cosmology it represents the integral over the lapse.

[2]We can alternatively remove the pre-factor $x^{-1/2}$, with a change of variables $z = \sqrt{x}$, transforming the integral into $I = 2\int_0^\infty e^{i(z^{-2}+z^2)}\mathrm{d}z$. See Figure 8.2b for the corresponding Picard-Lefschetz diagram.

$$\mathcal{J} = \left\{ (\lambda+1) + i\lambda\sqrt{\frac{1+\lambda}{1-\lambda}} \;\middle|\; \lambda \in (-1,1) \right\} \subset \mathbb{C}\,, \tag{8.4}$$

the integrand (ignoring the pre-factor) is non-oscillatory and absolutely convergent, *i.e.*,

$$e^{i(x^{-1}+x)}|_{\mathcal{J}} = \mathrm{Exp}\left[2i - \frac{2\lambda^2}{\sqrt{1-\lambda^2}}\right]\,. \tag{8.5}$$

while the Jacobian

$$\frac{\mathrm{d}x}{\mathrm{d}\lambda}\frac{1}{\sqrt{x}} \tag{8.6}$$

is polyonmial in λ. The integrand approaches a Gaussian near the saddle point, $\lambda = 0$, and vanishes at the boundaries, $\lambda = \pm 1$. The integration contour, on the other hand is a complex curve starting at $x = 0$ and moving to complex infinity $i\infty$. Note that this is an exact result, since the integral over arc from ∞ to $i\infty$ vanish, *i.e.*, we parametrize the arc by $x(\theta) = re^{i\theta}$ for $r > 0$, the integral along the arc is dominated by

$$\left|\int_{\mathrm{arc}} e^{i\left(x^{-1}+x\right)}\frac{\mathrm{d}x}{\sqrt{x}}\right| = r^{1/2}\left|i\int_0^{\pi/2} e^{i\left[re^{i\theta}+r^{-1}e^{-i\theta}+\theta/2\right]}\mathrm{d}\theta\right| \tag{8.7}$$

$$\leq r^{1/2}\int_0^{\pi/2} e^{-(r-r^{-1})\sin\theta}\mathrm{d}\theta \tag{8.8}$$

$$< r^{1/2}\int_0^{\pi/2} e^{-\frac{1}{2}(r-r^{-1})\theta}\mathrm{d}\theta \tag{8.9}$$

$$= \frac{2r^{3/2}}{r^2-1}\left[1 - e^{-\frac{\pi}{4}\left(r-\frac{1}{r}\right)}\right]\,, \tag{8.10}$$

which approaches 0 as $r \to \infty$. This generally true for polynomial exponents of conditionally convergent integrals.

In this chapter, we apply functional Picard-Lefschetz theory to the real-time (Lorentzian) path integral. We first describe the Feynman-Kac formula and summarize work on continuous-time regularization schemes for the real-time path integral. We subsequently propose a new definition of the real-time path integral on the Lefschetz thimble in complex function space. We construct a σ-measure for the path integral for a range of simple models, including the free particle, the (inverted) harmonic oscillator, and the relativistic particle. We extend these results to real-time perturbation theory and discuss when the naive Euclidean treatment on the thimble fails. Finally, we conjecture how these results can be extended beyond the perturbative regime. We, in particular, find a split between high and the lower frequencies. It seems that the Feynman propagator of many quantum mechanical phenomena can be studied by means of finite dimensional integrals.

8.1 Feynman's path integral

In the Schrödinger picture of quantum mechanics the wavefunction ψ of a particle with mass m in a potential V, evolves according to the Schrödinger equation

$$i\hbar\frac{\partial\psi}{\partial t} = \hat{H}\psi \tag{8.11}$$

$$= \left[-\frac{\hbar^2}{2m}\nabla^2 + V(x)\right]\psi\,, \tag{8.12}$$

with the Hamiltonian operator $\hat{H}$. Given the initial wavefunction $\psi(x,t_0) = \psi_0(x)$, the wavefunction is propagated forward, to time $t = t_1 > t_0$, by a convolution with the propagator [150], *i.e.*,

$$\psi(x_1,t_1) = \int G[x_1,t_1;x_0,t_0]\psi_0(x_0)\mathrm{d}x_0\,. \tag{8.13}$$

The propagator $G[x_1,t_1;x_0,t_0]$ is the amplitude to evolve from the point (x_0,t_0) to the point (x_1,t_1), formally written as

$$G[x_1,t_1;x_0,t_0] = \Theta(T)\langle x_1|e^{-\frac{i}{\hbar}HT}|x_0\rangle\,, \tag{8.14}$$

with the interval $T = t_1 - t_0$, and the Heaviside step function Θ. Feynman noticed that this propagator can be interpreted as the sum over the amplitudes corresponding to the possible paths between (x_0,t_0) and (x_1,t_1), *i.e.*,

$$G[x_1,t_1;x_0,t_0] = \sum_{(\text{paths})} \mathcal{A}(\text{particle moves form } (x_0,t_0) \text{ to } (x_1,t_1) \text{ following a particular path})\,. \tag{8.15}$$

This is a very fruitful idea, but has unfortunately been tricky to formalize [150].

8.1.1 Lattice regularization

Feynman, in his initial definition of the path integral, followed the definition of the Riemann integral and proposed a lattice regularization [150]. Dividing the interval T into N segments each lasting $\delta t = T/N$, we can write the propagator as

$$G[x_1,t_1;x_0,t_0] = \langle x_1|e^{-\frac{i}{\hbar}H\delta t}e^{-\frac{i}{\hbar}H\delta t}\ldots e^{-\frac{i}{\hbar}H\delta t}|x_0\rangle\,. \tag{8.16}$$

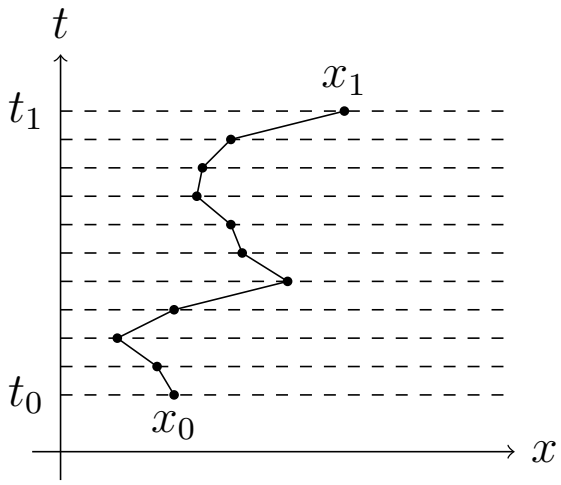


Figure 8.3: Lattice regularization of the path integral.

Using the fact that the position states $|x\rangle$ forms a complete set of states, *i.e.*,

$$\int \mathrm{d}x|x\rangle\langle x| = 1\,, \tag{8.17}$$

the propagator becomes an $(N-1)$-dimensional integral

$$G[x_1,t_1;x_0,t_0] = \left(\prod_{j=1}^{N-1}\int \mathrm{d}q_j\right)\langle x_1|e^{-\frac{i}{\hbar}H\delta t}|q_{N-1}\rangle\langle q_{N-1}|e^{-\frac{i}{\hbar}H\delta t}|q_{N-2}\rangle\ldots\langle q_1|e^{-\frac{i}{\hbar}H\delta t}|x_0\rangle\,, \tag{8.18}$$

where q_j represents the position of the particle at time $t = t_0 + j\delta t$. The history of the particle is approximated by a series of positions $(q_0, q_1, q_2, \ldots, q_N)$, with $q_0 = x_0$ and $q_1 = x_1$ (see Figure 8.3 for a schematic representation). In the continuum limit $\delta t \to 0$, we hope to recover the set of continuous paths interpolating between the points x_0 and x_1.

The transition amplitude $\langle q_{j+1}|e^{-\frac{i}{\hbar}H\delta t}|q_j\rangle$, for small δt, determines the nature of integral. For small δt, Trotter's formula enables us to evaluate this amplitude. Given the Hamiltonian operator

$$\hat{H} = \frac{\hat{p}^2}{2m} + V(\hat{x})\,, \tag{8.19}$$

we can use a complete set of momentum states to evaluate the transition amplitude

$$\langle q_{j+1}|e^{-\frac{i}{\hbar}\hat{H}\delta t}|q_j\rangle = \int \mathrm{d}p\, \langle q_{j+1}|e^{-\frac{i}{\hbar}\frac{\hat{p}^2}{2m}\delta t}|p\rangle\langle p|e^{-\frac{i}{\hbar}V(\hat{x})\delta t}|q_j\rangle \tag{8.20}$$

$$= \int \mathrm{d}p\, e^{-\frac{i}{\hbar}\frac{p^2}{2m}\delta t - \frac{i}{\hbar}V(q_j)\delta t}\langle q_{j+1}|p\rangle\langle p|q_j\rangle \tag{8.21}$$

$$= \int \mathrm{d}p\, e^{-\frac{i}{\hbar}\frac{p^2}{2m}\delta t - \frac{i}{\hbar}V(q_j)\delta t} e^{ip(q_{j+1}-q_j)} \tag{8.22}$$

$$= \sqrt{\frac{-2\pi m}{\delta t}} e^{i\left[\frac{m}{2}\frac{(q_{j+1}-q_j)^2}{\delta t^2} - V(q_j)\right]\delta t}, \tag{8.23}$$

where we use the inner product $\langle q|p\rangle = e^{ipq}$. Substituting the transition amplitude in the propagator, we obtain a representation of the propagator in terms of the position of the particle at times $t_j = t_0 + j\delta t$, *i.e.*,

$$G[x_1, t_1; x_0, t_0] = \left(\frac{-2\pi m}{\delta t}\right)^{\frac{N}{2}} \int \prod_{j=1}^{N-1} \mathrm{d}q_j\, e^{\frac{i}{\hbar}\sum_{j=0}^{N-1}\left[\frac{m}{2}\frac{(q_{j+1}-q_j)^2}{\delta t^2} - V(q_j)\right]\delta t}. \tag{8.24}$$

At this point, it is tempting to take the continuum limit $\delta t \to 0$ and write the propagator as a path integral

$$G[x_1, t_1; x_0, t_0] = \int_{x(0)=x_0}^{x(1)=x_1} \mathcal{D}x(t)\, e^{\frac{i}{\hbar}\int\left[\frac{m}{2}\dot{x}^2 - V(x)\right]\mathrm{d}t}, \tag{8.25}$$

with the infinite product of Lebesgue measures

$$\int_{x(0)=x_0}^{x(1)=x_1} \mathcal{D}x(t) = \lim_{N\to\infty}\left(\frac{-2\pi m}{\delta t}\right)^{\frac{N}{2}} \int \prod_{j=1}^{N-1} \mathrm{d}q_j\,. \tag{8.26}$$

Note that while the definition of the propagator is written in terms of the Hamiltonian operator $\mathcal{H}$, the 'path integral' is expressed in terms of the action of the particle.

The integral formulation of quantum mechanics has led to significant insights into quantum mechanics. It can be argued that it is the most elegant definition of quantum theory. It moreover plays a central role in our current understanding of quantum field theory, where one integrates over the possible field configurations, *i.e.*,

$$\int \mathcal{D}\varphi\, e^{\frac{i}{\hbar}\int \mathrm{d}^4x\left[\frac{1}{2}(\partial\varphi)^2 - V(\varphi)\right]}. \tag{8.27}$$

However, the lattice regularisation of the Lorentzian path integral is not mathematically satisfactory for a number of reasons [223]:

- The propagator expressed as the N-dimensional integral (8.24), is a conditionally convergent integral since the integrand is a pure phase and the Lebesgue measure of integration domain $\mathbb{R}^N$ diverges, *i.e.*,

$$\int \prod_{j=0}^{N-1} \mathrm{d}q_j = \infty \,. \tag{8.28}$$

  As a consequence, Fubini's theorem does not apply and the integral can depend on the order of integration or the regularization scheme[3]. See appendix 8.B.

- In the limit $N \to \infty$, the order of the limit and the integration is effectively exchanged (see equation (8.26)). Since the dominated convergence theorem does not apply to conditionally convergent integrals, the limit of the integral is not necessarily equal to the integral of the limit. See appendix 8.B.

These problems are often avoided with the $i\epsilon$-procedure, by letting $m \to m - i\epsilon$ for small positive ϵ, making the integral absolutely convergent. Alternatively, it is common to Wick rotate to imaginary time $t \to -it$. It should be noted that both procedures are of an ac hoc nature. The Wick rotation moreover, does not work for all physically motivated propagators. However, even if the integral is made absolutely convergent, the lattice regularization scheme remains unsatisfactory:

- The pre-factor in equation (8.26) is ill-defined, since the limit

$$\lim_{N\to\infty} \left(\frac{-2\pi m}{\delta t} \right)^{\frac{N}{2}} = \lim_{N\to\infty} \left(\frac{-2\pi m N}{T} \right)^{\frac{N}{2}} \tag{8.29}$$

  diverges for finite m and T.

- Finally and most importantly, the infinite product of Lebesgue measures $\mathcal{D}x$ is not a proper σ-measure, and, as a consequence, the path integral not a proper integral over paths (see appendices 8.A and 8.B). This is an important point since the philosophy of the calculation is based on the idea that the propagator can be written as a functional integral.

[3]We assume that the integral is convergent, since the definition of the propagator in terms of the Hamiltonian operator (8.14) is finite for non-zero T.

To see this, note that both the Lebesgue measure μ_n on $\mathbb{R}^n$ (with $\mathrm{d}\mu_n = \prod_{j=1}^n \mathrm{d}q_i$) and the limit $\mu_\infty = \lim_{n\to\infty} \mu_n$ on $\mathbb{R}^\infty$ (with $\mathrm{d}\mu_\infty = \prod_{j=1}^\infty \mathrm{d}q_j$) are positive translational invariant functions. It is trivial to see that the measure μ_n of the unit hypercube in $[0,1]^n \subset \mathbb{R}^n$ is one, *i.e.*,

$$\mu_n([0,1]^n) = \int_{[0,1]^n} \prod_{j=1}^n \mathrm{d}q_j = 1\,. \tag{8.30}$$

We can subdivide this unit hypercube into 2^n mutually disjoint smaller cubes with width $1/2$ and measure

$$\mu_n([0,1/2]^n) = \int_{[0,1/2]^n} \prod_{j=1}^n \mathrm{d}q_j = (1/2)^n\ . \tag{8.31}$$

The fact that the Lebesgue measure μ_n is a translationally invariant σ-measure ensures that the measure of the hypercube equals the sum of measure of the smaller cubes

$$\mu_n([0,1]^n) = 2^n \mu_n([0,1/2]^n) = 1\,. \tag{8.32}$$

This is no longer true for the infinite dimensional map μ_∞. The 'measure' of the infinite-dimensional hypercube is still evaluates to one, *i.e.*,

$$\mu_\infty([0,1]^\infty) = \lim_{n\to\infty} \mu_n([0,1]^n) = 1\,. \tag{8.33}$$

However, the 'measure' of the smaller cubes vanishes, *i.e.*,

$$\mu_\infty([0,1/2]^\infty) = \lim_{n\to\infty} \mu_n([0,1/2]^n) = 0\,. \tag{8.34}$$

This leads to an inconsistency, since the union of the smaller hypercube still equates to the larger hypercube, while the sum over the measures of the smaller hypercubes does no longer as it vanishes. More generally, any subset of the hypercube $[0,1]^\infty$ has vanishing measure μ_∞. This reasoning does not apply to the Wiener measure as the measure is not invariant under translations.

Richard Feynman was aware of these problems, as can be found in Feynman and Hibbs [150]:

> There are many ways to define a subset of all the paths between A and B. The particular definition we have used here may not be the best for some mathematical purposes.[...]. However, in the few such examples with which we have had experience the [lattice regulation] has been adequate. There may be other cases where [...] the present definition of a sum over all paths is just too awkward to use.
>
> A similar situation arises in ordinary integration, where sometimes the Riemann definition, is not adequate and recourse must be had to some other definition, such as that of Lebesgue. The need to redefine the method of integration does not destroy the concept of integration. So we feel that the possible awkwardness of the special definition of the sum over all paths may eventually require new definitions to be formulated. Nevertheless, the concept of the sum over all paths, like the concept of an ordinary integral, is independent of a special definition and valid in spite of the failure of such definitions.

Note that the infinite product of the Lebesgue measures is not a σ-measure since the product is translational invariant. This is inherent to the lattice regularization. We can only be avoided these problems by considering different regularization schemes. For the Euclidean path integral, this can be done by identifying the kinetic term and the infinite dimensional product of Lebesgue measures as a Wiener measure. In the Lorentzian case, such an identification is more complicated as the integral does not converge absolutely.

8.1.2 Continuous-time regularization

Since the inception of the path integral, there have been several attempts to develop a mathematically rigorous definition of the propagator as an integral over paths. One specific line of research has been aimed at relating the propagator to the Wiener measure of Brownian motion, as the Wiener measure is countably additive and leads to a well-defined functional integral over paths. These attempts are known as *continuous-time regularization schemes*. For an overview see [222].

The Feynman-Kac formula

Mark Kac proposed the first continuous-time regularization scheme. After the Wick rotation $t \to -it$, the infinite product of Lebesgue measures combined with the kinetic term of the action can be reinterpreted as the Wiener measure $\mu^{\nu}_{x_1,x_0}$. For sufficiently well behaved

potentials V, the Wick-rotated path integral is formalized by the Feynman-Kac formula (1951) [218]:

Theorem 1 (The Feynman-Kac formula) *Assuming the potential $V : \mathbb{R}^d \to \mathbb{R}$ is bounded from below and can be expressed as a sum of functions in $L^2(\mathbb{R}^d)$, we can define the Euclidean path integral in terms of the Wiener measure*

$$\int_{x(0)=x_0}^{x(T)=x_1} e^{-\frac{1}{\hbar}\int\left[\frac{m}{2}\dot{x}^2+V(x(t))\right]\mathrm{d}t}\mathcal{D}x(t) = \int_{C_{x_1,x_0}([0,T];\mathbb{R}^n)} e^{-\int_0^T V(x(t))\mathrm{d}t}\,\mathrm{d}\mu^{\hbar/m}_{x_1,x_0}(x)\,, \tag{8.35}$$

with the Wiener σ-measure $\mu^{\nu}_{x_1,x_0}(x)$ defined on the Borel σ-algebra of the continuous paths interpolating between the points x_0 and x_1, written as $C_{x_1,x_0}([0,T];\mathbb{R}^d)$ [248, 299, 177, 223, 8, 166].

The Feynman-Kac formula is a proper functional integral. The Wiener measure $\mu^{\nu}_{x_1,x_0}$ is not defined as the infinite product of Lebesgue measures. It is instead defined in terms of the characteristic functional of Brownian motion, using the Bochner-Minlos theorem (see appendices 8.C and 8.D). The kinetic term indicates that the Wiener measure is not translation invariant.

The Heisenberg uncertainty relation:
The Feynman-Kac formula leads to a paradox, which unveils a central property of the Wiener measure in relation to quantum mechanics. Canonical quantum mechanics is described in terms of operators such as the position $\hat{x}$ and the momentum operator $\hat{p}$. When two operators fail to commute, *i.e.*, $[\hat{x},\hat{p}] = i\hbar$, the two observables cannot be simultaneously probed with arbitrary accuracy. For the observables x and p, this is expressed by the Heisenberg uncertainty relation

$$\Delta x \Delta p \geq \frac{\hbar}{2}\,, \tag{8.36}$$

with Δx and Δp the uncertainty in measurements of x and p. This is an intrinsic property of quantum mechanics. The Feynman-Kac formulation, on the other hand, is described in terms of continuous paths $x(t) \in C_{x_1,x_0}([0,1],\mathbb{R})$. As a consequence, it seems like the expectation value of products of observables in the path integral formulation is independent of orderings. How can the two formulations be equivalent?

The resolution lays in the fact that Brownian motion, though almost everywhere continuous, is almost nowhere differentiable, *i.e.*, the corresponding Wiener measure vanishes

for differentiable paths. Lets consider a free particle, $V = 0$, of unit mass, $m = 1$, and the times $0 < t_- < t < t_+ < T$. The expectation value of the two orderings $\hat{x}\hat{p}$ and $\hat{p}\hat{x}$ correspond to the path integrals

$$\int_{x(0)=x_0}^{x(T)=x_1} x(t)\dot{x}(t_-)\mathrm{d}\mu^{\hbar}_{x_1,x_0} = \langle x_1|e^{-\hat{H}(T-t)/\hbar}\hat{x}e^{-\hat{H}(t-t_-)/\hbar}\hat{p}e^{-\hat{H}t_-/\hbar}|x_0\rangle\,, \tag{8.37}$$

$$\int_{x(0)=x_0}^{x(T)=x_1} x(t)\dot{x}(t_+)\mathrm{d}\mu^{\hbar}_{x_1,x_0} = \langle x_1|e^{-\hat{H}(T-t_+)/\hbar}\hat{p}e^{-\hat{H}(t_+-t)/\hbar}\hat{x}e^{-\hat{H}t/\hbar}|x_0\rangle\,. \tag{8.38}$$

In the limit $t_+ \to t$ from above and $t_- \to t$ from below, the difference in the canonical formulation exhibits the commutation relation of $\hat{x}$ and $\hat{p}$, *i.e.*,

$$\langle x_1|e^{-\hat{H}(T-t)/\hbar}[\hat{x},\hat{p}]e^{-\hat{H}t/\hbar}|x_0\rangle = \hbar\langle x_1|e^{-\hat{H}T/\hbar}|x_0\rangle\,. \tag{8.39}$$

On the path integral side, we realize that the time derivative $\dot{x}$ is ill-defined since the Wiener measure has only support for non-differentiable paths. When substituting the finite difference approximation

$$\dot{x} = \lim_{t_+\to t}\frac{x(t_+)-x(t)}{t_+-t} \tag{8.40}$$

$$= \lim_{t_-\to t}\frac{x(t)-x(t_-)}{t-t_-}\,, \tag{8.41}$$

we recover the commutation relation

$$\lim_{\Delta t\to 0}\int\left[x(t)\frac{x(t)-x(t-\Delta t)}{\Delta t} - x(t)\frac{x(t+\Delta t)-x(t)}{\Delta t}\right]\mathrm{d}\mu^{\hbar}_{x_1,x_0}(x) \tag{8.42}$$

$$= \hbar\int \mathrm{d}\mu^{\hbar}_{x_1,x_0}(x) \tag{8.43}$$

after some evaluation. This demonstrates that the non-differentiable nature of Brownian motion is at the core of quantum mechanics. We expect that a proper definition of the real-time path integral will share this characteristic behaviour. For a physically motivated derivation of this result see Feynman and Hibbs chapter 7-5 [150].

Quantum fields:
The Feynman-Kac formula provides a mathematically rigorous definition for path integrals in Euclidean quantum mechanics. The measure of the functional integral has support in

the space of continuous non-differentiable paths. This can be extended to the path integral of Euclidean quantum field theory

$$\int \mathcal{D}\phi \, e^{-S[\phi]} \,. \tag{8.44}$$

However, it can be shown that the resulting measure does not have support in the space of continuous fields, interpolating between two configurations. The measure has instead support in the space of distributions:

Theorem 2 *For all $d \geq 1$, there exists a Gaussian measure on the space $\mathcal{F}^d$ of fields on $\mathbb{R}^d$ analogous to the Wiener measure for the Feynman-Kac formula. However, for $d \geq 2$, this measure is not supported on any space of ordinary functions, but rather on a space of distributions [166, 177].*

The implementation of the Euclidean path integral following the Feynman-Kac method is thus in tension with the original philosophy of the path integral formalism. The multi-dimensional Wiener measure is thus not satisfactory in Euclidean quantum field theory.

The attempt by Israel Gel'fand and Akiva Yaglom

The main obstacle to a continuous-time regularization of the real-time path integral is the fact that the integral (8.24), for finite N, is highly oscillatory and only conditionally convergent. That is to say, the integrand is a pure phase and the integral over the magnitude of the integrand diverges. In 1956, the mathematicians Israel Gel'fand and Akiva Yaglom attempted to solve this problem by adding an artificial diffusion ν [157]. They considered the integral

$$\left(\frac{-2\pi m}{\delta t}\right)^{\frac{N}{2}} \prod_{j=1}^{N-1} \int \mathrm{d}q_j e^{\frac{i}{\hbar}\sum_{j=0}^{N-1}\left[\frac{m}{2}\frac{(q_{j+1}-q_j)^2}{\delta t^2}-V(q_j)\right]\delta t - \frac{1}{2\nu}\sum_{j=0}^{N-1}\frac{(q_{j+1}-q_j)^2}{\delta t}} \,, \tag{8.45}$$

in which the the term $-\frac{1}{2\nu}\sum_{j=0}^{N-1}\frac{(q_{j+1}-q_j)^2}{\delta t}$ makes the integral absolutely convergent. Since Brownian motion is almost nowhere differentiable, they combined the kinetic term and the regulator into a complex mass $m' = \frac{\hbar}{\nu} - im$. In analogy with the Feynman-Kac formula, they identified the limit $N \to \infty$ with the Wiener measure

$$\int_{C_{x_1,x_0}([0,T];\mathbb{R}^n)} e^{-\int_0^T V(x(t))\mathrm{d}t} \, \mathrm{d}\mu^{\hbar/m'}_{x_1,x_0}(x) \,, \tag{8.46}$$

obtaining a one-parameter family of well-defined functional integrals. As a final step, they removed the regulator in the limit $\nu \to \infty$. The hope was that the complex Wiener measure $\mu^{\hbar/m'}_{x_1,x_0}$ converged to a proper σ-measure for the Feynman path integral, as $\nu \to \infty$. However, as shown by Robert Cameron (1960) [69], the complex Wiener measure is only countably additive for $\mathrm{Re}[m'] > 0$. Since $\mu^{\hbar/m'}_{x_1,x_0}$ is not a σ-measure in the required limit, Cameron showed that the Gel'fand-Yaglom scheme fails to define the real-time Feynman path integral.

The Itô scheme

Following the work by Robert Cameron, Kiyosi Itô attempted to fix the Gel'fand and Yaglom scheme by adding yet another term in the action [213]. Whereas Gel'fand and Yaglom considered the regulator $-\frac{1}{2\nu}\int \dot{x}^2 \mathrm{d}t$, Itô proposed the regulator $-\frac{1}{2\nu}\int [\ddot{x}^2 + \dot{x}^2]\,\mathrm{d}t$, *i.e.*,

$$\int \mathcal{D}x\, e^{\frac{i}{\hbar}S[x] - \frac{1}{2\nu}\int_0^T (\ddot{x}^2 + \dot{x}^2)\mathrm{d}t}\,. \tag{8.47}$$

This is a working continuous-time regularization scheme, as it leads to a σ-measure over paths in the limit $\nu \to \infty$. This way, Itô managed to recover of the Feynman propagator for a non-relativistic particle in a constant, a linear and a quadratic potential. However, it is not clear whether this can be considered as a completely satisfactory solution. The second order derivative $\ddot{x}$ in the regulator, changes the support of the σ-measure to once differentiable paths. It is thus not clear whether the integral reproduces the Heisenberg uncertainty relation between the position $\hat{x}$ and momentum operator $\hat{p}$. The propagator for finite ν moreover does not satisfy a Schrödinger-like equation as it includes second order derivatives with respect to time. The measure should, for this reason, only be considered in the limit $\nu \to \infty$ [223]. The constant, linear and quadratic potentials can, moreover, be completely studied in terms of classical solutions (for the path integral is Gaussian).

The coherent state scheme

More recently Ingrid Daubechies and John Klauder (1985), inspired by Itô's regularization scheme, proposed a continuous-time regularization scheme for the phase-space formulation of Feynman path integral [88]. The analysis is comparable to Itô's proposal in that they considered the regulator $-\frac{1}{2\nu}\int [\dot{p}^2 + \dot{x}^2]\,\mathrm{d}t$, with $p = m\dot{x}$ the canonical momentum ,

$$\int \mathcal{D}x\mathcal{D}p\, e^{\frac{i}{\hbar}\int_0^T (\dot{q}p - H[x;p])\mathrm{d}t - \frac{1}{2\nu}\int_0^T (\dot{x}^2 + \dot{p}^2)\mathrm{d}t} \tag{8.48}$$

For finite ν, the corresponding σ-measure has support for continuous position and momentum paths. It is thus not clear how this scheme relates to the Heisenberg uncertainty relation. The implications of this proposal are still being explored [222, 223].

8.2 Functional Picard-Lefschetz theory

The analysis in the previous section demonstrates that the original lattice regularization scheme does not lead to a satisfactory definition of the Feynman path integral, *i.e.*, the resulting expression is not a well-defined integral. In the subsequent years, several attempts were made to deform the integral in the complex plane and relate the result to the Wiener measure. Notwithstanding progress in continuous-time regularization schemes, the physics community has not reached a fully satisfactory definition of the real-time path integral. In practice, it is often claimed that the real-time path integral is ill-defined and that one needs to rely on the Euclidean theory. We here use functional Picard-Lefschetz theory to propose a new definition of the real-time path integral. For a similar analysis of the Picard-Lefschetz structure of a view non-relativistic toy-models see [306].

Picard-Lefschetz theory is designed to express oscillatory integrals of the form

$$\int_{\mathbb{R}} e^{if(x)} \mathrm{d}x \tag{8.49}$$

as a sum of non-oscillatory absolutely convergent integrals. We analytically continue the integrand into the complex plane and deform the integration contour to the steepest descent contours of the relevant saddle points. For an exposition of the theory applied to one-dimensional integral see chapter 3. We here consider the functional extension.

Consider the propagator of a quantum mechanical system, written as the 'Lorentzian path integral'

$$G[x_1, x_0, 1] = \int_{x(0)=x_0}^{x(1)=x_1} e^{iS[x]} \mathcal{D}x \tag{8.50}$$

with the action $S : C_{x_1,x_0}([0,1],\mathbb{R}) \to \mathbb{R}$ and $C_{x_1,x_0}([0,1],\mathbb{R})$ the manifold of continuous real-valued trajectories $x : [0,1] \to \mathbb{R}$ with the boundary conditions $x(0) = x_0, x(1) = x_1$, *i.e.*,

$$C_{x_1,x_0}([0,1],\mathbb{R}) = \{\text{continuous } x : [0,1] \to \mathbb{R} | x(0) = x_0, x(1) = x_1\}. \tag{8.51}$$

In this chapter we only consider analytic actions. We will without loss of generality always consider propagation over the domain $[0,1]$. The 'Lorentzian path integral' is performed over the space of continuous paths $C_{x_1,x_0}([0,1],\mathbb{R})$.

Given the action S, we define the Lefschetz thimble. We start by analytically continuing the action S to complex-valued paths

$$C_{x_1,x_0}([0,1],\mathbb{C}) = \{\text{continuous } x : [0,1] \to \mathbb{C} | x(0) = x_0, x(1) = x_1\}\,. \tag{8.52}$$

We write the action in terms of its real and imaginary part, *i.e.*,

$$iS[x] = h[u+iv] + iH[u+iv], \tag{8.53}$$

with $x = u + iv$ where $u \in C_{x_1,x_0}([0,1],\mathbb{R})$ and $v \in C_{0,0}([0,1],\mathbb{R})$, the real part $h : C_{x_1,x_0}([0,1],\mathbb{C}) \to \mathbb{R}$ and the imaginary part $H : C_{0,0}([0,1],\mathbb{C}) \to \mathbb{R}$.

Picard-Lefschetz theory provides a constructive procedure to obtain the Lefschetz thimble $\mathcal{J} \subset C_{0,0}([0,1],\mathbb{C})$ to which the original integration domain $C_{0,0}([0,1],\mathbb{R})$ can be deformed. Given the functional h, we define the downward flow

$$\gamma_\lambda : C_{x_1,x_0}([0,1],\mathbb{C}) \to C_{x_1,x_0}([0,1],\mathbb{C}) \tag{8.54}$$

by the flow equation

$$\frac{\partial \gamma_\lambda(f)}{\partial \lambda} = -\nabla_{u+iv} h(\gamma_\lambda(f)) \tag{8.55}$$

with $\gamma_0(f) = f$ for the complex path $f \in C_{x_1,x_0}([0,1],\mathbb{C})$ and $\lambda \in I \subset \mathbb{R}$. In the flow equation, the functional gradient is defined as

$$\nabla_{u+iv} h = \frac{\delta h}{\delta u} + i\frac{\delta h}{\delta v}\,. \tag{8.56}$$

Note that the gradient requires the definition of a metric on the complex function space. We here use the l^2-norm on function spaces to measure distances. When acting with the flow on the original integration domain, we obtain a one-parameter family of integration contours

$$D_\lambda = \gamma_\lambda(C_{x_1,x_0}([0,1],\mathbb{R}))\,. \tag{8.57}$$

In the limit $\lambda \to \infty$, we obtain the Lefschetz thimble

$$\mathcal{J} = \lim_{\lambda\to\infty} D_\lambda\,. \tag{8.58}$$

When the exponent has saddle points in $C_{x_1,x_0}([0,1],\mathbb{C})$, corresponding to (complex) classical solution, the limit exists and the thimble can be interpreted as a fixed point of the downward flow. It follows that the thimble consists of a collection of steepest descent contours corresponding to the set of relevant saddle points. The exponent evaluated along each steepest descent contour assumes its maximum at the saddle point[4]. The Lefschetz thimble is the optimal deformation of the integration contour, as the integrand is non-oscillator along the flow. To see this write the flow as $\gamma_\lambda(f) = u_\lambda + iv_\lambda$ with

$$\frac{\partial u_\lambda}{\partial \lambda} = -\frac{\delta h}{\delta u}\,, \tag{8.59}$$

$$\frac{\partial v_\lambda}{\partial \lambda} = -\frac{\delta h}{\delta v}\,. \tag{8.60}$$

The flow preserves H since

$$\frac{\mathrm{d}H(u_\lambda + iv_\lambda)}{\mathrm{d}\lambda} = \int \left[\frac{\delta H}{\delta u}\frac{\partial u_\lambda}{\partial \lambda} + \frac{\delta H}{\delta v}\frac{\partial v_\lambda}{\partial \lambda}\right] \mathrm{d}t \tag{8.61}$$

$$= -\int \nabla_{u+iv} h \cdot \nabla_{u+iv} H \,\mathrm{d}t \tag{8.62}$$

$$= 0\,, \tag{8.63}$$

with the multi-dimensional Cauchy-Riemann equations $\nabla_{u+iv} h \cdot \nabla_{u+iv} H = 0$, where the inner product is the sum of the product of the real and imaginary part. The h-functional strictly decreases along the flow, since

$$\frac{\mathrm{d}h(u_\lambda + iv_\lambda)}{\mathrm{d}\lambda} = \int \left[\frac{\delta h}{\delta u}\frac{\partial u_\lambda}{\partial \lambda} + \frac{\delta h}{\delta v}\frac{\partial v_\lambda}{\partial \lambda}\right] \mathrm{d}t \tag{8.64}$$

$$= -\int \left[\left(\frac{\delta h}{\delta u}\right)^2 + \left(\frac{\delta h}{\delta v}\right)^2\right] \mathrm{d}t\,. \tag{8.65}$$

8.2.1 The free particle

It is instructive to evaluate the Lefschetz thimble for a few simple toy-models. We first analyze the Gaussian random walk since this is a finite dimensional problem corresponding to the lattice regionalization scheme. We subsequently consider the imaginary-time and real-time propagator of the free non-relativistic particle.

[4]At a Stokes phenomenon, the steepest descent contour can contain a series of saddle points which are local maxima of the contour.

The Gaussian random walk

A Gaussian random walk, interpolating between the points x_0 and x_1 with N steps consists of a vector $(q_0, q_1, \ldots, q_N) \in \mathbb{R}^{N+1}$ with $q_0 = x_0$ and $q_N = x_1$ for which the difference between neighbouring points $q_{j+1} - q_j$ is Gaussian distributed, *i.e.*,

$$p[\{q_j\}] \propto e^{-\frac{1}{2\sigma^2}\sum_{j=0}^{N-1}(q_{j+1}-q_j)^2}\,, \tag{8.66}$$

with the standard deviation σ (see appendix 8.D). The expectation value of a function $U[\{q_j\}]$ is given by the $(N-1)$-dimensional integral

$$\langle U\rangle \propto \int_{\mathbb{R}^{N-1}} U[\{q_j\}]p[\{q_j\}]\prod_{j=1}^{N-1}\mathrm{d}q_j\,. \tag{8.67}$$

A Picard-Lefschetz analysis of this integral amounts to deforming the integration domain $D = \mathbb{R}^{N-1}$ in the complex plane $\mathbb{C}^{N-1}$ to the contours of steepest descent corresponding to the relevant saddle points. The exponent $f : \mathbb{R}^{N+1} \to \mathbb{R}$ is given by the sum

$$f(\{q_i\}) = -\frac{1}{2\sigma^2}\sum_{j=0}^{N-1}(q_{j+1}-q_j)^2\,. \tag{8.68}$$

The exponent has a unique saddle point, since the difference equation

$$\frac{\partial f}{\partial q_j} = \frac{q_{j-1} - 2q_j + q_{j+1}}{\sigma^2} = 0 \tag{8.69}$$

for $j = 1, \ldots, N-1$, is solved by the linear interpolation

$$\bar{q}_j = x_0 + \frac{j}{N}(x_1 - x_0)\,. \tag{8.70}$$

Note that the saddle point is relevant since $(\bar{q}_0, \bar{q}_1, \ldots, \bar{q}_N) \in \mathbb{R}^{N+1}$.

Analytically continuing the exponent f into the complex plane by considering complex paths, $q_j = u_j + iv_j$, we write f in terms of a real and an imaginary part, *i.e.*,

$$f(\{q_j\}) = h(\{u_j + iv_j\}) + iH(\{u_j + iv_j\}) \tag{8.71}$$

with

$$h(\{u_j + iv_j\}) = -\frac{1}{2\sigma^2}\sum_{j=0}^{N-1}\left[(u_{j+1}-u_j)^2 - (v_{j+1}-v_j)^2\right]\,, \tag{8.72}$$

$$H(\{u_j + iv_j\}) = -\frac{1}{\sigma^2}\sum_{j=0}^{N-1}(u_{j+1}-u_j)(v_{j+1}-v_j)\,. \tag{8.73}$$

The gradient of the real part, h, is given by

$$\nabla_{u+iv} h = (\partial_{u_1} h + i\partial_{v_1} h, \ldots, \partial_{u_{N-1}} h + i\partial_{v_{N-1}} h) \tag{8.74}$$

with the partial derivatives

$$\frac{\partial h}{\partial u_i} = \frac{u_{i-1} - 2u_i + u_{i-1}}{\sigma^2}\,, \quad \frac{\partial h}{\partial v_i} = -\frac{v_{i-1} - 2v_i + v_{i-1}}{\sigma^2}\,. \tag{8.75}$$

Since the gradient flow evaluated on the original integration domain $D = \mathbb{R}^{N-1}$ is directed in the real direction, *i.e.*, $\nabla_{u+iv} h|_D \in \mathbb{R}^{N-1}$, we conclude that the Lefschetz thimble $\mathcal{J}$ coincides with the original integration domain D. The Picard-Lefschetz analysis is thus trivial. This is to be expected as the integrand on the original integration domain is real and convex. Note that the saddle point $\{\bar{q}_j\}$ is included in the thimble $\mathcal{J}$. As we shall see, this also applies to Brownian motion in the continuum limit $N \to \infty$.

The Brownian bridge

The probability distribution of a non-relativistic particle in imaginary-time, also known as the Brownian bridge, is colloquially given by

$$p(x(t)) = e^{-\frac{1}{2}\int_0^1 \dot{x}^2 \mathrm{d}t}\,, \tag{8.76}$$

with x a real-valued continuous function interpolating between x_0 and x_1 (see Figure 8.4). This corresponds to the integration domain of continuous real-valued functions $D = C_{x_1,x_0}([0,1],\mathbb{R})$. The exponent

$$f(x(t)) = -\frac{1}{2}\int \dot{x}^2 \mathrm{d}t \tag{8.77}$$

can be analytically continued into the complex plane by considering complex-valued paths $x \in C_{x_1,x_0}([0,1],\mathbb{C})$. Differentiating with respect to the function x, we obtain the equation of motion

$$\ddot{x} = 0 \tag{8.78}$$

which is solved by the linear motion

$$\bar{x}(t) = x_0 + t(x_1 - x_0)\,. \tag{8.79}$$

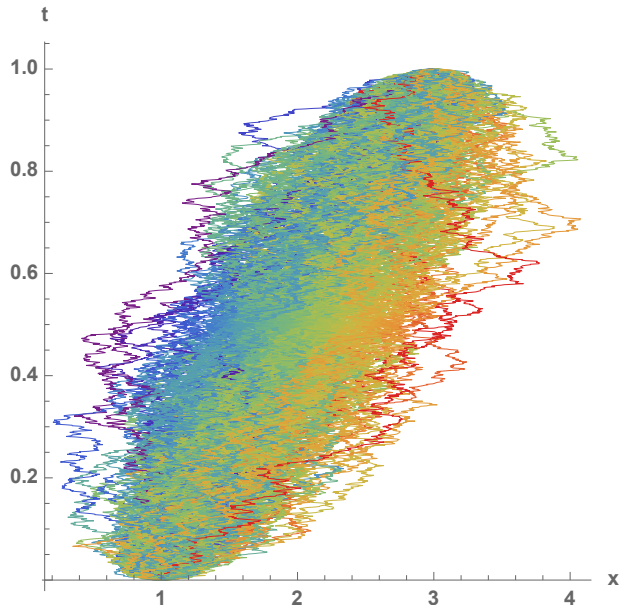


Figure 8.4: Two hundred realizations of a Brownian bridge interpolating between $x(0) = 1$ and $x(1) = 3$.

The saddle point is again relevant as it is included in the original integration domain D.

Writing the path in terms of a real and an imaginary part, *i.e.*, $x = u + iv$ with the boundary conditions $u(0) = x_0, u(1) = x_1$ and $v(0) = v(1) = 0$, the h-function takes the form

$$h[x] = -\frac{1}{2}\int_0^1 \left(\dot{u}^2 - \dot{v}^2\right) \mathrm{d}t\,, \tag{8.80}$$

leading to the gradient flow

$$\nabla_{u+iv} h = \frac{\delta h}{\delta u} + i\frac{\delta h}{\delta v} \tag{8.81}$$

$$= \ddot{u} - i\ddot{v}\,. \tag{8.82}$$

This is the functional analogue of the gradient derived for the random walk.

Note that the gradient ∇h evaluated on real paths is real-valued. As a consequence the thimble $\mathcal{J}$ again coincides with the original integration domain, *i.e.*,

$$\mathcal{J} = C_{x_1,x_0}([0,1],\mathbb{R}) \subset C_{x_1,x_0}([0,1],\mathbb{C})\,. \tag{8.83}$$

Note that saddle point $\bar{x}$ is included in the thimble $\mathcal{J}$ and that the h-function evaluated on the thimble is maximal at the saddle point, *i.e.*,

$$h[x]|_{\mathcal{J}} = -\frac{1}{2}\int \dot{u}^2 \mathrm{d}t\,, \tag{8.84}$$

is maximal for the linear path interpolating between x_0 and x_1. The measure on the thimble $\mu_{\mathcal{J}}$ is the Wiener measure μ_{x_1,x_0}.

The free non-relativistic particle

For the Gaussian random walk and the Brownian bridge, the integrand is real-valued and the Picard-Lefschetz analysis is trivial. For the real-time path integral, the integrand is complex-valued and oscillatory. As a consequence, the Lefschetz thimble takes a more complicated form. The simplest example of a real-time path integral is the free non-relativistic particle.

The propagator of the free particle is colloquially expressed as the path integral

$$G[x_1, x_0, 1] = \int_{x(0)=x_0}^{x(1)=x_1} e^{\frac{i}{\hbar}\int_0^1 \frac{m}{2}\dot{x}^2 \mathrm{d}t} \mathcal{D}x\,, \tag{8.85}$$

where the 'integral' ranges over the domain $D = C_{x_1,x_0}([0,1],\mathbb{R})$. The exponent is given by the imaginary functional

$$f[x] = \frac{i}{\hbar}\int_0^1 \frac{m}{2}\dot{x}^2 \mathrm{d}t\,. \tag{8.86}$$

Variation with respect to the path gives the equation of motion

$$m\ddot{x} = 0\,, \tag{8.87}$$

which is solved by the linear interpolation

$$\bar{x}(t) = x_0 + t(x_1 - x_0)\,. \tag{8.88}$$

The saddle point is relevant as it is included in the original integration domain D.

The analytic continuation of the exponent is defined on the complex-valued paths $C_{x_1,x_0}([0,1],\mathbb{C})$. Writing f in terms of a real and an imaginary part, *i.e.*, $x = u + iv$

with the boundary conditions $u(0) = x_0, u(1) = x_1, v(0) = 0$ and $v(1) = 0$, the exponent can be written in terms of a real and an imaginary part

$$f[x] = h[u + iv] + iH[u + iv]\,, \tag{8.89}$$

with

$$h[u + iv] = -\frac{m}{\hbar}\int_0^1 \dot{u}\dot{v}\mathrm{d}t\,, \tag{8.90}$$

$$H[u + iv] = \frac{m}{2\hbar}\int_0^1 \left[\dot{u}^2 - \dot{v}^2\right]\mathrm{d}t\,. \tag{8.91}$$

The gradient of the h-function is given by

$$\nabla_{u+iv} h = \ddot{v} + i\ddot{u}\,. \tag{8.92}$$

We see that the gradient restricted to the original integration domain is imaginary. The original integration domain is thus not the thimble. From the symmetry of the gradient, it follows that the Lefschetz thimble is given by the functions

$$\mathcal{J} = \{\bar{x} + (1 + i)u | u \in C_{x_1,x_0}([0, 1], \mathbb{R})\} \subset C_{x_1,x_0}([0, 1], \mathbb{C})\,. \tag{8.93}$$

This is in agreement with the thimble obtained by [306]. One can alternatively obtain this thimble by flowing the original integration domain (for more details see the harmonic oscillator).

The thimble $\mathcal{J}$ is the steepest descent contour emanating from the saddle point $\bar{x}$. The thimble corresponds to paths in a rotated spacetime. Observe that the h-function evaluated along the thimble, is equivalent to the the Euclidean theory

$$h[x]|_{\mathcal{J}} = -\frac{m}{\hbar}\int_0^1 \dot{u}^2\mathrm{d}t\,. \tag{8.94}$$

The imaginary part is constant, *i.e.*,

$$H[x]|_{\mathcal{J}} = \frac{m}{2\hbar}\int_0^1 \dot{\bar{x}}^2\mathrm{d}t = \frac{m(x_1 - x_0)^2}{2\hbar}\,. \tag{8.95}$$

It follows that we can relate the integral along the thimble $\mathcal{J}$ to the Wiener measure, *i.e.*, the complex measure along the thimble is given by

$$\mu_{\mathcal{J}}(x) = e^{\frac{i}{\hbar}\frac{m(x_1-x_0)^2}{2}}\mu_{0,0}^{2\hbar/m}(u)\,, \tag{8.96}$$

in agreement with [306]. By analytically continuing $\mu_{\mathcal{J}}$ back to the original integration domain D, we obtain a well-defined path integral measure. This procedure is analogous to the Feynman-Kac formula. However, it does not rely on an analytic continuation of the time variable and applies to more general situations.

We define the map $\Gamma : C_{x_1,x_0}([0,1],\mathbb{R}) \to \mathcal{J}$ by $\Gamma(\bar{x}+\delta x) = \bar{x} + (1+i)\delta x$. Define the measure

$$\mu_D(A) = \int_{\mathcal{J}} \mathbf{1}_A(x) \mathrm{d}\mu_{\mathcal{J}}(x) \tag{8.97}$$

$$= \frac{1}{2}(1-i) \int_{\mathcal{J}} \mathbf{1}_A\left(\bar{x} + (1+i)u\right) \mathrm{d}\mu_{\mathcal{J}}(u) \ . \tag{8.98}$$

with A an element of the Borel σ-algebra of $C_{x_1,x_0}([0,1],\mathbb{R})$ and $\mathbf{1}_A$ the Dirac measure ($\mathbf{1}_A(x) = 1$ if $x \in A$ and 0 otherwise). For the mutually disjoint sets A_j it follows that $\mathbf{1}_{\cup_{j=1}^{\infty} A_j} = \sum \mathbf{1}_{A_j}$. Since the analytic continuation of a sum is the sum of the analytic continuation of its parts, we see that

$$\mu_D(\cup_{j=1}^{\infty} A_j) = \int_{\mathcal{J}} \mathbf{1}_{\cup_{j=1}^{\infty} A_j}(x) \mathrm{d}\mu_{\mathcal{J}}(x) \tag{8.99}$$

$$= \int_{\mathcal{J}} \sum_{j=1}^{\infty} \mathbf{1}_{A_j}(x) \mathrm{d}\mu_{\mathcal{J}}(x) \tag{8.100}$$

$$= \sum_{j=1}^{\infty} \int_{\mathcal{J}} \mathbf{1}_{A_j}(x) \mathrm{d}\mu_{\mathcal{J}}(x) \tag{8.101}$$

$$= \sum_{j=1}^{\infty} \mu_D(A_j) \, . \tag{8.102}$$

The map μ_D is thus a complex σ-measure (see appendix 8.A for the definitions).

It is instructive to evaluate the propagator corresponding to the free particle in a double slit experiment (see Figure 9.36). Assume that the particles reach the slit at time $t = \frac{1}{2}$. Consider the class of paths which pass through the upper slit $[a_1, b_1]$ and lower slit $[a_2, b_2]$ with $[a_1, b_1] \cap [a_2, b_2] = \emptyset$ at time $t = 1/2$, *i.e.*,

$$A_j = \{x \in C_{x_1,x_0}([0,1],\mathbb{R}) | x(1/2) \in [a_j, b_j]\} \tag{8.103}$$

for $j = 1, 2$. Since the Dirac measure $\mathbf{1}_{A_j}(x)$ and the analytic continuation only depends on the position of the path at time $t = 1/2$, *i.e.*, $\mathbf{1}_{A_j}(x) = \mathbf{1}_{[a_j,b_j]}(x(1/2))$, we can use

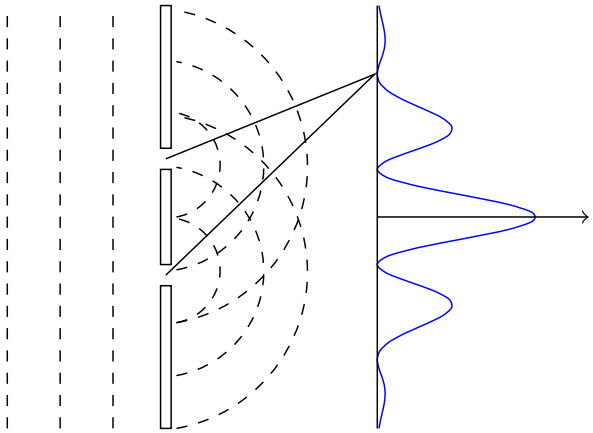


Figure 8.5: The double slit experiment.

Wiener's theorem to write the functional integral as a one-dimensional oscillatory integral

$$\mu_D(A_j) = \int_{\mathcal{J}} \mathbf{1}_{A_j}(x)\mathrm{d}\mu_{\mathcal{J}}(x) \tag{8.104}$$

$$= \frac{1}{2}(1-i)\int_{\mathcal{J}} \mathbf{1}_{A_j}\left(\bar{x}+(1+i)u\right)\mathrm{d}\mu_{\mathcal{J}}(u) \tag{8.105}$$

$$= \frac{1}{2}(1-i)e^{\frac{i}{\hbar}\frac{m(x_1-x_0)^2}{2}}\int_{\mathcal{J}} \mathbf{1}_{[a_i,b_i]}\left(\bar{x}(1/2)+(1+i)u(1/2)\right)\mathrm{d}\mu_{0,0}^{2\hbar/m}(u) \tag{8.106}$$

$$= \frac{1}{2}(1-i)e^{\frac{i}{\hbar}\frac{m(x_1-x_0)^2}{2}}\frac{m}{2\pi\hbar}\int_{\mathbb{R}} \mathbf{1}_{[a_i,b_i]}\left(\bar{x}_{1/2}+(1+i)u_{1/2}\right)e^{-\frac{m}{\hbar}u_{1/2}^2}\mathrm{d}u_{1/2} \tag{8.107}$$

$$= e^{\frac{i}{\hbar}\frac{m(x_1-x_0)^2}{2}}\frac{m}{2\pi\hbar}\int_{\mathbb{R}} \mathbf{1}_{[a_i,b_i]}\left(x_{1/2}\right)e^{\frac{i}{\hbar}\frac{m}{2}\left(x_{1/2}-\bar{x}_{1/2}\right)^2}\mathrm{d}x_{1/2} \tag{8.108}$$

$$= e^{\frac{i}{\hbar}\frac{m(x_1-x_0)^2}{2}}\frac{m}{2\pi\hbar}\int_{a_j}^{b_j} e^{\frac{i}{\hbar}\frac{m}{2}\left(x_{1/2}-\bar{x}_{1/2}\right)^2}\mathrm{d}x_{1/2} \tag{8.109}$$

where $x(1/2), \bar{x}(1/2)$, and $u(1/2)$ are written as $x_{1/2}, \bar{x}_{1/2}$ and $u_{1/2}$. Formally, we can express the measure of the set A_j, $\mu_D(A_j)$, in terms of error functions. The measure of the union of the two disjoint sets

$$\mu_D(A_1 \cup A_2) = e^{\frac{i}{\hbar}\frac{m(x_1-x_0)^2}{2}}\frac{m}{2\pi\hbar}\int_{[a_1,b_1]\cup[a_2,b_2]} e^{\frac{i}{\hbar}\frac{m}{2}\left(x_{1/2}-\bar{x}_{1/2}\right)^2}\mathrm{d}x_{1/2} \tag{8.110}$$

gives the interference pattern of the double slit experiment.

8.2.2 The harmonic oscillator

The non-relativistic harmonic oscillator is another example for which we can evaluate the thimble exactly. We study both the harmonic and the inverted harmonic oscillators in real-time. The inverted harmonic oscillator turns out to be rather similar to the free particle. For the harmonic oscillator, we obtain a slightly more involved Lefschetz thimble.

The real-time path integral for the harmonic oscillator is given by

$$\int_{x(0)=x_0}^{x(1)=x_1} e^{iS[x]}\mathcal{D}x\,, \tag{8.111}$$

with the action

$$S[x] = \int \left[\frac{\dot{x}^2(t)}{2} - Ax(t) - \frac{\omega^2}{2}x^2(t)\right] \mathrm{d}t\,, \tag{8.112}$$

for real A and ω^2. The equation of motion

$$0 = \frac{\delta S[x]}{\delta x(t)} = -\left(\ddot{x} + A + \omega^2 x\right)\,, \tag{8.113}$$

is solved by the saddle point

$$\bar{x}(t) = -\frac{A}{\omega^2} + \frac{1}{\omega^2 \sin\omega}\left[(A + x_0\omega^2)\sin\left[\omega(1-t)\right] + (A + x_1\omega^2)\sin\left[\omega t\right]\right]\,. \tag{8.114}$$

The saddle point is relevant since it is included in the original integration domain.

Writing the complex path as $x = \bar{x} + u + iv$, with $u, v \in C_{0,0}([0,1],\mathbb{R})$, the exponent $f = iS$ can be expressed in terms of a real and an imaginary part

$$iS[x] = iS[\bar{x}] + h[u+iv] + iH[u+iv]\,, \tag{8.115}$$

with

$$h[u+iv] = -\int \left[\dot{u}\dot{v} - \omega^2 uv\right] \mathrm{d}t\,, \tag{8.116}$$

$$H[u+iv] = S[\bar{x}] + \frac{1}{2}\int \left[\dot{u}^2 - \dot{v}^2 - \omega^2(u^2 - v^2)\right] \mathrm{d}t\,. \tag{8.117}$$

Note that the h-functional is independent of the linear term in the potential, since $\bar{x}$ is a critical point of the action S.

The gradient of the functional h takes the simple form

$$\nabla_{u+iv} h[u+iv] = [\ddot{v} + Bv] + i\left[\ddot{u} + \omega^2 u\right] . \tag{8.118}$$

The downwards flow $\gamma_\lambda(f) = \bar{x} + u_\lambda + iv_\lambda$ can be written in terms of the coupled partial differential equations

$$\partial_\lambda u_\lambda = -\left[\ddot{v}_\lambda + \omega^2 v_\lambda\right] , \tag{8.119}$$

$$\partial_\lambda v_\lambda = -\left[\ddot{u}_\lambda + \omega^2 u_\lambda\right] , \tag{8.120}$$

with the initial condition $\bar{x} + u_0 + iv_0 = f \in C_{x_1,x_0}([0,1],\mathbb{C})$. By changing coordinates to the Fourier modes

$$u_\lambda = \sum_n \alpha_n(\lambda)\sin(\pi n t) , \tag{8.121}$$

$$v_\lambda = \sum_n \beta_n(\lambda)\sin(\pi n t) , \tag{8.122}$$

the system simplifies to the set of ordinary differential equations

$$\alpha_n'(\lambda) = ((\pi n)^2 - \omega^2)\beta_n(\lambda) , \tag{8.123}$$

$$\beta_n'(\lambda) = ((\pi n)^2 - \omega^2)\alpha_n(\lambda) , \tag{8.124}$$

$$\tag{8.125}$$

which is solved by the exponential

$$\vec{\alpha}_n(\lambda) = e^{M\lambda}\vec{\alpha}_n(0) , \tag{8.126}$$

with $\vec{\alpha}_n = (\alpha_n, \beta_n)$ and

$$M = \begin{pmatrix} 0 & (\pi n)^2 - \omega^2 \\ (\pi n)^2 - \omega^2 & 0 \end{pmatrix} , \tag{8.127}$$

leading to the flow

$$\gamma_\lambda = \bar{x} + \sum_n e^{M\lambda}\vec{\alpha}_n(0) \cdot \begin{pmatrix} 1 \\ i \end{pmatrix} \sin(\pi n t) . \tag{8.128}$$

The behaviour of the flow is determined by the eigenvalues

$$\lambda_\pm = \pm((\pi n)^2 - \omega^2) \tag{8.129}$$

and the corresponding eigenvectors

$$v_\pm = (1, \pm 1) , \tag{8.130}$$

of the matrix M.

Inverted harmonic oscillator

Using the flow, we evaluate the thimble. For the inverted harmonic oscillator, *i.e.*, $\omega^2 < 0$, we obtain a positive and a negative eigenvalue. As a consequence, the steepest descent flow emanating from the saddle point $\bar{x}$ consists of the paths $x = \bar{x} + f$ with $f \in C_{0,0}([0,1], \mathbb{C})$, for which the upwards flow terminates in the saddle point $\bar{x}$. In terms of the flow, the thimble consists of the f for which

$$\lim_{\lambda \to -\infty} \gamma_\lambda(f) = \bar{x} \,. \tag{8.131}$$

For these paths, $x = \bar{x} + f$, the coefficients in the Fourier analysis are parallel to the eigenvector corresponding to the positive eigenvalue. For the inverted harmonic oscillator this is the eigenvector $v_+ = (1,1)$. We thus conclude that the thimble is given by the set of complex paths

$$\mathcal{J} = \{\bar{x} + (1+i)u | u \in C_{0,0}([0,1], \mathbb{R})\} \,. \tag{8.132}$$

The flow restricted to the thimble takes the simple form

$$\bar{x} + u_\lambda + i v_\lambda = \bar{x} + (1+i) \sum_n \alpha_n(0) e^{((\pi n)^2 - \omega^2)\lambda} \sin(\pi n t) \,. \tag{8.133}$$

The flow amplifies the fluctuations. A generic initial path flows to a path with an infinite number of oscillations with an infinite amplitude.

The steepest ascent contour is given by the paths moving in the orthogonal direction

$$\mathcal{K} = \{\bar{x} + (1-i)u | u \in C_{0,0}([0,1], \mathbb{R})\} \,. \tag{8.134}$$

Note that the Lefschetz thimble is similar to the thimble of free particle. We can interpreted the thimble, $\mathcal{J}$, as the set of paths on a complex spacetime.

The exponent evaluated along the thimble

$$iS[x]|_{\mathcal{J}} = iS[\bar{x}] - \int \left(\dot{u}^2 - \omega^2 u^2\right) \mathrm{d}t \,, \tag{8.135}$$

is convex and assumes its maximum at the classical solution $\bar{x}$. In analogy with the exponent of the Brownian bridge, we can construct a countable additive measure on the thimble

$$\mu_{\mathcal{J}}(A) = e^{iS[\bar{x}]} \int_A e^{\omega^2 \int_0^1 u^2 \mathrm{d}t} \mathrm{d}\mu_{x_1,x_0}(u) \tag{8.136}$$

for any element A of the Borel σ-algebra of $\mathcal{J}$.

After continuing the measure to the original integration domain, we obtain the path integral measure

$$\mu_D(A) = \int_{\mathcal{J}} \mathbf{1}_A(x) \mathrm{d}\mu_{\mathcal{J}}(x)\,. \tag{8.137}$$

For more details see section 8.3.

Harmonic oscillator

For the harmonic oscillator, *i.e.*, $\omega > 0$, the analysis is a bit more involved. For the high frequency modes, *i.e.*, $|n| > \frac{\omega}{\pi}$, the eigenvalue λ_+ is positive and λ_- is negative. This is the same as for the inverted harmonic oscillator. For a finite number of low frequency modes, for which $|n| < \frac{\omega}{\pi}$, the two eigenvalues switch role. As a consequence, the thimble consists of the paths

$$\mathcal{J} = \left\{ \bar{x} + (1-i) \sum_{|\pi n|<\omega} a_n \sin(\pi n t) + (1+i) \sum_{|\pi n|>\omega} a_n \sin(\pi n t) | a_n \in \mathbb{R} \right\}. \tag{8.138}$$

The integer, n, for which the thimble switches from $(1-i)$ to $(1+i)$ is the Maslov-Morse index of the classical trajectory [16]. The thimble is thus no longer the paths in a complex spacetime.

We can write any element $x \in \mathcal{J}$ in terms of a low and a high frequency part, *i.e.*, $x = \bar{x} + x_l + x_h$ where the low frequency part $x_l = (1-i)u_l$ and the high frequency part $x_h = (1+i)u_h$ with u_r and u_l real-valued paths. The exponent evaluate along the thimble takes the form

$$iS[x]|_{\mathcal{J}} = iS[\bar{x}] + \int \left(\dot{u}_l^2 - \omega^2 u_l^2\right) \mathrm{d}t - \int \left(\dot{u}_h^2 - \omega^2 u_h^2\right) \mathrm{d}t\,. \tag{8.139}$$

The high-frequency part has a negative-sign kinetic term and corresponds to a Wiener-like measure. The low-frequency part has a positive kinetic term. However since u_l consists of only a finite number of Fourier modes, we can interpret this term as a finite dimensional Lebesgue measure. We thus construct the countably additive σ-measure

$$\mu_{\mathcal{J}}(A) = e^{iS[\bar{x}]} \int_A e^{\omega^2 \int_0^1 u_h^2 \mathrm{d}t}\, \mathrm{d}\mu_l(u_l) \times \mathrm{d}\mu_{x_1,x_0}(u_h)\,, \tag{8.140}$$

with the low frequency Lebesgue measure

$$\mu_l(A_l) = \int_{A_l} e^{\frac{1}{2}\sum_{|\pi n|<\omega}((n\pi)^2-\omega^2)a_n^2} \prod_{|\pi n|<\omega} \mathrm{d}a_n \,, \tag{8.141}$$

with A_l an element of the Borel σ-algebra corresponding to the set of low frequency paths. This is in agreement with [306].

By continuing $\mu_{\mathcal{J}}$ back to the original integration domain, we obtain the path integral σ-measure

$$\mu_D(A) = \int_{\mathcal{J}} \mathbf{1}_A(x) \mathrm{d}\mu_{\mathcal{J}}(x) \,. \tag{8.142}$$

For more details see section 8.3.

8.2.3 The relativistic particle

The propagator of the relativistic particle moving in a potential V is colloquially written as the path integral

$$G[x_1; x_0] = \int_0^\infty \mathrm{d}N \int_{x^\mu(0)=(t_0,x_0)}^{x^\mu(1)=(t_1,x_1)} \mathcal{D}x^\mu e^{\frac{i}{\hbar}\int\left[\frac{m}{2N}(\dot{t}^2-\dot{x}^2)+N\left(\frac{m}{2}+V[x^\mu(s)]\right)\right]\mathrm{d}s} \tag{8.143}$$

with N the Schwinger time of the particle. Depending on the complexity of the potential, we can evaluate the Lefschetz thimble.

The free relativistic particle

When the potential vanishes, $V = 0$, the equations of motion $\delta f = 0$, is solved by the two saddle points

$$(\bar{x}^\mu(s), \bar{N}) = \left(x_0^\mu + s(x_1^\mu - x_0^\mu), \pm\sqrt{(x_0^\mu - x_1^\mu)^2}\right) \,, \tag{8.144}$$

with the boundary conditions $x_0^\mu = (t_0, x_0), x_1^\mu = (t_1, x_1)$, and the exponent

$$f[x^\mu, N] = i\frac{m}{2\hbar}\int \left[\frac{\dot{t}^2 - \dot{x}^2}{N} + N\right] \mathrm{d}s \,. \tag{8.145}$$

For time-like separated spacetime points x_0^μ, x_1^μ, only the saddle point with positive $\bar{N}$ is relevant since it is included in the original integration domain and the real part h is symmetric with respect to $N \mapsto -N$. For space-like separated spacetime points x_0^μ, x_1^μ the saddle points in the lapse are imaginary.

For the free particle, the integrand factorizes into a term depending on the spatial $x(s)$ and temporal $t(s)$ path. We evaluate the thimble by first flowing the t and x component for general complex N. We can subsequently flow the integration domain for N to obtain the complete Lefschetz thimble $\mathcal{J}$. The time dependent component of the exponent is given by the Gaussian

$$f_t[t, N] = i\frac{m}{2\hbar}\int \frac{\dot{t}^2}{N}\mathrm{d}s\,. \tag{8.146}$$

In analogy with the non-relativistic free particle we obtain the thimble

$$\mathcal{J}_t(N) = \{\bar{t} + (1+i)\sqrt{N}u_t | u_t \in C_{0,0}([0,1],\mathbb{R})\}\,. \tag{8.147}$$

The spatial component with the exponent

$$f_x[x, N] = -i\frac{m}{2\hbar}\int \frac{\dot{x}^2}{N}\mathrm{d}s \tag{8.148}$$

leads to the thimble

$$\mathcal{J}_x(N) = \{\bar{x} + (1-i)\sqrt{N}u_x | u_x \in C_{0,0}([0,1],\mathbb{R})\}\,. \tag{8.149}$$

The exponent evaluated on the manifold $\mathcal{J}_t(N) \times \mathcal{J}_x(N)$ takes the form

$$f[x^\mu, N]|_{\mathcal{J}_t(N)\times\mathcal{J}_x(N)} = i\frac{m}{2\hbar}\left[\frac{(x_0^\mu - x_1^\mu)^2}{N} + N\right] - \frac{m}{\hbar}\int(\dot{u}_t^2 + \dot{v}_t^2)\mathrm{d}s\,. \tag{8.150}$$

Since the fluctuations u_t, u_x decouple from the Schwinger time N, we can flow the integration domain for N by considering the exponent

$$f_N[N] = i\frac{m}{2\hbar}\left[\frac{(x_0^\mu - x_1^\mu)^2}{N} + N\right]\,, \tag{8.151}$$

leading to the thimble

$$\mathcal{J}_N = \left\{ N_r + iN_i \middle| N_i = \frac{N_r - \sigma}{\sqrt{\frac{2\sigma}{N_r} - 1}} \text{ for } N_r \in (0, 2\sigma) \right\}\,, \tag{8.152}$$

with the proper distance $\sigma = \sqrt{(x_0^\mu - x_1^\mu)^2}$. See Figure 8.6 for an sketch of $\mathcal{J}_N$ for time-like and space-like separated x_0^μ and x_1^μ.

We conclude that the Lefschetz thimble of the free relativistic particle is given by the manifold

$$\mathcal{J} = \{(\bar{t} + (1+i)\sqrt{N}u_t, \bar{x} + (1-i)\sqrt{N}u_x, N) | u_t, u_x \in C_{0,0}([0,1], \mathbb{R}) \text{ and } N \in \mathcal{J}_N\}\,. \tag{8.153}$$

The exponent f evaluated along the thimble takes the form

$$f[x^\mu, N]|_{\mathcal{J}} = i\frac{m\sigma}{\hbar} - \frac{m}{2\hbar}\frac{2(N_r - \sigma)^2}{\sqrt{2\sigma N_r - N_r^2}} - \frac{m}{\hbar}\int (\dot{u}_t^2 + \dot{v}_t^2)\,. \tag{8.154}$$

The part corresponding to the path (u_t, u_x) leads to a Wiener measure. The part corresponding to the Schwinger time is a convex function leading to a one-dimensional σ-measure. We thus can construct the σ-measure on the thimble

$$\mathrm{d}\mu_{\mathcal{J}}(x^\mu, N) = e^{\frac{i}{\hbar}m\sigma} e^{-\frac{m}{2\hbar}\frac{2(N_r-\sigma)^2}{\sqrt{2\sigma N_r - N_r^2}}} \mathrm{d}\mu(N) \times \mathrm{d}\mu^{2\hbar/m}_{(t_1,x_1),(t_0,x_0)}(u_t, u_x)\,. \tag{8.155}$$

Note that this integral has support in the space of continuous non-differentiable paths. This should be contrasted with the multi-dimensional Wiener measure used in quantum field theory.

Given the measure on the thimble $\mathcal{J}$, we can construct the measure on the original integration domain

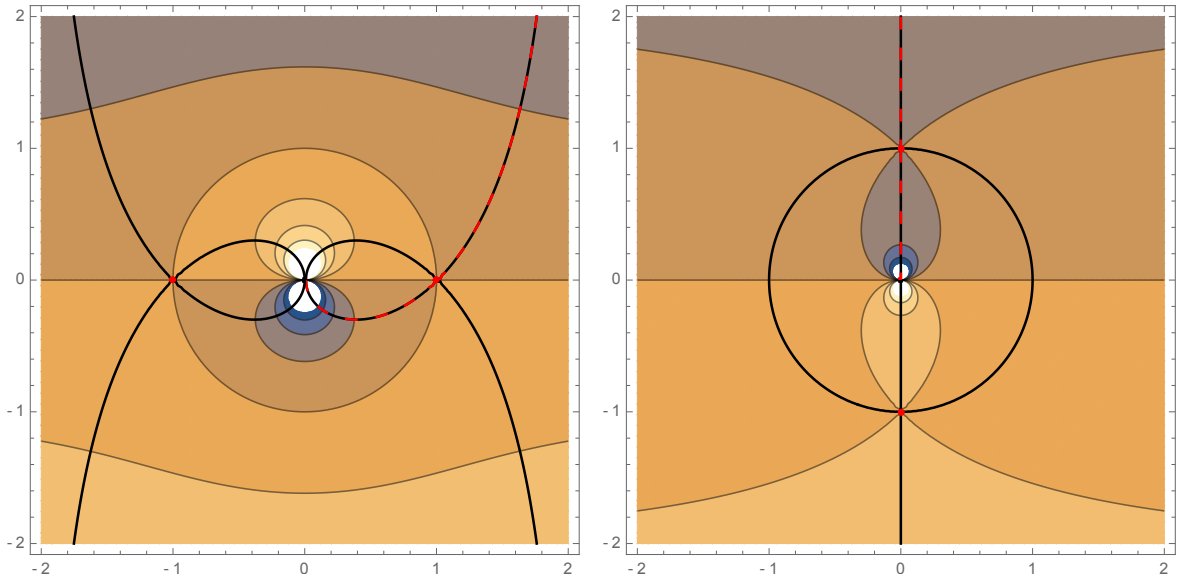
$$\mu_D(A) = \int_{\mathcal{J}} \mathbf{1}(A) \mathrm{d}\mu_{\mathcal{J}}(A)\,, \tag{8.156}$$

as for the non-relativistic particle. For more details see section 8.3.

8.2.4 Perturbation theory

In the previous sections we evaluated the Lefschetz thimble $\mathcal{J}$ and constructed the corresponding σ-measure for several toy-models. Unfortunately, we can only evaluate the Lefschetz thimble exactly for a limited set of path integrals, as the action

$$S = \int \left[\frac{m}{2}\dot{x}^2 - V(x)\right] \mathrm{d}t \tag{8.157}$$



(a) Time-like separated boundary conditions. (b) Space-like separated boundary conditions.

Figure 8.6: The Lefschetz thimble in the complex N-plane. The saddle points $\bar{N}$ (in red), the lines of steepest ascent and descent (in black), and the h-function $\mathrm{Re}[f_N]$ on the background. The Lefshetz thimble $\mathcal{J}$ is represented by the red dashed line.

with a more general potential, induces a downward flow consisting of non-linear partial differential equation. In this section, we use perturbation theory to extend the analysis of the thimble to more general quantum mechanical models. In the next section, we consider the structure of the thimble in the non-perturbative regime.

For the free non-relativistic particle, the Lefschetz thimble corresponding to the propagation from position x_0 to x_1 in a unit of time is given by

$$\mathcal{J} = \{\bar{x} + (1+i)u | u \in C_{x_1,x_0}([0,1],\mathbb{R})\} \subset C_{x_1,x_0}([0,1],\mathbb{C})\,, \tag{8.158}$$

with $\bar{x}$ the linear interpolation,

$$\bar{x}(t) = x_0 + t(x_1 - x_0)\,. \tag{8.159}$$

Writing the path on the thimble as the classical trajectory, $\bar{x}$, plus a fluctuation,

$$x = \bar{x} + (1+i)u\,, \tag{8.160}$$

the measure on the thimble is given by

$$\mu_{\mathcal{J}}(x) = e^{\frac{i}{\hbar}\frac{m(x_1-x_0)^2}{2}} \mu_{0,0}^{2\hbar/m}(u)\,, \tag{8.161}$$

with the Wiener measure $\mu_{0,0}^{2\hbar/m}$.

We write the propagator for the particle moving in the potential V in terms of the free theory by expanding the potential term

$$G[x_1;x_0;1] = \int_{x(0)=x_0}^{x(1)=x_1} \mathcal{D}x\ e^{\frac{i}{\hbar}\int_0^1\left[\frac{m}{2}\dot{x}^2 - V(x)\right]\mathrm{d}t} \tag{8.162}$$

$$= \int_{x(0)=x_0}^{x(1)=x_1} \mathcal{D}x\ e^{\frac{i}{\hbar}\int_0^1 \frac{m}{2}\dot{x}^2\mathrm{d}t} \sum_{n=0}^{\infty} \frac{\left[-\frac{i}{\hbar}\int_0^1 V(x)\mathrm{d}t\right]^n}{n!} \tag{8.163}$$

$$= G_0[x_1;x_0;1] + G^{(1)}[x_1;x_0;1] + G^{(2)}[x_1;x_0;1] + \ldots \tag{8.164}$$

where

$$G_0[x_1;x_0;1] = \int_{x(0)=x_0}^{x(1)=x_1} e^{\frac{i}{\hbar}\int_0^1 \frac{m}{2}\dot{x}^2\mathrm{d}t}\mathcal{D}x \tag{8.165}$$

$$G^{(1)}[x_1;x_0;1] = -\frac{i}{\hbar}\int_{x(0)=x_0}^{x(1)=x_1} e^{\frac{i}{\hbar}\int_0^1 \frac{m}{2}\dot{x}^2\mathrm{d}t}\left[\int_0^1 V(x(s))\mathrm{d}s\right]\mathcal{D}x \tag{8.166}$$

$$G^{(2)}[x_1;x_0;1] = -\frac{1}{2\hbar^2}\int_{x(0)=x_0}^{x(1)=x_1} e^{\frac{i}{\hbar}\int_0^1 \frac{m}{2}\dot{x}^2\mathrm{d}t}\left[\int_0^1 V(x(s))\mathrm{d}s\int_0^1 V(x(s'))\mathrm{d}s'\right]\mathcal{D}x\,. \tag{8.167}$$

The full propagator thus consists of the sum of weak-values of the potentials. The free propagator, G_0, is the amplitude that the particle travels from x_0 to x_1, without scattering with the potential. The n^{th} order propagator, $G^{(n)}$, is the amplitude that the particle arrives at x_1 after being scattered at n points in its path.

In perturbation theory, it is natural to define these real-time functional integrals by deforming the integration domain $C_{x_1,x_0}([0,1],\mathbb{R})$ to the Lefschetz thimble $\mathcal{J}$ of the free theory, given by equation (8.158)[5]. Unfortunately, this cannot be done for general potentials, *e.g.*, polynomial potentials will generally not converge on the free thimble $\mathcal{J}$ (see the

[5]Note that is also possible to construct a thimble for every scattering propagator $G^{(n)}$ by gluing together free thimbles. In this approach, the scattering amplitude will have different thimbles. We will not consider this glued thimble here.

analysis in Section 8.2.5). We, for this reason, will restrict our analysis to localized rational potentials, such as

$$V(x) = \frac{1}{1+x^{2n}} \tag{8.168}$$

with $n \geq 1$, for which the analytic continuation vanishes for large $|x|$ in any direction in the complex plane. Note that most functions can be well represented with a Padé approximation.

In general, we can deform the original integration domain to the free thimble, by continuously deforming the real path $x = \bar{x} + \delta x$ to the complex path

$$x_\rho = \bar{x} + \text{Exp}\left[i\frac{\pi\rho}{4}\right]\delta x\,, \tag{8.169}$$

with $\rho \in [0,1]$. For every time $t \in [0,1]$, we continuously rotate the real line, $\mathbb{R}$, around the point $\bar{x}(t)$ to the thimble, *i.e.*,

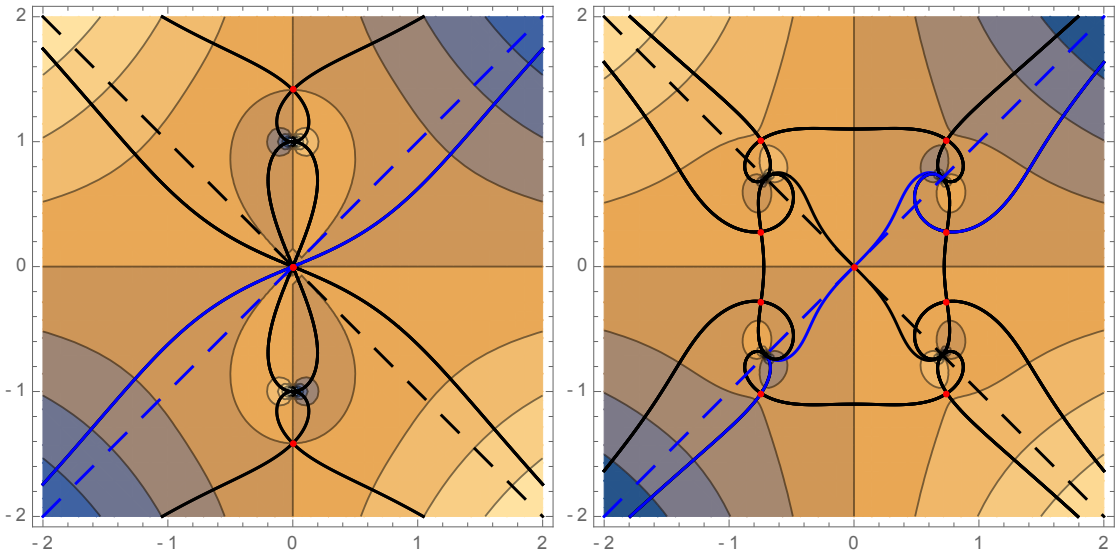
$$\bar{x}(t) + e^{i\frac{\pi\rho}{4}}\mathbb{R}\,. \tag{8.170}$$

The scattering propagator, $G^{(n)}$, evaluated over the real-valued paths coincides with the path integral over the thimble, when the analytic continuation of the potential, V, is holomorphic between the two integration domains. For example, the integral over the real-valued paths coincides with the integral over the free thimble corresponding to the stationary path, $x_0 = x_1 = 0$, $\bar{x} = 0$, for the potential $V = \frac{1}{1+x^2}$, since the poles $x = \pm i$ do not lay between the original integration domain and the thimble. However, when the path $\bar{x}$ includes points laying outside the domain $[-1,1]$, this is no-longer true. For example, for the boundary conditions $x_0 = -2, x_1 = 0$ – with the linear interpolation $\bar{x} = -2 + 2t$ – the deformation for $t \in [0,1/2]$ passes the pole at $x = i$. The real-valued paths can still be deformed to the thimble, however, there is an additional contribution from the pole. In the special case considered here, the pole leads to an correction $\pm\pi$ for the time spend outside of the domain $[-1,1]$, for every potential present in the scattering propagator, *i.e.*, in this case the single scattering propagator

$$G^{(1)}[0;-2;1] = -\frac{i}{\hbar}\int_{x(0)=-2}^{x(1)=0} e^{\frac{i}{\hbar}\int_0^1 \frac{m}{2}\dot{x}^2 \mathrm{d}t}\left[\int_0^1 \frac{1}{1+x(s)^2}\mathrm{d}s\right]\mathcal{D}x \tag{8.171}$$

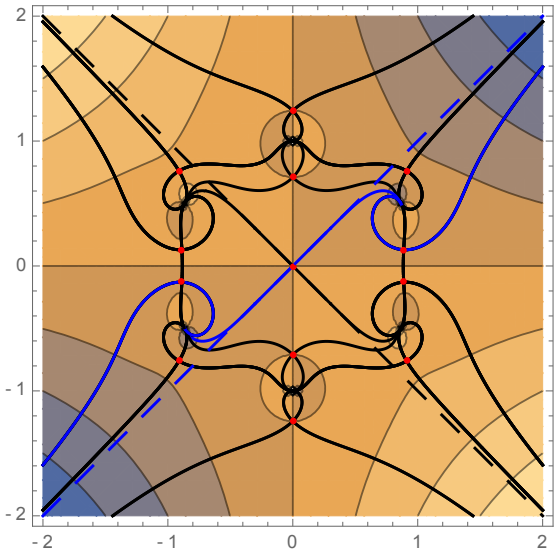
$$= -\frac{i}{\hbar}\int_{\mathcal{J}}\left[\int_0^1 \frac{1}{1+x(s)^2}\mathrm{d}s\right]\mathrm{d}\mu_{\mathcal{J}} + \frac{\pi}{2}. \tag{8.172}$$

Note that we can on the thimble change the order of integration, as the integral is absolutely convergent and the Fubini's theorem applies.

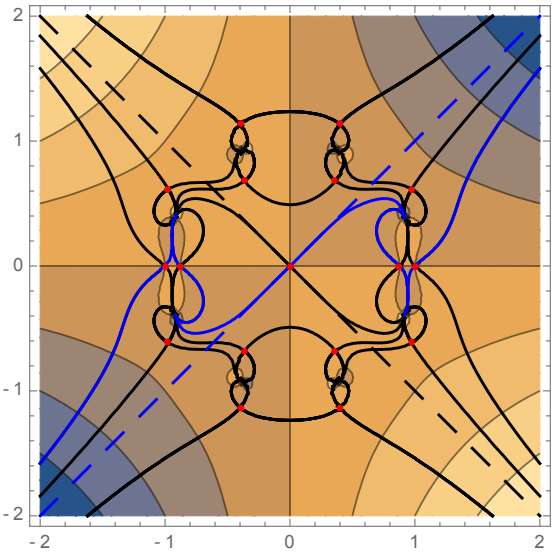


(a) $n = 1$

(b) $n = 2$



(c) $n = 3$



(d) $n = 4$

Figure 8.7: The Picard-Lefschetz diagram for the toy-model for $n = 1, 2, 3, 4$. The red points are the saddle points of the full integral. The dotted blue line and the dotted black line are the Lefschetz thimble and the steepest ascent contour of Gaussian. The sold blue and black lines are the steepest ascent and descent contours corresponding to the saddle points. The background colour is the h-function. The Lefschetz thimble of the full integral is represented by the solid blue curves.

This phenomenon is especially apparent in the one-dimensional toy-model

$$\int_{-\infty}^{\infty} e^{i\left[x^2+\frac{1}{1+x^{2n}}\right]}\,, \tag{8.173}$$

for $n \geq 1$. The quadratic term corresponds to the kinetic term, and the rational function represents the potential term in the action. We here compare the integral along the Lefschetz thimble with the integral along thimble of the Gaussian integral. See Figure 8.7 for the Picard-Lefschetz analysis of the cases $n = 1, 2, 3, 4$. The exponent

$$f(x) = i\left[x^2 + \frac{1}{1+x^{2n}}\right] \tag{8.174}$$

always has a relevant saddle point at the origin in the complex plane (the red point at the origin), as the first order derivative of the exponent,

$$f'(x) = i\left(2x - \frac{2nx^{2n-1}}{(1+x^{2n})^2}\right)\,, \tag{8.175}$$

vanishes for $x = 0$. In the neighborhood of the saddle point, the effect of the potential is small. The thimble is close to the thimble corresponding to the Gaussian term, $\mathcal{J}_0 = (1+i)\mathbb{R}$. The Gaussian term resembles the free particle. In Figure 8.7 the steepest descent and the steepest ascent contours of the Gaussian integral are represented by the blue and black dotted lines.

The full integral has $2n$ poles, at the roots of unity of $x^{2n} = 1$. These poles lead to $4n$ additional saddle points, with a corresponding set of steepest descent and steepest ascent contours (the solid blue and black curves). The real part

$$h[x] = \mathrm{Re}\left[f(x)\right]\,, \tag{8.176}$$

plotted on the background, enables us to distinguish the various steepest ascent and descent contours. The Lefschetz thimble of the full integral is given by the blue curves. Note that the Picard-Lefschetz diagram far away from the origin resembles the diagram of the Gaussian integral, as the quadratic term dominates over the rational term for large $|x|$.

In these diagrams, we observe that:

- For $n = 1$, the Lefschetz analysis consists of only one relevant saddle point. The thimble of the full theory can be continuously deformed to the thimble of the Gaussian. This is a situation in which we can evaluate the full integral along the Gaussian thimble.

- For $n = 2$, two of the poles of the full integral lie on the Gaussian thimble. The integral along the Gaussian thimble is highly oscillatory and ambiguous, as the integral can be both deformed above and below the pole. The natural definition of the integral, for which the integrand vanishes near the poles, passes below the lower left and above the upper right pole. This integral is inequivalent to the real-time integral, as this choice of the Gaussian thimble cannot be deformed to the thimble of the full integral or to the real line. Note that the integral along the Gaussian thimble selects the upper left and lower right saddle point, whereas the thimble of the real integral selects two other saddle points. The difference between the two integrals is equal to the residues of the two relevant poles, as can be proven with Cauchy's theorem.

- For $n = 3$, the Gaussian thimble is inequivalent to the thimble of the full integral. Perturbation theory along the Gaussian thimble will converge but will differ from the real integral. The difference is given by the residue of the lower left and upper right poles.

- For $n = 4$, the Gaussian thimble is again inequivalent to the integral along the real line. Note that in this case, the real integral includes four real-valued saddle points which are not included in an integral along the Gaussian thimble.

This toy-model demonstrates that the perturbative propagator of the Lorentzian theory, describing the motion of a particle in a localized potential, can, in general, be written in terms of Wiener measure along the free thimble $\mathcal{J}_0$. However, for generic potentials, the integral along the thimble does not suffice to evaluate the real-time path integral. We, in general, need to include the contribution of the singularities of the analytic continuation of the potential.

It follows from the Schwinger representation of the relativistic propagator, that the perturbative analysis of the non-relativistic particle in terms of the free thimble and the singularities, naturally extends to the relativistic particle. The propagator is an integral over the Schwinger time and for fixed Schwinger time, the analysis is equivalent to the one of the non-relativistic particle.

8.2.5 Beyond perturbation theory

In the perturbative analysis, in the previous section, we observed that the evaluation of the real-time path integral along the free thimble requires an analysis of the singularities of the analytic continuation of the potential. We, moreover, observed that a perturbative analysis

only makes sense for localized potentials, which do not affect the asymptotic properties of the h-function for large $|x|$.

These complications do not arise in the Picard-Lefschetz analysis of the full theory. The downwards flow

$$\frac{\partial \gamma_\lambda(f)}{\partial \lambda} = -\nabla_{u+iv} h(\gamma_\lambda(f)) \tag{8.177}$$

is well-defined for arbitrary analytic actions. It follows from Morse-Smale theory [256, 303] that the flow of the original integration domain,

$$D_\lambda = \gamma_\lambda(C_{x_1,x_0}([0,1],\mathbb{R}))\,, \tag{8.178}$$

generally converges to the Lefschetz thimble

$$\lim_{\lambda\to\infty} D_\lambda = \mathcal{J} \tag{8.179}$$

in terms of a set of steepest descent contours $\mathcal{J}_j$ corresponding to the relevant saddle points $\bar{x}_j$ for some index j. See Figure 8.8 for the general structure of the thimble in complex function space. Note that when the action, S, has no saddle points, there does not exist a classical limit of the quantum theory.

Along the steepest descent contours of the thimble, $\mathcal{J}_i$, the integrand is non-oscillatory and convex, as the downward flow preserves the imaginary part of the exponent. In the Picard-Lefschetz analysis of the harmonic oscillator, we observed that the analytic continuation of the kinetic term restricted to the thimble can consist of the sum of both positive and negative modes. The thimble of the harmonic oscillator consists of a finite number low-frequency modes, for which the kinetic term contributed positively, and an infinite number of high-frequency modes, with a negatively kinetic term. When the kinetic term only contributes positively to the exponent for a finite number of Fourier modes, we can use the Wiener measure to construct a σ-measure on the thimble. In this situation, the thimble satisfies all the properties one might wish to have for a 'Euclidean theory'. The thimble in this situation leads to a proper definition of the corresponding functional integral on the thimble. As we discuss in section 8.3, this measure on the thimble generally induces a well-defined σ-measure on the space of real-valued paths. We here conjecture that there exists a large class of actions, consisting of only first-order derivatives in time, for which the Lefschetz thimble leads to a continuous-time regularization of the real-time path integral. We anticipate that this class will include the Lorentzian path integral for gravity. In the following discussion, we will indicate why we think that this is a reasonable conjecture. We are currently in the process of developing a rigorous proof.

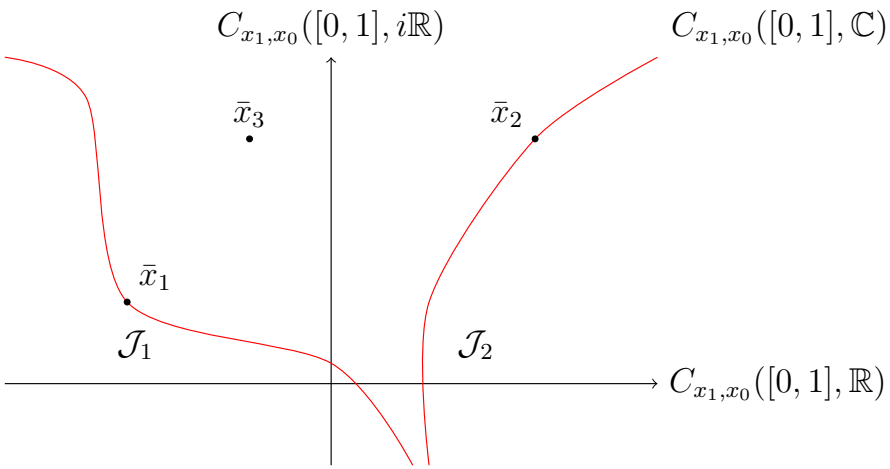


Figure 8.8: The complex function plane $C_{x_1,x_0}([0,1],\mathbb{C})$, with three saddle points $\bar{x}_1, \bar{x}_2, \bar{x}_3$ and the Lefschetz thimble $\mathcal{J} = \mathcal{J}_1 + \mathcal{J}_2$ consisting of two steepest descent contours $\mathcal{J}_1$ and $\mathcal{J}_2$ (red lines).

Consider the action

$$S = \int \left[\frac{m}{2}\dot{x}^2 - V(x)\right] \mathrm{d}t\,. \tag{8.180}$$

We will for simplicity assume the potential to be a polynomial, *i.e.*,

$$V[x] = \sum_{k=1}^{K} a_k x^k\,, \tag{8.181}$$

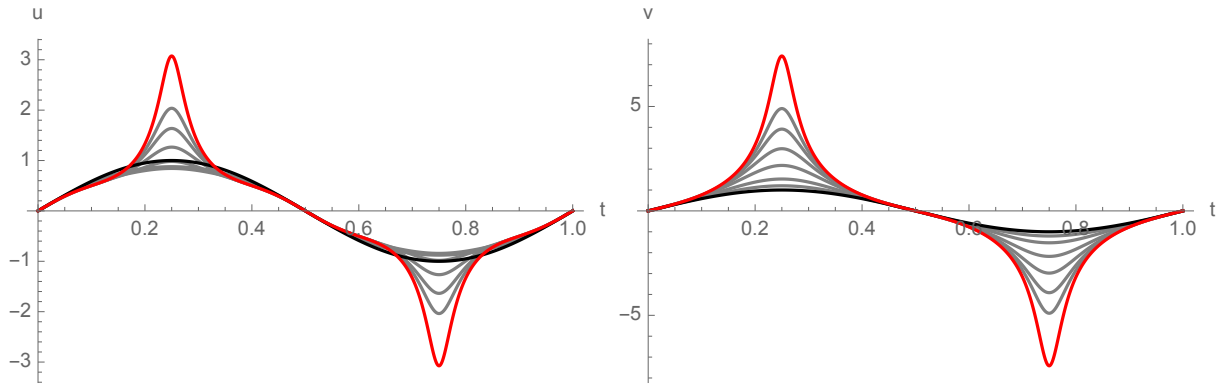
for some positive integer K and coefficients $a_k \in \mathbb{R}$ for $k = 1, \ldots K$. The action will generally have a large set of saddle points, $\bar{x}_j$, corresponding to real and complex solutions of the equation of motion

$$m\ddot{x} = -V'[x]\,. \tag{8.182}$$

Writing the downward flow as $\gamma_\lambda = u_\lambda + iv_\lambda$, with u_λ and v_λ two real-valued paths, the flow equation can be expressed as

$$\partial_\lambda u_\lambda = -\left[\ddot{v}_\lambda + \frac{\delta h_V}{\delta u}\right]\,, \tag{8.183}$$

$$\partial_\lambda v_\lambda = -\left[\ddot{u}_\lambda + \frac{\delta h_V}{\delta v}\right]\,, \tag{8.184}$$



(a) u_λ. (b) v_λ

Figure 8.9: The flow of the $\delta x = (1+i)\sin(2\pi t)$ by the quartic potential $V[x] = x^4$. The fluctuation δx a $\lambda = 0$ (the black curve) flows to $\lambda = 0.08$ (the red curve).

with h_V the h-function corresponding to the potential, *i.e.*,

$$h_V[u+iv] = \mathrm{Re}\left[i\int V[u+iv]\right]\mathrm{d}t\,. \tag{8.185}$$

Near the (non-degenerate) saddle points of the action, $\bar{x}_j$, the flow approaches that of a (inverted) harmonic oscillator. The thimble will consists of a finite number of Fourier modes with a kinetic term contributing positively and an infinite number of Fourier modes contributing negatively to the exponent. The modes for which the kinetic term contributes negatively lie in the lower left or the upper right velocity quadrant, *i.e.*,

$$\{x = u+iv | (\dot{u}(t) > 0 \text{ and } \dot{v}(t) > 0) \text{ or } (\dot{u}(t) < 0 \text{ and } \dot{v}(t) < 0) \text{ for all } t \in [0,1]\}\,. \tag{8.186}$$

Away from the saddle point, some of the modes originally contributing negatively might flow to a sector in which their kinetic term contributes positively. In order for the Lefschetz thimble to enable a continuous-time regularization scheme we need to prove that such a transition will only occur for a finite number of modes.

Consider the steepest descent contour $\mathcal{J}_j$ corresponding to the relevant saddle point $\bar{x}_j$. For potentials, which are dominated by a polynomial it follows from general considerations that there exists a threshold n_t on the frequencies of the paths, such that the flow equation for of all Fourier modes with higher frequencies, $n > n_t$, are dominated by either the second order derivative of the mode and leading power of the polynomial, $a_K x^K$. We will now show that for these high-frequency modes, the sign of the kinetic term is preserved by the downward flow.

When the flow equation is dominated by the second order derivative term, it follows from the analysis of the harmonic oscillator that the sign of the kinetic term is preserved. The flow only amplifies the fluctuation and does not lead to additional maxima and minima in the real and imaginary part of the complex path. When the potential term starts to take over, the path at different times t decouples and the flow becomes ultra-local. In this regime, the Picard-Lefschetz analysis reduces to an infinite set of one-dimensional flows which can be analyzed with the methods developed in chapter 3. The ultra-local evolution also preserves the sign of the kinetic term since it generally stretches the path. See Figure 8.9 for an example of the ultra-local flow of a quartic potential.

The general behavior we expect to proof for the functional integral can be illustrated with the one-dimensional toy-model

$$\int e^{i\left[n^2x^2-\sum_{k=2}^{K}a_kx^k\right]}\mathrm{d}x\,. \tag{8.187}$$

The exponent has a saddle point at the origin, $\bar{x}=0$, and there generally exists a threshold n_t such that the steepest descent thimble corresponding to the saddle point for all $n>n_t$ is solely determined by the quadratic term n^2x^2 and the highest power in the potential term a_Kx^K. For these integrals, the thimble takes a simple form (see Figure 8.10). Near the origin, the steepest descent contour is close to the Gaussian thimble $(1+i)\mathbb{R}$ passing diagonally through the saddle point from the lower left to the upper right quadrant. When the leading term of the potential a_Kx^K takes over, the direction of the flow changes. The steepest descent contour, however, remains in the lower left and upper right quadrant. The Gaussian term always contributes negatively to the exponent along the thimble.

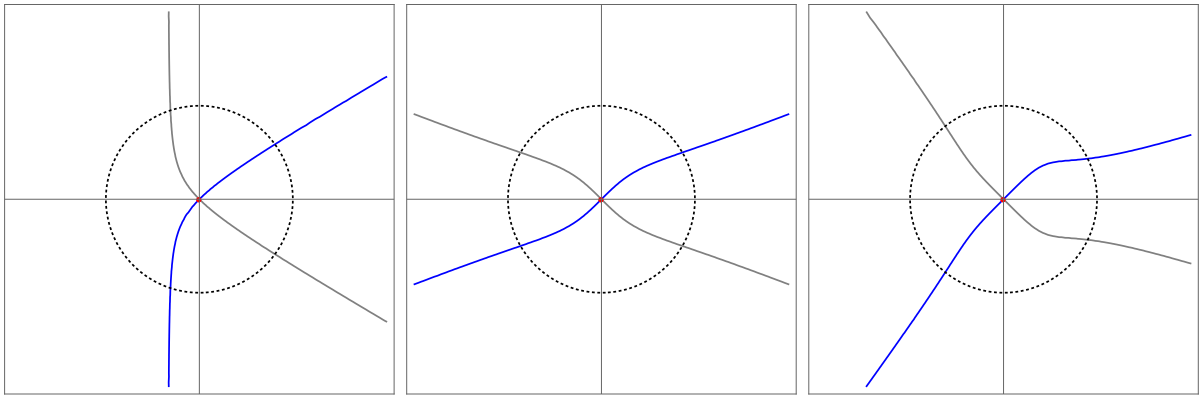
8.3 The path integral measure on real-valued paths

We analytically continue the measure on the thimble $\mu_{\mathcal{J}}$ back to the original integration domain $D=C_{x_1,x_0}([0,1],\mathbb{R})$, by considering the analytic continuation of the exponential function. Consider the class of paths $A_{p,t'}$ in D consisting of the paths passing through point $p\in\mathbb{R}$ at time $0<t'<1$

$$A_{p,t'}=\{x\in C_{x_1,x_0}([0,1],\mathbb{R})|x(t')=p\}\,, \tag{8.188}$$

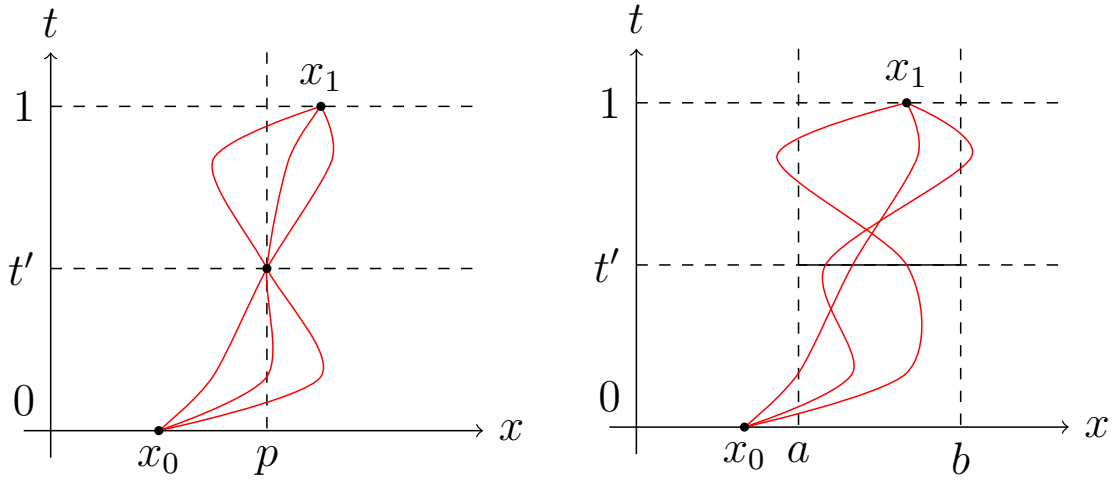
(see left panel of Figure 8.11). We want to define the measure μ_D of the set $A_{p,t'}$. First evaluate the expectation value of the exponential function along the Lefschetz thimble by

$$\left\langle e^{ix(t')q}\right\rangle=\int_{\mathcal{J}}e^{ix(t')q}\mathrm{d}\mu_{\mathcal{J}}(x)\,. \tag{8.189}$$



(a) $K = 3$. (b) $K = 4$. (c) $K = 5$.

Figure 8.10: The downward flow for the exponent $f(x) = i(n^2x^2 - a_K x^K)$ for varying K. The blue line is the thimble, the grey line is the steepest ascent contour and the dotted circle is the regime in which the higher order term takes dominates over the quadratic term.



(a) The set of pinned paths $A_{p,t'}$. (b) The set of pinned paths $A_{[a,b],t'}$.

Figure 8.11: The class of pinned paths $A_{p,t'}$ (left) and $A_{[a,b],t'}$ (right) consisting of the continuous paths from $x(0) = x_0$ to $x(1) = x_1$ passing either through the point p or through the interval $[a, b]$ at time t'.

This is a well-defined function of $q \in \mathbb{R}$, since the analytic continuation of the exponential function does not have singularities in the complex plane. If we now inverse Fourier transform this expectation value with respect to q, we obtain the measure μ_D of the set $A_{p,t'}$, *i.e.*,

$$\mu_D(A_{p,t'}) = \int \frac{\mathrm{d}q}{2\pi} e^{-ipq} \left\langle e^{ix(t')q} \right\rangle = \langle \delta(x(t') - p) \rangle \,. \tag{8.190}$$

The measure of the union of two disjoint sets $A_{p_1,t'} \cup A_{p_2,t'}$, with $p_1 \neq p_2$, is equal to the sum of the individual measures, *i.e.*,

$$\mu_D(A_{p_1,t'} \cup A_{p_2,t'}) = \int \frac{\mathrm{d}q}{2\pi} \left[e^{-ip_1 q} + e^{-ip_2 q} \right] \left\langle e^{ix(t')q} \right\rangle \tag{8.191}$$

$$= \mu_D(A_{p_1,t'}) + \mu_D(A_{p_2,t'}) \,. \tag{8.192}$$

It follows that the measure μ_D is countably additive. The the paths passing through the slit $[a, b]$ at time t' is given by the union of the disjoint sets

$$A_{[a,b],t'} = \cup_{p \in [a,b]} A_{p,t'} \,, \tag{8.193}$$

see right panel of Figure 8.11 for an illustration. The measure of this larger set is given by the double integral

$$\mu_D(A_{[a,b],t'}) = \int_a^b \left[\int \frac{\mathrm{d}q}{2\pi} e^{-ipq} \left\langle e^{ix(t')q} \right\rangle \right] \mathrm{d}p \,. \tag{8.194}$$

Since the Borel σ-algebra of the set of paths in $C_{x_1,x_0}([0,1],\mathbb{R})$ is generated by the set, $A_{[a,b],t'}$, the measure μ_D is a complex σ-measure on the original integration domain $C_{x_1,x_0}([0,1],\mathbb{R})$. We now rigorously define the real-time path integral in terms of the complex measure μ_D, *i.e.*,

$$\int_{x(0)=x_0}^{x(1)=x_1} Q[x] e^{iS} \mathcal{D}x = \int_{C_{x_1,x_0}([0,1],\mathbb{R})} Q[x] \mathrm{d}\mu_D(x) \,. \tag{8.195}$$

Note that this measure differs from both the lattice regularization and the continuous-time regularization scheme used in the Feynman-Kac formula.

8.4 Conclusion

The Feynman path integral is often argued to be the most elegant quantization procedure, since it makes direct contact with the classical theory, and has been successfully and

elegantly applied to gauge theories. It, moreover, plays an important role in the formulation of quantum gravity. Path integrals are nowadays widely used in modern theoretical physics, even though the mathematical definition in terms of a lattice regularization is not mathematically rigorous. We belief that the lack of a rigorous definition, is not merely a pedantic point, as it is important to know the limitations of quantization methods in the investigation of new physical phenomena. Especially when the phenomenon is difficult measure in an experiment and when conventional tricks fail, such as in the case of quantum gravity.

In the decades after Feynman's proposal, there have been several attempts to construct a mathematically rigorous definition by relating the path integral to the Wiener measure of Brownian motion. The most famous proposal is the Feynman-Kac formula for simple quantum mechanical experiments in imaginary-time. This formulation is mathematically satisfactory. However, it is not always clear whether the required Wick rotation $t \mapsto \pm it$ is a sensible procedure. Moreover, when generalized to path integrals over the quantum fields in Euclidean quantum field theory, we find that the integral measure has support in the space of distributions. We believe this to be a serious problem, as the mathematical formulation is in tension with the philosophy of the formalism.

Following Kac's proposal, there have been several attempts to extend this result to real-time path integrals. This, in particular, includes Itô's and Klauder's proposal using a regulator including a second order derivative in time. Unfortunately, it is not clear whether the path integral defined in terms of such a higher order regulator reproduces a Green's function of the Schrödinger equation for general potentials. It is moreover not clear whether such a path integral is compatible with the Heisenberg uncertainty relation.

We here use functional Picard-Lefschetz theory to propose a different continuous-time regularization scheme of the real-time path integral. We deform the original integration domain, consisting of real-valued paths interpolating between two points, into the Lefschetz thimble in the complex plane. We show that the thimble leads to a well defined σ-measure in terms of the Wiener measure for the free non-relativistic particle, the (inverted) harmonic oscillator, and the free relativistic particle. We subsequently extend these results using perturbation theory to more general localized potentials. Finally, we conjecture that this proposal extends to more general real-time path integrals by studying the downward flow of the full theory. We are currently in the process of proving for this conjecture. By deforming this measure back to the original integration domain of continuous real-valued paths, we hope to define a new mathematically rigorous definition for the Feynman path integral over real-valued paths.

We show that the procedure extends to relativists theories, by integrating over the

Schwinger time of the particle. The path integral formulation, defined on the thimble, is the Feynman propagator with the corresponding causal structure. This should be contrasted with canonical quantum field theory, in which the propagator is defined as the inverse of the Klein-Gordon or Dirac equation. In this formulation, there exists a range of possible propagators with different properties, among which the Feynman propagator is usually selected.

8.A Measure theory

A mathematically rigorous understanding of integration theory relies on the study of measures. We here quickly review the main concepts of measure theory relevant to the study of path integrals. For a thorough introduction in modern measure and integration theory see [3, 115, 130, 70, 33].

A measure is a generalization of the notion of volume, *i.e.*, a measure is a non-negative map from a collection of subset satisfying a series of properties. Not every collection of subsets allows for a consistent definition of a measure. It is for this reason important to first define a 'nice' collection of subsets known as an algebra.

Definition 1 (The algebra) *A collection $\mathcal{A}$ of subsets of a set Ω is an algebra if*

1. $\Omega \in \mathcal{A}$
2. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
3. $A, B \in \mathcal{A} \Rightarrow A \cup B \in \mathcal{A}$

where A^c is the complement $\Omega \backslash A$.

A measure is a map from an algebra to the real positive numbers including ∞, with the property that the measure of the empty set vanishes and that the measure of the union of two disjoint sets equals the sum of the measures of the two sets.

Definition 2 (A finitely additive measure) *A finitely additive measure μ on an algebra $\mathcal{A}$ is a mapping $\mu : \mathcal{A} \to [0, \infty]$ which satisfies*

1. $\mu(\emptyset) = 0$

2. $A, B \in \mathcal{A}$ *pairwise disjoint* $\Rightarrow \mu(A \cup B) = \mu(A) + \mu(B)$

These two properties are trivially satisfied for volumes in Euclidean space.

Finitely additive measures are useful in the study of sets. However, the definition of the integral requires the slightly more restricted notion of the σ-algebra and of the countably additive measure. A σ-algebra is an algebra which is closed under the countable union of sets

Definition 3 (The σ-algebra) *A collection $\mathcal{A}$ of subsets of a set Ω is a σ-algebra if*

1. $\Omega \in \mathcal{A}$
2. $A \in \mathcal{A} \Rightarrow A^c \in \mathcal{A}$
3. $A_n \in \mathcal{A}, n \in \mathbb{N} \Rightarrow \cup_{n=1}^{\infty} A_n \in \mathcal{A}$

Every σ-algebra is an algebra, however there exist algebras which are only finitely additive, as demonstrated in the two examples:

Example 1 *Let $\Omega = \mathbb{R}$ and consider the algebra $\mathcal{A}$ consisting of the disjoint unions of the intervals of the form $(-\infty, a], (a, b], (b, \infty), \emptyset, \mathbb{R}$. This is certainly an algebra since $\mathcal{A}$ contains $\mathbb{R}$, the empty set, complements and finite unions. However $\mathcal{A}$ is not countably additive since the union of the intervals $(\frac{i-1}{i}, \frac{i}{i+1}] \in \mathcal{A}$ for $i = 1, 2, \ldots$ given by $(0, 1)$, is not included in the algebra $\mathcal{A}$.*

Example 2 *Let $\Omega = \mathbb{R}$ and consider the algebra $\mathcal{A}$ consisting of the finite sets $\{p_1, p_2, \ldots, p_n\}$ and their complements $\{p_1, p_2, \ldots, p_n\}^c$ with $p_1, \ldots, p_n \in \mathbb{R}$. This is an algebra since the whole set $\mathbb{R}$ is the complement of the empty set, the algebra is closed under complements and finite unions. The algebra is however not a σ-algebra since the countable union of points $\cup_{i=1}^{\infty}\{i\} = \mathbb{N}$ consists of an infinite number of points and is thus not contained in the algebra $\mathcal{A}$.*

A countably additive measure on a σ-algebra is a finitely additive measure for which the measure of the countable union of disjoint sets coincides with the sum of their measures:

Definition 4 (The countably additive measure) *A countably additive, or σ-measure μ on a σ-algebra $\mathcal{A}$ is a mapping $\mu : \mathcal{A} \to [0, \infty]$ which satisfies*

1. $\mu(\emptyset) = 0$
2. $A_n \in \mathcal{A}, n \in \mathbb{N}$, *pairwise disjoint* $\Rightarrow \mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$

This property is also true for volumes in Euclidean space as formalized by the Lebesgue measure on $\mathbb{R}$.

Definition 5 (The Lebesgue measure) *The length of the interval $I = [a, b]$ is given by $l(I) = b - a$. The Lebesgue outer measure μ^* on $E \subset \mathbb{R}$ is defined as the infimum*

$$\mu^*(E) = inf\left\{\sum_{n=1}^{\infty} l(I_n) | (I_n)_{n\in\mathbb{N}} \text{ is a sequence of closed intervals for which } E \subset \cup_{n=1}^{\infty} I_n\right\}$$

That is to say the Lebesgue outer measure of a subset E is the minimal length of the cover of E. The outer measure μ^* induces the Lebesgue σ-algebra $\mathcal{A}$ which is defined to be the collection of sets E satisfying

$$\mu^*(A) = \mu^*(A \cap E) + \mu^*(A \cap E^c) \tag{8.196}$$

for every $A \subset \mathbb{R}$. The Lebesgue outer measure restricted to the Lebesgue σ-algebra is defined to be the Lebesgue measure μ, i.e., $\mu(E) = \mu^*(E)$ if $E \in \mathcal{A}$.

Countably additive measures on σ-algebras suffice for the definition of integrals. However, in the study of real-time path integrals, it is handy to extend the notion of measures to signed and complex measures.

Definition 6 (signed measures) *A map $\mu : \mathcal{A} \to [-\infty, \infty]$ on a σ-algebra $(\Omega, \mathcal{A})$ is a signed σ-measure when it satisfies the properties:*

1. *The map is either bounded from above or below, i.e., either $\mu(E) \neq -\infty$ or $\mu(E) \neq \infty$ for all $E \in \mathcal{A}$,*
2. *The map acting on the empty set vanishes, i.e., $\mu(\emptyset) = 0$,*
3. *The map is countably additive, i.e., $\mu(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$ for all pairwise disjoint $E_i \in \mathcal{A}$.*

We see that a positively valued signed measure is a ordinary σ-measure. The difference of two finite σ-measures is an example of a signed measure. Conversely, given a signed measure, the Hahn decomposition theorem allows us to uniquely write the signed measure as the difference of two σ-measures.

Theorem 3 (Hahn decomposition theorem) *Given a signed measure μ on the σ-algebra $(\Omega, \mathcal{A})$, there exist two σ-measurable subsets P and N of Ω, such that*

1. *$P \cup N = \Omega$ and $P \cap N = \emptyset$,*
2. *For every $E \in \mathcal{A}$ such that $E \subset P$, the measure $\mu(E) \geq 0$,*
3. *For every $E \in \mathcal{A}$ such that $E \subset N$, the measure $\mu(E) \leq 0$.*

It is evident that P is called the positive, while N is called the negative set. The decomposition of Ω into (P, N) is essentially unique, in the sense that if both (P, N) and (P', N') satisfy these three conditions, the sets P, P' and N, N' only differ by a set of measure 0.

Now, given a signed σ-measure μ and the corresponding Hahn decomposition (P, N), we can uniquely write $\mu = \mu_P - \mu_N$ with both μ_P and μ_N ordinary σ-measures. In particular we can define μ_P and μ_N by $\mu_P = \mu(E)$ for all $E \subset P$ and $\mu_P = 0$ for all $E \subset N$ and $\mu_N = -\mu(E)$ for all $E \subset N$ and $\mu_N = 0$ for all $E \subset P$. This is known as the Jordan decomposition. Observe that when a signed measure is secretly positive, the measure is an ordinary measure since $N = \emptyset$ and μ_N is not defined.

Given the signed measures, we define complex measures.

Definition 7 (complex measures) *A complex σ-measure $\mu : \mathcal{A} \to \mathbb{C}$ on a σ-algebra $(\Omega, \mathcal{A})$ satisfies the properties:*

1. *The empty set has vanishing measure $\mu(\emptyset) = 0$*
2. *The measure is countably additive, i.e., $\mu(\cup_{i=1}^{\infty} E_i) = \sum_{i=1}^{\infty} \mu(E_i)$ for all pairwise disjoint $E_i \in \mathcal{A}$.*

It is trivial to show that the complex measures are an extension of the signed-measures. It it moreover easy to observe that the real and imaginary parts of a complex measure are signed measures. We can thus write

$$\mu = \mu_r + i\mu_i \tag{8.197}$$

where μ_r and μ_i are the signed measures defined by $\mu_r(E) = \mathrm{Re}[\mu(E)]$ and $\mu_i(E) = \mathrm{Im}[\mu(E)]$ for all $E \in \mathcal{A}$.

8.B Integration theory

Given a countably additive measure μ on a σ-algebras $\mathcal{A}$ of a set Ω, we construct the integral of measurable functions. The integral over a piecewise constant function

$$f = \sum_{k=1}^{r} \alpha_k \mathbf{1}_{A_k} \,, \tag{8.198}$$

where $\alpha_k \in \mathbb{R}$, $A_k \in \mathcal{A}$ are a pairwise disjoint sets of a measure space, and where $\mathbf{1}_A$ is the Dirac measure ($\mathbf{1}_A(x) = 1$ if $x \in A$ and $\mathbf{1}_A(x) = 0$ if $x \notin A$), is directly defined in terms of the measure

$$\int_\Omega f \,\mathrm{d}\mu = \sum_{k=1}^{r} \alpha_k \mu(A_k) \,. \tag{8.199}$$

By approximating a measurable function $g : \Omega \to \mathbb{R}$ (a function respecting the algebra structure) with series of piecewise constant functions, the integral $\int_\Omega g \,\mathrm{d}\mu$ can be defined using a limiting procedure when the function is integrable, *i.e.*,

$$\int_\Omega |g| \,\mathrm{d}\mu < \infty \,. \tag{8.200}$$

In practice it is common to write the integral as

$$\int_\Omega g(x) \,\mathrm{d}\mu(x) = \int_\Omega g(x) \,\mu(\mathrm{d}x) \,, \tag{8.201}$$

or in the Lebesgue integral over the real line

$$\int_{\mathbb{R}} g(x) \,\mathrm{d}\mu(x) = \int_{\mathbb{R}} g(x) \,\mathrm{d}x \,, \tag{8.202}$$

where we use the notation $\mathrm{d}\mu(x) = \mathrm{d}x$.

Singed and complex measures

Integrals over singed and complex measures on $\mathbb{R}^n$ are defined using the decomposition in terms of σ-measures. Given a singed measure μ with the decomposition $\mu = \mu_P - \mu_N$, we define the integral over an integrable function f as

$$\int f \mathrm{d}\mu = \int f \mathrm{d}\mu_P - \int f \mathrm{d}\mu_N \,. \tag{8.203}$$

When working with a complex measure μ we decompose the measure in terms of the real and imaginary signed measures $\mu = \mu_r + i\mu_i$ and evaluate the integral as

$$\int f \mathrm{d}\mu = \int f \mathrm{d}\mu_r + i \int f \mathrm{d}\mu_i \,. \tag{8.204}$$

Absolutely convergent integrals

Assuming the integral is absolutely convergent, *i.e.*,

$$\int |f| \mathrm{d}\mu < \infty \,, \tag{8.205}$$

we can use the dominated convergence theorem and Fubini's theorem to manipulate limits and integration orders.

Theorem 4 (The dominated convergence theorem) *Given a family of real-valued measurable functions $\{f_n\}_{n=1}^{\infty}$ with a well defined point-wise limit*

$$\lim_{n\to\infty} f_n(x) = f(x) \;\forall\; x \,, \tag{8.206}$$

which is dominated by an integrable function g, i.e., $|f_n(x)| < g(x)$ for all n and for almost all x with $\int |g| \mathrm{d}x < \infty$, the limit of the integral converges to the integral of the limit, i.e.,

$$\lim_{n\to\infty} \int f_n(x) \mathrm{d}\mu(x) = \int f(x) \mathrm{d}\mu(x) \,. \tag{8.207}$$

Theorem 5 (Fubini's theorem) *An absolutely convergent multi-dimensional integrals can be evaluated iteratively. That is to say, for the two-dimensional case*

$$\int f(x,y) \mathrm{d}\mu(x,y) = \int \left[\int f(x,y) \mathrm{d}\mu(x) \right] \mathrm{d}\mu(y) \tag{8.208}$$

$$= \int \left[\int f(x,y) \mathrm{d}\mu(y) \right] \mathrm{d}\mu(x) \tag{8.209}$$

if

$$\int |f(x,y)| \, \mathrm{d}\mu(x,y) < \infty \,. \tag{8.210}$$

For conditionally convergent integrals, the result can depend on the order of integration or the regularization scheme. See for example:

Example 3 *The two-dimensional integral*

$$\iint_{[1,\infty)^2} \frac{x-y}{(x+y)^3}\,\mathrm{d}\mu(x,y)\,, \tag{8.211}$$

is only conditionally convergent. As a consequence, the two orderings of integration give distinct results, i.e.,

$$\int_1^\infty \left[\int_1^\infty \frac{x-y}{(x+y)^3}\mathrm{d}\mu(y)\right]\mathrm{d}\mu(x) = -\frac{1}{2}\,, \tag{8.212}$$

$$\int_1^\infty \left[\int_1^\infty \frac{x-y}{(x+y)^3}\mathrm{d}\mu(x)\right]\mathrm{d}\mu(y) = \frac{1}{2}\,. \tag{8.213}$$

8.C Characteristic functions

A σ-measure on a space Q for which $\mu(Q) = 1$, is known as a probability measure. Given a probability measure, it useful to evaluate the characteristic function C defined by

$$C(t) = \int e^{itx}\,\mathrm{d}\mu(x)\,. \tag{8.214}$$

The characteristic function is sometimes seen as the Fourier transform of the measure. From the definition of the characteristic function, straightforward to proof the four properties

- $C(0) = 1$
- $|C(t)| \leq 1$
- $\sum_{j,k=1}^{N,N} \alpha_j^* \alpha_k C(t_k - t_j) \geq 0$, for a general finite set of complex numbers $\alpha_j \in \mathbb{C}$
- $C(t)$ is continuous

The first property corresponds to the fact that the probability density is normalized. The second property follows from the observation that

$$\left|\int e^{ikx}\mathrm{d}\mu(x)\right| \leq \int \left|e^{ikx}\right|\mathrm{d}\mu(x) = 1\,. \tag{8.215}$$

The third property can be proven from the definition of the characteristic function

$$\sum_{j,k=1}^{N,N} \alpha_j^* \alpha_k C(t_k - t_j) = \left\langle \sum_{j,k=1}^{N,N} \alpha_j^* e^{-it_j x} \alpha_k e^{it_k x} \right\rangle = \left\langle \left| \sum_k \alpha_k e^{it_k x} \right|^2 \right\rangle \geq 0 \,. \tag{8.216}$$

The fourth property can be proven by constructing an upper bound for $|C(t) - C(s)|$ and showing that for every $|C(t) - C(s)| < \epsilon$ we can construct a $\delta > 0$ such that $|t - s| < \delta$. Interestingly enough, the converse is also true. That is to say, given a function C satisfying these four properties, there exists a corresponding probability measure. For finite dimensional measures, this theorem is known as Bochner's theorem. For the infinite dimensional case, the statement is known as the Bochner-Minlos theorem.

8.C.1 Bochner's theorem

Bochner's theorem, states that the four properties of the characteristic function uniquely specify the corresponding probability measures.

Theorem 6 (Bochner's theorem) *Given a function C satisfying these four conditions,*

- $C(0) = 1$
- $|C(t)| \leq 1$
- $\sum_{j,k=1}^{N,N} \alpha_j^* \alpha_k C(t_k - t_j) \geq 0, \quad \alpha_j \in \mathbb{C}$ *and* $\forall$ *positive integers* N
- $C(t)$ *is continuous*

there exists a unique probability measure $\mu(x)$ for which $C(t)$ is the characteristic function

$$C(t) = \int e^{itx} \, \mathrm{d}\mu(x) \,. \tag{8.217}$$

It turns out that the characteristic function is sometimes better behaved than the corresponding probability measure. It is for this reason useful to study measures in terms of characteristic functions. Bochner's theorem applies to measures on finite dimensional spaces. In the next section we present the extension of Bochner's theorem to measures on infinite dimensional spaces.

8.C.2 Bochner-Minlos theorem

The Bochner's theorem specifies four properties which are necessary and sufficient for the existence of a probability measure on finite dimensional spaces. In order for the measure of infinite-dimensional integrals to exist, we need an extension known as Bochner-Minlos theorem which applies to sequences and functions.

Denote a sequence of real numbers by $\{t_1, t_2, t_3, \dots\} = \{t_n\}_{n=1}^{\infty}$ with $t_i \in \mathbb{R}$. In the space of sequences we define a collection of sequences with nice convergence properties known as the **nuclear space** $\mathcal{N}$.

Definition 8 (Nuclear space) *A nuclear space is a set of sequences* $\{t_n\}$ *for which*

$$\lim_{n\to\infty} n^r |t_n| = 0 \tag{8.218}$$

or equivalently

$$\sum_{n=0}^{\infty} n^r |t_n|^2 < \infty \tag{8.219}$$

for all real $r \geq 0$.

In other words, a nuclear space is a set of sequences which fall off faster than any polynomial. An example of an sequence in a nuclear space is given by $t_n = e^{-an}$ for $a > 0$. A series of sequences is denoted by $\{t_n^j\}$. We say that the sequence converges $\lim_{j\to\infty}\{t_n^j\} = \{t_n\}$ if

$$\lim_{j\to\infty} \sum_{n=1}^{\infty} n^r |t_n^j - t_n| = 0 \tag{8.220}$$

for any $r \geq 0$.

Given a nuclear space $\mathcal{N}$ we can define the **dual nuclear space** $\mathcal{N}'$.

Definition 9 (Dual nuclear space) *A dual nuclear space is a set of sequences* $\{x_n\}$ *which are polynomial bounded*

$$|x_n| \leq A + Bn^C \tag{8.221}$$

for some constants A, B *and* C.

For sequences $\{x_n\}$ in the dual space $\mathcal{N}'$ the inner product with an element $\{t_n\} \in \mathcal{N}$ is finite

$$\sum_{n=1}^{\infty} t_n x_n < \infty \,. \tag{8.222}$$

Note that whereas the nuclear space $\mathcal{N}$ consists of sequences which are well behaved and quickly approach 0, the dual nuclear space contains sequences which diverge wildly, but for which the inner product is still well defined.

Given the notion of nuclear and dual nuclear spaces we can state the Bochner-Minlos theorem for sequences

Theorem 7 (Bochner-Minlos theorem for sequences) *Suppose we have a functional $C(\{t_n\})$ depending on sequences $\{t_n\}$ that belong to a nuclear space $\mathcal{N}$, and which satisfied the following conditions*

- $C(\{0\}) = 1$
- $|C(\{t_n\})| \leq 1$
- $\sum_{q,p=1}^{N,N} \alpha_p^* \alpha_q C(\{t_{n(p)} - t_{n(q)}\}) \geq 0, \quad \forall \alpha_i \in \mathbb{C}$ *and* $\forall N < \infty$
- $C(\{t_n\})$ *is continuous (e.g. on* $\mathcal{N}$*)*

then there exits a unique probability measure μ with support on sequences $\{x_n\}$ that belong to the dual $\mathcal{N}'$ of the nuclear space $\mathcal{N}$, for which $C(\{t_n\})$ is the characteristic functional

$$C(\{t_n\}) = \int e^{i \sum_{n=1}^{\infty} t_n x_n} \, \mathrm{d}\mu(\{x_n\}) \,. \tag{8.223}$$

We thus observe that we can use a characteristic function C acting on the well-behaved nuclear space $\mathcal{N}$ to proof the existence of a measure μ on the ill-behaved dual nuclear space. This is the power of the Bochner-Minlos theorem. Note that the measure μ has in general a smaller support than the dual space, *i.e.*, the dual space covers the support of the measure.

We can extend the theorem to function by extending the nuclear and dual nuclear space. Considering a function as a linear combination of an orthogonal set of basis functions. That

is to say, given the basis functions $\{h_n(x)\}$ for which

$$\int h_n(x)h_m(x)\mathrm{d}x = \delta_{nm} \tag{8.224}$$

$$\sum_{n=1}^{\infty} h_n(x)h_n(y) = \delta(x-y) \tag{8.225}$$

the function f belongs to the nuclear space $\mathcal{N}$ if

$$f(x) = \sum_{n=1}^{\infty} t_n h_m(x) \tag{8.226}$$

for $\{t_n\} \in \mathcal{N}$. An example of a valid set of basis functions are the Hermit polynomials. Note that we can go back to the sequence by projecting f onto the basis functions

$$t_n = \int h_n(x)f(x)\mathrm{d}x\,. \tag{8.227}$$

A similar relation holds for the dual nuclear space, *i.e.*, given $\{x_n\} \in \mathcal{N}'$ we can define

$$\phi(x) = \sum_{n=1}^{\infty} x_n h_n(x) \tag{8.228}$$

as part of the dual nuclear space. Note that ϕ is in general a distribution. The distributions in the dual nuclear space are however sufficiently well-behaved that if $f \in \mathcal{N}$ and $\phi \in \mathcal{N}'$ the inner product is well-defined

$$\int f(x)\phi(x)\mathrm{d}x = \sum_{n=1}^{\infty} x_n \int h_n(x)\phi(x)\mathrm{d}x = \sum_{n=1}^{\infty} x_n t_n < \infty\,. \tag{8.229}$$

The inner product $\int f(x)\phi(x)\mathrm{d}x$ with $f \in \mathcal{N}$ and $\phi \in \mathcal{N}'$ is sometimes written as $\phi(f)$.

Using the correspondence between sequences and functions in nuclear and dual nuclear spaces, we can rewrite the Bochner-Minlos theorem for functions, which gives us a measure usable for path integrals.

Theorem 8 (Bochner-Minlos theorem for functions) *If $C\{f\}$ is a functional defined for all f in some nuclear space $\mathcal{N}$ and satisfies the conditions*

- $C(\{0\}) = 1$

- $|C(f)| \leq 1$
- $\sum_{p,q=1}^{N,N} \alpha_p^* \alpha_q C\{f_p - f_q\} \geq 0$ *for all* $\alpha_p \in \mathbb{C}$ *and* $\forall N < \infty$
- $C\{f\}$ *is continuous for all* $f \in \mathcal{N}$

then there exists a uniquely probability measure μ *with support in the dual nuclear space* $\mathcal{N}$*, for which* $C\{f\}$ *is a characteristic functional*

$$C\{f\} = \int e^{i \int f(x)\phi(x)\,\mathrm{d}x} \mathrm{d}\mu(\phi)\,. \tag{8.230}$$

8.D Infinite dimensional measures

Brownian motion is the random motion of particles suspended in a fluid (a liquid or a gas) resulting from their collision with the fast-moving molecules in the fluid. The first modern observation of the phenomenon is often attributed to Robert Brown, *On the particles contained in the pollen of plants, and on the general existence of active molecules in organic and inorganic bodies* (1827). While he studied pollen grains suspended in water under a microscope, he observed minute particles, executing a jittery motion. The phenomenon was subsequently described and modeled by Einstein (1905) [123] and von Smoluchowski (1906) [332] as a means to prove the existence of atoms. The subject was subsequently put on a mathematically rigorous basis by Wiener.

A discrete version of Brownian motion is provided by the Gaussian random walk

$$x_n = x_{n-1} + \eta_n \tag{8.231}$$

where η_n is a set of identically independently Gaussian distributed stochastics with vanishing mean $\langle \eta_n \rangle = 0$ and the correlation function $\langle \eta_n \eta_m \rangle = \sigma^2 \delta_{nm}$. A random walk starting at the origin is after iterations n is located at $x_n = \sum_{i=1}^n \eta_i$. By symmetry, the mean position at iteration n is the initial position $\langle x_n \rangle = 0$. The variance of the random walk at iteration n is given by $\langle x_n^2 \rangle = n\sigma^2$. Observe that the particle is on average a distance $\sqrt{n}\sigma$ separated from the starting point.

The sequence of pushes $\{\eta_i\}_{i=1}^n$ which a particle has undergone after n iterations can be described by the multidimensional Gaussian distribution

$$p(\{\eta_i\}) = \frac{\mathrm{Exp}\left[-\sum_{i=1}^n \frac{\eta_i^2}{2\sigma^2}\right]}{(2\pi\sigma^2)^{n/2}}\,. \tag{8.232}$$

Since the random walk is fully determined by the pushes, and since $\eta_i = x_i - x_{i-1}$, the probability distribution for the random walk $\{x_i\}_{i=1}^n$ is given by

$$p(\{x_i\}) = \frac{\mathrm{Exp}\left[-\sum_{i=1}^n \frac{(x_i - x_{i-1})^2}{2\sigma^2}\right]}{(2\pi\sigma^2)^{n/2}} \,. \tag{8.233}$$

Brownian motion is the continuum limit of the Gaussian random walk. Let the time between the iterations be given by Δt. The time at iteration n is given by $t = n\Delta t$, by which the variance at time t, precisely after a integer number of iterations, can be written as $\langle x^2(t)\rangle = \frac{\sigma^2}{\Delta t} t$. We now take the continuum limit $\Delta t \to 0$ while diminishing the pushes $\sigma \to 0$ such that the variance $\langle x^2(t)\rangle = \frac{\sigma^2}{\Delta t} t = 2Dt$ remains fixed. In this limit the evolution equation reduces to the stochastic differential equation

$$\frac{\mathrm{d}x}{\mathrm{d}t} = \eta(t) \tag{8.234}$$

where η is the white noise stochastic process with zero mean $\langle \eta(t)\rangle = 0$ and correlation function $\langle \eta(t)\eta(t')\rangle = 2D\delta(t-t')$ for all times t and t'. The constant D is often interpreted as the diffusion rate. Note that $\langle x(t)\rangle = 0$ given we started the process at the origin, which can be deduces from symmetry and that $\langle x(t)x(u)\rangle = 2D\min(t,u)$ which can be deduced from the correlation function of white noise. It is common to set the diffusion rate D to $1/2$ to simplify the equations.

The probability density for the path $x(t)$, naively in the path integral sense, takes the form

$$p[x(t)] \propto e^{-\frac{1}{4D}\int \dot{x}^2 \mathrm{d}t} \,. \tag{8.235}$$

This is however not mathematically rigorous as Brownian motion is by construction almost everywhere non-differentiable. A proper definition of the Wiener measure for Brownian motion can be derived using the Bochner-Minlos theorem.

8.D.1 The Wiener measure

As we derived above, standard Brownian motion (starting in the origin with $D = 1/2$) satisfies the four properties

1. $x(0) = 0$, assuming the particle starts from the origin

2. $\langle x(t)\rangle = 0$ by symmetry

3. $\langle x(t)x(u)\rangle = \min(t,u)$ due to the properties of white noise

4. The process is Gaussian, meaning that the trajectory is completely determined by the two-point correlation function.

In fact, it can be proven that these four properties suffice to define Brownian motion without reference to the stochastic differential equation or the white noise source term. These four properties can be summarized in the characteristic function

$$C[s(t)] = \langle e^{i\int s(t)x(t)\mathrm{d}t}\rangle = e^{-\frac{1}{2}\int s(t)s(u)\,\min(t,u)\mathrm{d}t\mathrm{d}u}\,. \tag{8.236}$$

That is to say, the first property is a convention. The second equation can be obtained by evaluating the functional derivative

$$\langle x(t)\rangle = \left.\frac{\delta C[s]}{i\delta s(t)}\right|_{s=0} = 0\,. \tag{8.237}$$

The third property follows from the second order functional derivative

$$\langle x(t)x(u)\rangle = \left.\frac{\delta^2 C[s]}{i\delta s(t) i\delta s(u)}\right|_{s=0} = \min(t,u)\,. \tag{8.238}$$

Finally, since the characteristic function only depends on the two-point correlation function it is definitely Gaussian.

Now, it is clear that the characteristic function C satisfies the conditions of the Bochner-Minlos theorem. It thus follows that there exists a probability measure $\mu_W(x)$ for Brownian motion, known as the Wiener measure, for which C is a characteristic function, *i.e.*,

$$C[s] = \int e^{i\int s(t)x(t)\mathrm{d}t}\,\mathrm{d}\mu_W(x)\,. \tag{8.239}$$

The characteristic function obtained from the naive path integral formulation is given by

$$C[s] = \int e^{i\int s(t)x(t)\mathrm{d}t} e^{-\frac{1}{2}\int \dot{x}^2(t)\mathrm{d}t}\,\mathcal{D}x\,. \tag{8.240}$$

We thus see that whereas the components of $e^{-\frac{1}{2}\int \dot{x}^2(t)\mathrm{d}t}\,\mathcal{D}x$ are ill-defined, they can be interpreted as the Wiener measure $\mathrm{d}\mu_W(x)$.

It is useful to mention two famous properties of Brownian motion. Brownian motion is almost everywhere continuous and almost nowhere differentiable. The trajectory is thus extremely jittery and the trajectory has, in particular, no memory of its velocity (as can be seen from the stochastic differential equation).

So far we studied the standard Wiener measure, which starts at the origin and moves with unit standard deviation. We can generalize the Wiener measure by transforming the Brownian motion to

$$\tilde{x}(t) = \sqrt{\nu} x(t - t_0) + x_0 \,. \tag{8.241}$$

This generalized Brownian motion starts at $\tilde{x}(t_0) = x_0$ and moves with a diffusion rate ν. The corresponding Wiener measure is denoted by $\mu^{\nu}_{x_0}(x)$. The generalized Wiener measure $\mu^{\mu}_{x_0}(x)$ is defined over the set of continuous paths starting at x_0, given by $C_{x_0}([0,T],\mathbb{R})$.

Definition 10 *The set $C_{x_0}([0,T];A)$ consists of all continuous paths $\gamma : [0,T] \to A$ with $T > 0$ and A the target space for which the path starts at x_0, i.e., $\gamma(0) = x_0$.*

We can add an additional requirement by stating that $\tilde{x}(t_1) = x_1$. This is known as a Brownian bridge. The Wiener measure for a Brownian bridge is denoted by $\mu^{\nu}_{x_1,x_0}(x)$ and is defined by

$$\mathrm{d}\mu^{\nu}_{x_1,x_0}(x) = \delta(x(t_1) - x_1)\, \mathrm{d}\mu^{\nu}_{x_0}(x) \,. \tag{8.242}$$

The Wiener measure for Brownian bridges $\mu^{\nu}_{x_1,x_0}$ is defined on the set of continuous paths starting at x_0 and ending at x_1, given by $C_{x_1,x_0}([0,T],\mathbb{R})$.

Finally it should be noted that the expectation value function depending on the position of the particle at a finite number of times, can be written as a finite dimensional integral using Wiener's theorem.

Theorem 9 (Wiener's theorem) *For each vector $x_0 \in \mathbb{R}^d$ and each pair of positive numbers σ and t, there exists a unique measure $\mu^{\sigma}_{x_0}$ on the Borel σ-algebra in $C_{x_0}([0,T];\mathbb{R}^d)$ such that the following conditions hold. For each sequence $0 = t_0 < t_1 < \cdots < t_n \leq T$ of real numbers and each non-negative measurable function f on $(\mathbb{R}^d)^n$, we have*

$$\int_{C_{x_0}([0,T];\mathbb{R}^d)} f(x(t_1), x(t_2), \ldots, x(t_n))\, \mathrm{d}\mu^{\sigma}_{x_0}(x) \tag{8.243}$$

$$= \left[\prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma(t_i - t_{i-1})}}\right] \int_{(\mathbb{R}^d)^n} e^{-\frac{1}{2\sigma}\sum_{i=1}^{n} \frac{(x_i - x_{i-1})^2}{t_i - t_{i-1}}} f(x_1, x_2, \ldots, x_n)\, \mathrm{d}x_1 \ldots \mathrm{d}x_n \,. \tag{8.244}$$

This turns out to be a very useful theorem in the development of the Feynman-Kac formula.

Part II

Classical cosmology

Chapter 9

Caustics, diffraction and oscillatory path integrals for radio astronomy

> A "trick" can be played on Nature by slowing down the light that takes shorter paths: glass of just the right thickness is inserted so that all the paths will take exactly the same time. This causes all of the arrows to point in the same direction, and to produce a whopping final arrow-lots of light! Such a piece of glass made to greatly increase the probability of light getting from a source to a single point is called a focusing lens
>
> Richard Feynman

Interference is one of the most universal phenomena in nature. In classical physics, the linear superposition of sound waves, surface waves, radio waves, light or gravitational waves all exhibit the same characteristic patterns of constructive and destructive interference. Interference is also fundamental to quantum physics. The basic quantum amplitudes describing particles or fields are most elegantly formulated as path integrals – sums over trajectories weighted by the phase factor $e^{i\mathcal{S}/\hbar}$, with $\mathcal{S}$ the action and $\hbar$ Planck's constant. As ubiquitous as interference and interference patterns are, they are generally hard to compute. The oscillatory integrals involved are only conditionally and not absolutely convergent, meaning they converge slowly and artefacts such as dependence on unphysical cutoffs may be hard to avoid. Likewise, if the integrals are performed iteratively, as is often the only practicable method, conditional convergence is in general insufficient to guarantee uniqueness, since the order in which partial integrals are taken can affect the result.

In quantum mechanics, these difficulties run deep. In fact, so far they have thwarted

all efforts to rigorously define nontrivial real-time Feynman path integrals, even in non-relativistic quantum mechanics [150]. The only available existence proofs involve a Wick rotation from real, Lorentzian time to imaginary, Euclidean time, which maps the phase factor to a real Boltzmann weight (for a recent review see, *e.g.*, [223]). Unfortunately, securing mathematical rigour this way comes at a high price: One can neither impose the correct initial or final conditions nor describe the system's dynamics in real time, where observations and experiments actually take place, having instead to rely on analytic continuation from imaginary time. Even when one can analytically continue back to real time, this may only be feasible for certain quantities. such as perturbative S-matrix elements and, even then, it may be very difficult. Furthermore, for many theories of interest, including general relativity and quantum condensed matter models with a "sign problem," *e.g.* the Hubbard model, the Wick rotation trick simply does not work.

This chapter represents a step towards a new, broadly applicable method for defining and computing Lorentzian path integrals. Here, we study the interference of relativistic waves, emitted from coherent sources and propagating through a region in which the refractive index varies in space, *i.e.*, a lens. As we shall show, the quantum mechanical path integral amplitude reduces, in this case, to an ordinary, finite dimensional integral.

The study of optical interference patterns dates back over two centuries, long predating Maxwell's equations, but remains of enduring interest. Starting in the 1970's, Berry, Nye and collaborators studied examples of "diffraction catastrophes" – the characteristic patterns created by diffraction about each of Thom's stable caustic catastrophes, and compared intricate mathematical calculations with beautiful experiments [41, 39, 42, 43, 40]. Recently, the need to accurately and efficiently compute similar patterns has arisen in radioastronomy where bright, coherent sources of radio waves like pulsars and fast radio bursts are being detected in rapidly growing numbers [301, 9, 217]. These objects are beacons lighting up the universe. They will potentially provide a vast new source of information for astrophysics and cosmology. Typically, they are lensed by diffuse astrophysical plasmas intervening along the line of sight. Since plasma lensing is strongest at long wavelengths, this lensing must be modeled in the full, wave optics regime [250, 83, 244]. Although challenging, such modeling will likely be vital to our ability to draw precise inferences from these sources [267].

Motivated by this contemporary need, we shall use astrophysical plasma lensing as our main example. However, as should be clear to the reader, the principles involved are far more broadly relevant. The interference patterns created by astrophysical plasma lenses and observed over astronomical or even cosmological distances are governed by exactly the same physics at play in Young's double slit experiment or X-ray crystallography. This is both a striking example of universality in physics and a reminder of how the universe

increasingly provides us with a powerful laboratory for studying fundamental physics.

Spatial variations in the refractive index of astrophysical plasmas can arise due to turbulence in the interstellar medium or other sources of heating [283, 212]. Pulsar observations have provided examples where plasma lensing amplifies the brightness of a coherent radio source by factors approaching a hundred [269, 244, 45]. It has been pointed out that plasma lensing is likely to play an important role in the phenomenology of Fast Radio Bursts (FRBs) [83, 244]. So-called Extreme Scattering Events (ESEs), where the brightness of radio sources is seen to change by factors of a few, are also thought likely to be due to as yet unexplained plasma lensing [151, 268]. Also gravitational lenses can lead to interference phenomena and amplifications of sources [258]. Recently, there has been growing interest in the idea that coherent gravitational wave pulses and trains, emitted from black hole or neutron star mergers, could be gravitationally lensed and thereby magnified. In this situation it will again be important to go beyond geometric optics and include wave diffraction [87]. In all these examples, when the line of sight between source and observer passes through a caustic of a lens, at a given frequency, the observed intensity may be enhanced leading to a pulse in frequency, time, or both. These situations have mainly been studied for one-dimensional lenses near fold and cusp caustics [267, 131]. Here, we shall explore more complex, two-dimensional examples including the swallowtail, elliptic and hyperbolic umbilic catastrophes which we describe below.

There is already an extensive astrophysical literature on the computation of interference patterns in wave optics [81, 80], but published methods tend to converge slowly [172]. They are expensive to implement and the results are sometimes inconclusive. In this chapter, we shall present faster and more reliable methods. Our approach builds on Picard-Lefschetz theory, a general, exact approach to multidimensional oscillatory integrals based upon saddle point and steepest descent techniques (for an introduction [341]; for applications to quantum cosmology, see Chapters **??** and to relativistic quantum mechanics, see [135]). As we shall show, our methods allow for the fast and reliable computation of even very intricate "diffraction catastrophe" patterns. The calculations of these patterns by Berry, Nye and collaborators were an analytical *tour de force*, but relied heavily on the particularities of Thom's canonical "normal forms" of catastrophes, and the mathematical properties of the related special functions, with each case treated separately. Unfortunately, while the normal forms represent the correct universal forms locally, they diverge at large distances. Hence, they are unrealistic as models for natural lenses. Realistic modeling requires a more versatile method which can be efficiently and straightforwardly implemented numerically. We present just such a method here.

Our method applies uniformly, with modest restrictions, to generic lens models. It is simple to implement numerically and computes interference patterns in polynomial time.

As far as we have been able to check, our results agree perfectly with those aforementioned. The only requirement of our method is that it should be possible to analytically extend the interference phase into the complexified space of the spatial coordinates over which the integral is taken. Such functions embrace a very large class of lens models including, for example, any rational function, and should be more than sufficient for most modeling purposes. For simplicity, we shall not consider phases which possess branch cuts in the space of complexified coordinates. However, there are physical cases of interest where such phases do occur and an extension of our approach to this more general setting is an interesting problem for the future.

Instead of using specific properties of special functions and symmetries, our method exploits Cauchy's theorem to exactly transform an integral of an oscillatory phase factor into a sum of absolutely convergent integrals taken over "Lefschetz thimbles" in the space of complexified coordinates. These "thimble" integrals are fast to compute numerically, requiring only polynomial time. They are insensitive to numerical cutoffs and may be performed iteratively in any order with no change to the result. In this chapter, we demonstrate the efficacy of our method by computing the interference patterns for one- and two-dimensional thin lenses. We study the most observationally accessible catastrophes, both in their "normal forms" and in a set of more realistic, localized lens models where these catastrophes appear. Our one-dimensional numerical code, capable of handling generic one-dimensional lenses, is now publicly available online [1].

A simple example of the type of integral we are interested in is

$$\Psi(\mu,\alpha,\nu) = \left(\frac{\nu}{\pi}\right)^{1/2} \int_{-\infty}^{\infty} \mathrm{d}x\, e^{i\phi(x)\nu}, \quad \text{where} \quad \phi(x) = (x-\mu)^2 + \frac{\alpha}{1+x^2}. \tag{9.1}$$

Here, $\Psi(\mu,\alpha,\nu)$ is the amplitude whose square $|\Psi(\mu,\alpha,\nu)|^2$ gives the intensity of light observed at a position, frequency and lens strength controlled by the parameters μ, ν and α. The control parameter μ is determined by the transverse positions of the observer and the source relative to the lens (See Fig. 9.1 and Eq. (9.13) below). The frequency of the waves is proportional to ν so the spacing of interference fringes shrinks as ν is increased. The eikonal limit is $\nu \to \infty$. Finally, α controls the strength of the lens which, in this example, is taken to have a Lorentzian profile. The integral (9.1) is analytically intractable. However, it is simple to compute numerically, for reasonable values of ν, μ and α, using the methods we shall describe below.

In the eikonal limit of large ν, only real saddle point solutions – real stationary points of the phase $\phi(x)$ – contribute significantly to the amplitude. Each one corresponds to

[1] See https://github.com/jfeldbrugge/Picard_Lefschetz_Integrator.

a particular ray. For $\alpha < 1$ the lens is "weak" and there is only one real solution of $\partial_x \phi(x) = 0$. Hence there is only one contributing ray at each value of μ. For $\alpha > 1$, the lens is "strong:" for a finite range of μ values centred on zero, there are three real solutions of $\partial_x \phi(x) = 0$ hence three contributing rays. Correspondingly, one finds three images of the source in this range of μ. The values of μ bounding this range mark a transition from three contributing rays (*i.e.* three real saddles in the phase) to one. At these values of μ, a maximum and a minimum of $\phi(x)$ merge into a *cubic* stationary point (*i.e.* a point of inflexion), creating the simplest "fold" catastrophe. If we now decrease the strength of the lens α towards unity, the two "fold" catastrophes approach the point $\mu = 0$ where they merge to form a "cusp" catastrophe, in which there is a *quartic* stationary point in the phase $\phi(x)$. Since the phase (viewed as a function of x) is flatter in the vicinity of higher order stationary points, there is less destructive interference. The intensity of light grows more rapidly as ν is increased as compared to the intensity from a quadratic saddle , so that "folds" become increasingly bright compared to the unlensed image and "cusps" become even brighter. While higher order catastrophes are rarer, their brightness makes them easier to detect. This has encouraged the conjecture, yet to be verified [267, 301, 83, 86, 172], that the brightest sources seen may be those which happen to be lensed into high order catastrophes.

In order to emphasize the foundational character of the physics at play and by way of a pedagogical introduction, we show how the standard Fresnel-Kirchhoff integral (see, *e.g.*, Ref. [51], Chapter 8), central to the description of lensing in radioastronomy and in optics[2], can be derived directly from Feynman's path integral for a massless particle propagating through a refractive medium, *i.e.*, one in which the speed of light varies across space. Our main focus in this chapter is on dispersive but non-dissipative lensing, in which the lensing phase factor always has modulus unity. However, the methods we use may equally well be applied to dissipative (lossy) lensing, in which the plasma dispersion relation is complex (for a review of dispersion relations, for example in water or in the ionosphere, see, *e.g.*, [214] Ch. 7). In this more general circumstance, the "phase factor" over which the Fresnel-Kirchhoff integral is taken has a varying modulus.

As an illustration of such a case, as well as to provide a foretaste of the use of our method in describing quantum mechanical interference, in Appendix 9.B we examine Young's famous double-slit experiment. We consider a thin, flat one-dimensional lens which modu-

[2]The integral formula has a fascinating history of successive approximate derivations and subsequent critiques, reviewed in Ref. [51], Chapter 8. Exact solutions of Maxwell's equations (or their scalar version) representing quasi-realistic interference patterns created by diffraction around physical obstacles of various types are still few in number, and are reviewed in Chapter 11 of the same work. It would be interesting to revisit these solutions and, perhaps to find others, using the ideas we develop here.

lates the intensity rather than the phase of the light passing through it. We model the lens with a smooth function which allows very little light through except in two narrow regions comprising the slits. We calculate the resulting interference pattern by deforming the contour onto the relevant Lefschetz thimbles numerically, observing how different real and complex saddle points become relevant and irrelevant, as one moves across the observational screen, through an intricate sequence of Stokes phenomena. Using this smooth lens model, we can also study in detail the emergence of the classical limit as Planck's constant $\hbar$ is taken to zero, so that the de Broglie wavelength becomes small. In this limit we find as expected that only the real, classical saddles contribute and all interference effects disappear.

Recently, Dunne, Unsal and collaborators have been pursuing a very interesting, related program in quantum field theory and quantum mechanics, based upon Euclidean path integrals [29, 35, 120, 37, 38]. See also, the closely related work of [295], and earlier work of [306].

The outline of this chapter is as follows. In Section 9.1 we show how the Fresnel-Kirchhoff integral and Fermat's principle follow from the relativistic path integral for a massless particle, *i.e.*, a spinless photon, moving in a medium with a variable speed of light. In Section 9.2 we discuss the Fresnel-Kirchhoff integral for thin astrophysical lenses, putting the answer into a canonical dimensionless form. We then discuss the intensity in the geometric optics limit, along with the occurrence of critical points and caustics. We introduce catastrophe theory, describing the "normal form" of critical points of increasing complexity and their relation to observable parameters. In Section 9.3 we discuss Picard-Lefschetz theory for a one-dimensional lens – first in the geometric optics limit and then beyond, to include diffraction. We introduce the key concept of "flowing" the integration contour into the complex plane, in order to find the set of relevant Lefschetz thimbles upon which the integral becomes absolutely convergent. We describe a simple and powerful numerical code which implements this idea. In section 9.4 we numerically compute the interference patterns of the seven elementary catastrophes, giving a comprehensive analysis of their "unfoldings." In section 9.5 we turn to localized lens models, which are analytically intractable. In section 9.6 we anticipate possible applications to the study of Fast Radio Bursts, which is an exciting current prospect. Section 9.7 concludes. Appendix 9.A provides some instructive background on the simplest (Gaussian) oscillatory integrals - both one- and two-dimensional, and Appendix 9.B tackles Young's famous double-slit experiment.

9.1 From Feynman to Fermat to Fresnel-Kirchhoff

Imagine a bright source emitting coherent electromagnetic waves which traverse an astrophysical plasma on their way to our telescopes on earth. Let us describe the propagation in terms of the elementary *quanta* of such waves, considered to be relativistic particles. The Feynman path integral over these particle's trajectories in spacetime yields the quantum mechanical amplitude to propagate from the source to any particular location. The square of the amplitude yields the intensity, determining the interference pattern in position and frequency. As we shall see, one or more classical trajectories dominate the amplitude: these dominant trajectories obey Fermat's "principle of least time." For simplicity, we shall ignore polarization effects, taking the elementary quanta to be spinless. We shall furthermore study only the simplest dispersion relation for astrophysical plasmas, valid in the high frequency regime – generalizations to more complex and realistic dispersion relations should be straightforward. Our derivation emphasizes the fundamental nature of the physics involved - as we shall show, the Fresnel-Kirchhoff integral (see, *e.g.*, Ref. [51], Chapter 8, 8.3.3 (28)) follows directly from the Feynman path integral. We hope the reader will enjoy the directness and economy of this approach compared to more standard (and cumbersome) derivations based on Maxwell's equations, or their scalar counterpart.

We start from the dispersion relation in a tenuous plasma (see, *e.g.*, [214] Section 7.9)

$$\omega^2 = k^2c^2 + \omega_p^2(\boldsymbol{x}). \tag{9.2}$$

Here, ω and k are the angular frequency and wavenumber of the waves, c is the speed of light and $\omega_p(\boldsymbol{x})$ is the plasma frequency at position $\boldsymbol{x}$, determined by the local density of electrons, assumed to vary across space on scales much larger than the wavelength of the electromagnetic waves. Notice that (9.2) takes exactly the same form as the dispersion relation for a relativistic particle whose mass varies with spatial position.

The dispersion relation (9.2) yields a phase propagation speed

$$v_p(\boldsymbol{x}) \equiv \frac{\omega}{k} = c\sqrt{1 + \frac{\omega_p^2(\boldsymbol{x})}{k^2c^2}}, \tag{9.3}$$

which is greater than the speed of light. This should be no cause for concern, as the analogy with a massive particle assures us, since information only propagates at the group velocity, $\boldsymbol{v}_g \equiv \boldsymbol{\nabla}_{\boldsymbol{k}}\omega$ whose magnitude $c_g = c^2/c_p$ is always less than the speed of light.

The quanta of these waves may be described as relativistic particles, following parameterized worldlines in spacetime: $x^\mu(\lambda) = (ct(\lambda), \boldsymbol{x}(\lambda))$. Reparameterizations $\lambda \to \tilde{\lambda}(\lambda)$

are generated by a Hamiltonian, and reparameterization invariance corresponds to the constraint that the Hamiltonian vanishes, $\mathcal{H} = 0$. The correct expression for the Hamiltonian $\mathcal{H}$ may be read off from the dispersion relation (9.2), using the correspondence $p_\mu = (p_0, \boldsymbol{p}) \leftrightarrow \hat{p}_\mu = -i\hbar\partial_\mu = \hbar(-\omega/c, \boldsymbol{k})$:

$$\mathcal{H} = -p_0^2\, c^2 + p^2\, c^2 + \hbar^2\, \omega_p^2(\boldsymbol{x}). \tag{9.4}$$

The first order (phase space) action, with the initial and final spacetime locations of the particle held fixed, is:

$$S[x; x^\mu(0), x^\mu(1)] = \int_0^1 d\lambda\, \left(p_0\, \dot{x}^0 + \boldsymbol{p}\cdot\dot{\boldsymbol{x}} - \tau(\lambda)\mathcal{H}\right). \tag{9.5}$$

where dots denote derivatives with respect to λ, taken to run from 0 to 1 as the the particle trajectory runs from the initial spacetime point $x^\mu(0) \equiv (c\, t_i, \boldsymbol{x}_i)$ to the final point $x^\mu(1) \equiv (c\, t_f, \boldsymbol{x}_f)$. The 'einbein' $\tau(\lambda)$ serves as a Lagrange multiplier enforcing the Hamiltonian constraint and ensuring the action is reparameterization invariant (it transforms under reparameterization so that $d\lambda\, \tau(\lambda)$ is invariant). Varying the action with respect to the momenta yields Hamilton's equations for the momenta $p_0\, c^2 = -\dot{x}^0/(2\tau)$ and $\boldsymbol{p}\, c^2 = \dot{\boldsymbol{x}}/(2\tau)$. Varying with respect to τ yields the constraint $\mathcal{H} = 0$. The energy $E = -p_0 c$ is conserved because the action is invariant under constant translations of the time x^0.

In seeking to derive Fermat's principle, we face a conundrum. If the initial and final times t_i and t_f are held fixed, how can the total time $t_f - t_i$ possibly vary? The resolution is that, for a monochromatic beam, we should fix the initial energy E, not the initial time t_i. We cannot fix both because of the time-energy uncertainty relation (which follows from the commutator $[\hat{p}_0, \hat{x}^0] = -i\hbar$). The action appropriate to fixing the initial energy is obtained by adding a boundary term. The latter must be chosen to ensure that the variation of the action is zero when the initial energy and the final time, as well as the initial and final spatial positions, are held fixed and the equations of motion are satisfied. The initial time is then free to vary, which is how Fermat's principle can arise. The required total action is:

$$S\left[\boldsymbol{x}; E, \boldsymbol{x}_i, t_f, \boldsymbol{x}_f\right] = p_0 x^0(0) + \int_0^1 d\lambda\, \left(p_0\, \dot{x}^0 + \boldsymbol{p}\, \dot{\boldsymbol{x}} - \tau(\lambda)\mathcal{H}\right), \tag{9.6}$$

with $\mathcal{H}$ given in (9.4).

Since the action (9.6) is quadratic in the momenta and linear in τ, we can integrate out those variables. At the relevant saddle, we may use Hamilton's equations for the

momenta, and the constraint, to obtain a reduced action expressed purely in terms of reparameterization-invariant quantities:

$$S_r\left[\boldsymbol{x}\right] = -Et_i - \int_{t_i}^{t_f} dt \frac{\hbar^2\omega_p^2(\boldsymbol{x}(t))}{E}. \tag{9.7}$$

Writing $t_i = t_f - \int_{t_i}^{t_f} dt$ where the final time t_f is held fixed, we find, up to an irrelevant constant phase,

$$S_r\left[\boldsymbol{x}\right] = E\int_{t_i}^{t_f} dt\, \left(1 - \frac{\hbar^2\omega_p^2(\boldsymbol{x})}{E^2}\right) = E\int_{\boldsymbol{x}_i}^{\boldsymbol{x}_f} \frac{|d\boldsymbol{x}|}{c}\left(1 - \frac{\hbar^2\omega_p^2(\boldsymbol{x})}{E^2}\right)^{\frac{1}{2}} = \int_{\boldsymbol{x}_i}^{\boldsymbol{x}_f} |d\boldsymbol{x}|\,|\boldsymbol{p}|, \tag{9.8}$$

where, again, we used Hamilton's equations for the momenta and the Hamiltonian constraint. Finally, we express the result in terms of the phase velocity (9.3), obtaining

$$S_r\left[\boldsymbol{x}\right] = E\int_{\boldsymbol{x}_i}^{\boldsymbol{x}_f} \frac{|\boldsymbol{dx}|}{c_p(\boldsymbol{x})}. \tag{9.9}$$

Note that, although the phase velocity $c_p(\mathbf{x})$ appearing here is always greater than the speed of light, nowhere in our derivation does any on-shell particle actually travel faster than light.

The reduced action (9.9) embodies Fermat's principle of least time or, more correctly, the principle that the time taken is stationary on dominant classical trajectories. The path integral over all paths, weighted by $e^{iS_r[\boldsymbol{x}]/\hbar}$, is the Fresnel-Kirchhoff integral we seek.

As an aside, note that one may, equally well, obtain the result (9.9) starting from the square root (Nambu-type) action for a particle with a spatially dependent mass $m(\boldsymbol{x})$, by making use of the correspondence $m(\boldsymbol{x})\,c^2 \leftrightarrow \hbar\omega_p(\boldsymbol{x})$, namely

$$\mathcal{S}\left[\boldsymbol{x}\right] = -\int_{t_i}^{t_f} dt\,\hbar\omega_p(\boldsymbol{x}(t))\left(1 - \frac{\dot{\boldsymbol{x}}(t)^2}{c^2}\right)^{\frac{1}{2}}, \tag{9.10}$$

where the dot now denotes a t derivative. This action is explicitly reparameterization invariant from the start. However, it is the action appropriate to fixing the initial time t_i whereas we need to fix the initial energy E. As before, we must supplement the action (9.10) by a boundary term, which turns out to be $+E(t_f - t_i)$. One can easily check that the identity $\partial S_{cl}/\partial t_i = E$ for Hamilton's principal function S_{cl} implies the total action is stationary, provided the desired boundary conditions and the equations of motion are fulfilled. Using $E = \hbar\omega_p/\left(1 - \dot{\boldsymbol{x}}^2/c^2\right)^{\frac{1}{2}}$, the total action reduces to (9.8) as before.

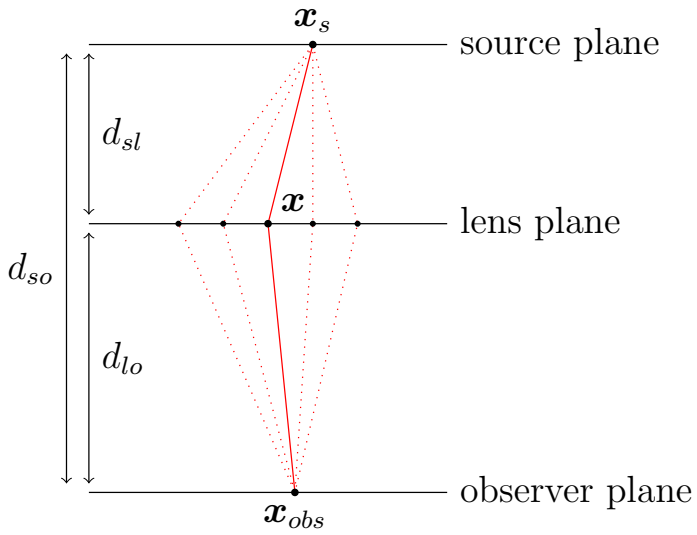


Figure 9.1: The geometry of interfering paths passing through a thin lens.

9.2 Evaluating the Fresnel-Kirchhoff integral

Consider now a radio wave quantum, as described above, traversing an astrophysical plasma from its initial position at the source $\boldsymbol{x}_i = \boldsymbol{x}_s$ to its final position at the observer $\boldsymbol{x}_f = \boldsymbol{x}_{obs}$. For simplicity we assume the plasma takes the form of a thin, flat lens, with the phase velocity $c_p(\boldsymbol{x}) = c$, the speed of light *in vacuo*, everywhere except on the lens (see Fig. 9.1). Let us redefine the spatial coordinates $\boldsymbol{x} \to (\boldsymbol{x}, z)$ to separate out the coordinates in the lens plane $\boldsymbol{x}$ from the normal coordinate z. The real classical paths are piecewise linear, with a possible bend at the lens, and the integral over these paths reduces to an ordinary integral over the lens plane [150]. The path integral amplitude for a (spinless) photon is obtained by integrating over all paths weighted by the phase factor $e^{iS_r[\boldsymbol{x}]/\hbar}$ obtained from (9.9):

$$\Psi(\boldsymbol{x}_{obs}, \boldsymbol{x}_s) = \int d\boldsymbol{x}\, \exp\left[i\omega \int_{\boldsymbol{x}_s}^{\boldsymbol{x}_{obs}} |d\boldsymbol{x}| \frac{n(\boldsymbol{x})}{c}\right]. \tag{9.11}$$

where we replaced E with $\hbar\omega$, ω being the angular frequency of the light, and the phase velocity $c_p(\boldsymbol{x})$ with $c/n(\boldsymbol{x})$ where c is the speed of light *in vacuo* and $n(\boldsymbol{x})$ is the refractive index. Notice that in replacing the energy with the angular frequency, $\hbar$ disappears from the interference phase, which can now be described in purely classical (wave) terms. For this reason, Fresnel and Kirchoff were able to describe interference using what is essentially a path integral, long before quantum mechanics was invented. For an astrophysical plasma,

as mentioned above, at high frequencu we have $n(\boldsymbol{x}) \approx 1 - \omega_p^2(\boldsymbol{x})/\omega^2$ where ω_p is the plasma frequency, given by $\omega_p^2(\boldsymbol{x}) \approx n_e(\boldsymbol{x})e^2/(m_e\epsilon_0)$, with $n_e(\boldsymbol{x})$, e, and m_e respectively the local electron density and the charge and mass of the electron in SI units (see *e.g.* [244]). We explicitly exhibit the $\boldsymbol{x}$-dependence since it governs the structure of the lens.

In the thin lens approximation, variations in the phase arise in part geometrically, from variations in the length of the straight line segments on either side of the lens, and in part from the passage through the lens. The former are straightforwardly computed using the Pythagorean theorem in the approximation that the relative horizontal displacements in Fig. 9.1, $|\mathbf{x}_{obs} - \mathbf{x}|$ and $|\mathbf{x}_s - \mathbf{x}|$ are much smaller than the vertical distances d_{lo} and d_{sl}. The latter are likewise computed approximately, noting that, to lowest order, the paths pass vertically through the lens so we may replace $\int n_e(\boldsymbol{x}, z))dz$ with $\Sigma_e(\boldsymbol{x})$, the electron surface density.

The path integral amplitude then becomes

$$\Psi(\boldsymbol{x}_{obs}, \boldsymbol{x}_s; \nu) \propto \int d\boldsymbol{x}\, \exp\left[i\frac{\omega}{2c}\left(\frac{(\boldsymbol{x}-\boldsymbol{\mu})^2}{\overline{d}} - \frac{\Sigma_e(\boldsymbol{x})e^2}{m_e\epsilon_0\omega^2}\right)\right], \tag{9.12}$$

where $\overline{d} = d_{sl}d_{lo}/(d_{sl}+d_{lo})$ is the reduced distance and $\boldsymbol{\mu} = (\mathbf{x}_s d_{lo} + \mathbf{x}_{obs} d_{sl})/d_{so}$ is a weighted average of the transverse displacements of the source and the observer. Notice that $\boldsymbol{x} - \boldsymbol{\mu}$ depends only on the *relative* displacements of the source, the lens and the observer, so that the answer is independent of the choice of origin for the transverse coordinates.

It is convenient to normalize the amplitude by dividing it by the amplitude obtained with the same geometry but no lens present. We may then write the resulting normalized amplitude as a dimensionless integral. Redefining $\boldsymbol{x} \to a\boldsymbol{x}$, $\boldsymbol{\mu} \to a\boldsymbol{\mu}$ where a is some convenient physical scale associated with the lens, we set $\nu = \omega a^2/(2c\overline{d}) = a^2/(2R_F^2)$ where $R_F = (\lambda\overline{d})^{\frac{1}{2}}$ is the Fresnel scale [43]. Notice that, because lensing alters the *angle* of propagation, the fringe spacing grows with the distance. Hence, it is the Fresnel scale – the geometric mean of the distance and the wavelength – rather than the wavelength which should be compared with the source dimensions to determine whether the interference pattern is observed in the heavily diffracted (low ν) or eikonal (high ν) regime. Finally, we define $\varphi(\mathbf{x}) = -\Sigma_e(\boldsymbol{x})e^2\overline{d}/(m_e\epsilon_0 a^2\omega^2)$ to obtain the normalized, dimensionless amplitude,

$$\Psi(\boldsymbol{\mu}; \nu) = \left(\frac{\nu}{\pi}\right)^{N/2} \int_{\mathbb{R}^N} \mathrm{d}\boldsymbol{x}\, \exp\left[i\phi(\boldsymbol{x}; \boldsymbol{\mu})\nu\right], \quad \text{with} \quad \phi(\boldsymbol{x}; \boldsymbol{\mu}) = (\boldsymbol{x}-\boldsymbol{\mu})^2 + \varphi(\boldsymbol{x}), \tag{9.13}$$

for an N-dimensional lens. Since $\nu \propto \omega$, we see that the eikonal limit is high frequency limit. However, the strength of the lens is controlled by φ which is proportional to ω^{-2}. Therefore the lens becomes stronger at lower frequencies where, of course, diffraction becomes

important. The highest magnifications attained involve a playoff between strong lensing, creating effects like caustics and catastrophes, and diffraction which tends to smear out intensity peaks. Hence, to model the most interesting regime for astrophysical plasma lenses, one must go beyond geometrical optics and include diffractive effects.

The intensity corresponding to the amplitude (9.13) is proportional to the probability for a photon to be detected at $\boldsymbol{\mu}$:

$$I(\boldsymbol{\mu};\nu) \propto |\Psi(\boldsymbol{\mu};\nu)|^2 \,. \tag{9.14}$$

The observed intensity should be normalized to the energy flux received by the detector, at each frequency, integrated over all observed $\boldsymbol{\mu}$. For a more detailed analysis see [57].

Except in special cases, the Fresnel-Kirchhoff integral (9.13) is not possible to evaluate analytically. At large ν (and with the dimensionless form of the lens $\varphi(\boldsymbol{x})$ held fixed) and in the geometric optics limit, one can easily model the intensity, as we shall explain. However, the most interesting regime for astrophysical plasma lenses occurs in the intermediate regime, where focusing and caustic catastrophes generate bright features whose peak intensity is controlled by diffraction. In this wave optics, intermediate-ν regime, there are characteristic patterns in the intensity, controlled by the topological character of the lens. Here, conventional integration techniques typically fail, and it is hard to capture the complex, oscillatory interference pattern numerically. For example, G. Grillo and J. M. Cordes [172] implemented a procedure based on Fourier methods but found this technique to generate numerical artifacts. Here, motivated by our earlier work on Picard-Lefschetz theory, we instead employ analytic continuation and Cauchy's theorem to unambiguously define and to evaluate the relevant oscillatory integrals. We have developed a custom numerical scheme (made available online [3].) which is fast and accurate, and applicable to a generic one dimensional oscillatory integral. A two dimensional version will be made available shortly. A nice feature of our method is that it typically becomes more efficient, *i.e.*, its convergence is improved, as the integrand becomes more oscillatory and difficult to handle via conventional techniques.

9.2.1 The geometric optics limit

In the limit of large ν, the Fresnel-Kirchhoff integral is dominated by real stationary points of the phase function ϕ which, except at special values of μ, are well-approximated by Gaussians. Furthermore, any interference between different stationary points leads to

[3] See https://github.com/jfeldbrugge/Picard_Lefschetz_Integrator.

oscillations in the intensity which, in the limit $\nu \to \infty$, become increasingly rapid. In the geometric optics approximation, one averages these oscillations away. Physically this averaging occurs through the incoherence of any realistic extended source, as we explain later. Although this chapter is devoted to the study of interference phenomena, it proves useful to begin by studying the geometric optics limit.

In the large ν (eikonal) limit, we focus on real critical points of the exponent, *i.e.*, those values of $\boldsymbol{x}$ for which

$$\nabla_{\boldsymbol{x}} \phi(\boldsymbol{x}; \boldsymbol{\mu}) = 0 \,, \tag{9.15}$$

considered as a function of the parameter $\boldsymbol{\mu}$. The critical points are generally smooth complex-valued functions of $\boldsymbol{\mu}$. In the eikonal limit, only the real critical points contribute because contributions from complex saddle points are exponentially suppressed. The critical points can be described in terms of the Lagrangian map $\boldsymbol{\xi} : X \to M$, mapping the points in the base space $\boldsymbol{x} \in X = \mathbb{R}^N$ to points in the parameter space $\boldsymbol{\mu} \in M$ according to the critical point condition

$$\nabla_{\boldsymbol{x}} \phi(\boldsymbol{x}; \boldsymbol{\mu})|_{\boldsymbol{\mu}=\boldsymbol{\xi}(\boldsymbol{x})} = 0 \,. \tag{9.16}$$

The Lagrangian map is determined by the gradient of the phase of the lens:

$$\nabla_{\boldsymbol{x}} \phi(\boldsymbol{x}; \boldsymbol{\mu}) = 2(\boldsymbol{x} - \boldsymbol{\mu}) + \nabla \varphi(\boldsymbol{x}) \quad \Longrightarrow \quad \boldsymbol{\mu} = \boldsymbol{\xi}(\boldsymbol{x}) = \boldsymbol{x} + \frac{1}{2} \nabla \varphi(\boldsymbol{x}) \,. \tag{9.17}$$

The Lagrangian map $\boldsymbol{\xi}$ determines the optical rays, giving a purely geometric description of the lens. Every point $\boldsymbol{x}$ is mapped to a point $\boldsymbol{\mu}$ in the space of observational parameters. In general, a point $\boldsymbol{\mu} \in M$ might be obtained from several points in $\boldsymbol{x} \in X$, *i.e.*, the Lagrangian map can be many-to-one. The regions in $\boldsymbol{\mu}$ where each point is obtained from n points in X are known as n-image regions. In multi-image regions, one adds the intensities due to each of the contributing paths: performing the relevant Gaussian integrals one finds for the normalized intensity

$$I(\boldsymbol{\mu}; \infty) = \sum_{\boldsymbol{x} \in \xi^{-1}(\boldsymbol{\mu})} \frac{2}{|\lambda_1(\boldsymbol{x})| \dots |\lambda_N(\boldsymbol{x})|} \,, \tag{9.18}$$

with $\boldsymbol{\xi}^{-1}$ the pre-image of the Lagrangian map and $\lambda_1, \dots, \lambda_N$ the eigenvalues of the deformation tensor,

$$\mathcal{M}_{ij}(\boldsymbol{x}) = \frac{\partial^2 \phi(\boldsymbol{x}; \boldsymbol{\mu})}{\partial x_i \partial x_j} = 2 \frac{\partial \xi_i(\boldsymbol{x})}{\partial x_j} \,, \tag{9.19}$$

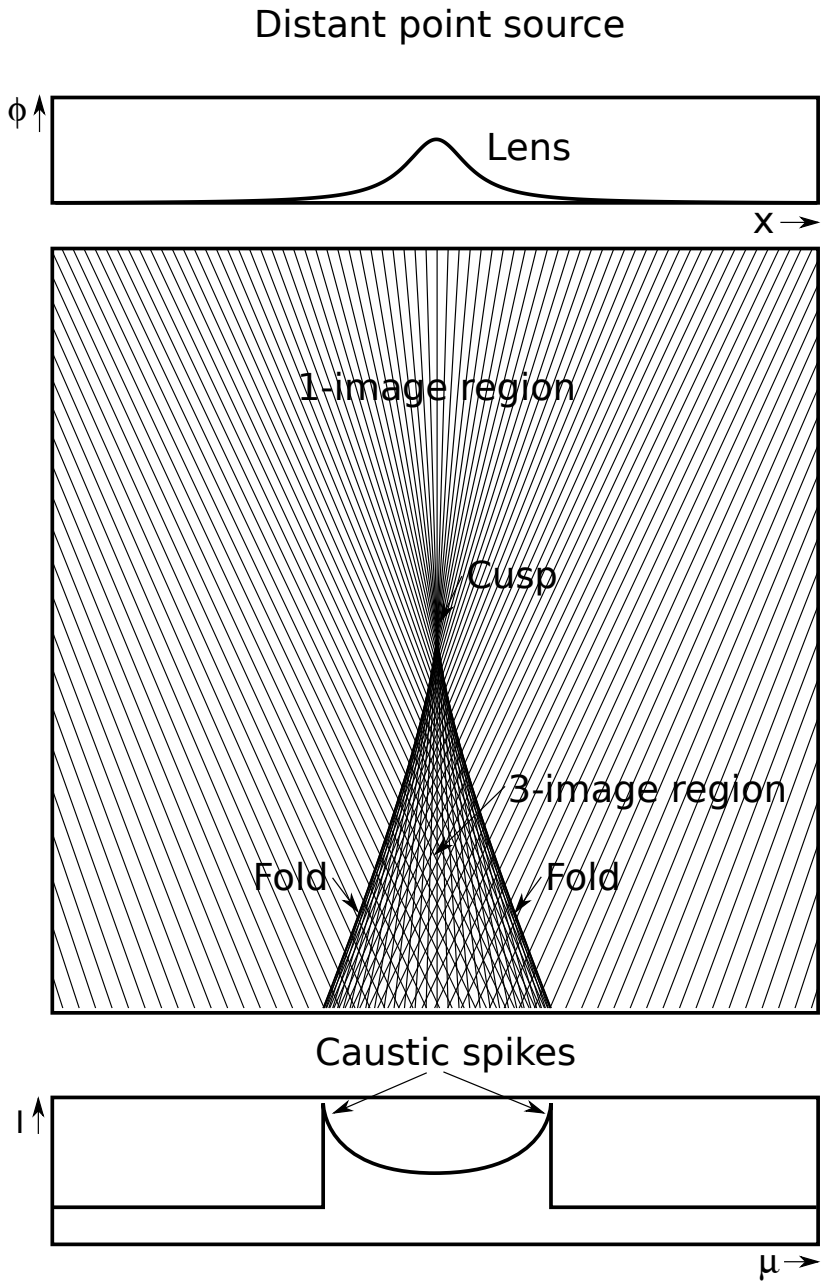


Figure 9.2: The Lagrangian map in geometrical optics. The image consists of two single and one triple image regions separated by a fold caustic at which the normalized intensity spikes.

evaluated at the relevant critical points $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$. Below, we shall study these multi-image regions in detail, at finite ν, where they exhibit intricate interference patterns.

At the boundaries between regions with a different number of images, at least one of the eigenvalue fields λ_i must vanish. At infinite ν this leads to an infinite spike in the normalized intensity map, signalling a caustic. See Fig. 9.2 for an illustration of the Lagrangian map corresponding to a one-dimensional lens with a single- and a triple-image region. The triple-image region is separated from the single-image regions by two fold caustics. At the fold caustic the normalized intensity profile diverges.

Formally, the Lagrangian map $\boldsymbol{\xi}$ forms a caustic at $\boldsymbol{x}_c \in X$ when the deformation tensor becomes singular, *i.e.*, its determinant vanishes. However, the X space is generally not observed. In the space M of observable parameters, the caustic at $\boldsymbol{x}_c$ appears at the point $\boldsymbol{\mu}_c = \xi(\boldsymbol{x}_c)$. For one-dimensional lenses caustics occur at isolated points. For higher-dimensional cases, the determinant of the deformation tensor vanishes on a manifold $X_c = \{\boldsymbol{x} \in X | |\mathcal{M}(\boldsymbol{x})| = 0\}$ which is mapped to a caustic set $M_c = \boldsymbol{\xi}(X_c)$ in the parameter space. Note that the set M_c is generally *not* a manifold, as it includes higher order caustics, such as cusps and swallowtails, at which the variety is non-differentiable and therefore M_c is not locally homeomorphic to Euclidean space. We shall discuss examples of this kind later, but note here that they are exactly the points at which the lensing integral exhibits the most interesting behaviour.

The geometric optics limit is attained in two stages: at short wavelengths, each real stationary point corresponds to a distinct image. As the wavelength is increased, each image itself forms an interference pattern, as illustrated in the Young's double slit experiment examined in Appendix 9.B. The limit of short wavelengths, in which phase coherence is maintained, is often called the eikonal approximation. However, when phase coherence is lost - for example, when the source size becomes larger than the spacing of its fringes, interference effects disappear altogether. This assumption of loss of coherence is implicit in the geometric optics limit. However, objects smaller than the Fresnel scale are still seen to scintillate, as a result of coherent interference effects on unresolved scales. This is reflected in the expression *stars twinkle, planets don't*. Interstellar scintillation typically occurs for sources smaller than about a micro arcsecond, corresponding to the Fresnel angle $\theta_F = \sqrt{\lambda/d}$ (with λ the wavelength and d the distance from the lens) on the sky. This condition is true for most FRBs and pulsars. Interplanetary scintillation due to the solar wind is commonly seen for many compact extragalactic radio sources at low frequencies [79]. In this case, the characteristic Fresnel angle for wavelengths of a few meters and distances of an astronomical unit is a fraction of an arcsecond. Ionospheric scintillation is strongest at the lowest frequencies, and is commonly seen at solar maximum or at equatorial locations near sunrise or sunset [317], and causes loss of lock in GPS. The Fresnel angular scale for

a screen at a distance of 200km at wavelengths of a meter is 8 arc minutes, causing all celestial sources except the sun and the moon to scintillate.

9.2.2 Catastrophe theory

Catastrophe theory is the mathematical classification of stable critical points. Caustics are classified by Lagrangian catastrophe theory [14, 15], which is a special application of the general theory. Given the definition of the Lagrangian map $\boldsymbol{\xi}$, the connection between caustics in optical systems and critical points is not surprising. For one-dimensional functions, the classification consists only of minima and maxima. The local minima and maxima of a one-dimensional function are stable, *i.e.*, the addition of a small perturbation merely leads to a displacement of the critical point. Degenerate critical points are not included, as they are not stable in one dimension. For example, a cubic critical point decomposes into a minimum and a maximum, or no critical point at all, when perturbed.

In the catastrophe theory of higher-dimensional functions, degenerate critical points are included because they are stable. René Thom (1972) proved [311] that the stable critical points with co-dimension[4] K less than or equal to 4 are classified by the seven "elementary catastrophes." These seven singularities suffice to classify the full range of caustics emerging in three-dimensional lenses. Thom named the seven catastrophes: *the fold, cusp, swallowtail, butterfly, and the elliptic, hyperbolic and parabolic umbilic.* The caustics were in the subsequent years connected and labeled by the Coxeter reflection groups (Arnol'd [22, 23]). The theory was subsequently applied to optical interference patterns by Berry and collaborators, and beautiful experiments were performed [43]. For a more recent theoretical investigation of catastrophe theory and caustics with applications to large-scale structure formation see [144]. Here we briefly review catastrophe theory and its application to oscillatory integrals.

Table 9.1 lists the seven "elementary catastrophes" and their unfoldings $\phi(\boldsymbol{x};\boldsymbol{\mu})$. The unfolding $\phi(\boldsymbol{x};\boldsymbol{\mu})$ evaluated at $\boldsymbol{\mu}=0$ is the normal form of the catastrophe, representing the archetypical form of the critical point near $\boldsymbol{x}=0$. We observe that the fold and the cusp respectively correspond to a cubic and quartic critical point of x. The parameter $\boldsymbol{\mu}$ represents the ways in which the caustic can decompose into lower-order caustics. In the case of the fold, we see that a linear perturbation decomposes the fold into a minimum and a maximum for $\mu<0$ and no critical point at all for $\mu>0$. The seven catastrophes

[4]The co-dimension of a caustic is roughly the dimensionality of the singularity. The stable critical points of a n-dimensional function are completely classified by the caustics with co-dimension smaller or equal to n.

Name	Symbol	K	N	$\phi(\boldsymbol{x};\boldsymbol{\mu})$
Maximum/minimum	$A_1^{\pm}$	0	1	$\pm x^2$
Fold	A_2	1	1	$x^3/3+\mu x$
Cusp	A_3	2	1	$x^4/4+\mu_2 x^2/2+\mu_1 x$
Swallowtail	A_4	3	1	$x^5/5+\mu_3 x^3/3+\mu_2 x^2/2+\mu_1 x$
Butterfly	A_5	4	1	$x^6/6+\mu_4 x^4/4+\mu_3 x^3/3+\mu_2 x^2/2+\mu_1 x$
Elliptic umbilic	D_4^-	3	2	$x_1^3-3x_1x_2^2-\mu_3(x_1^2+x_2^2)-\mu_2 x_2-\mu_1 x_1$
Hyperbolic umbilic	D_4^+	3	2	$x_1^3+x_2^3-\mu_3 x_1 x_2-\mu_2 x_2-\mu_1 x_1$
Parabolic umbilic	D_5	4	2	$x_1^4+x_1x_2^2+\mu_4 x_2^2+\mu_3 x_1^2+\mu_2 x_2+\mu_1 x_1$

Table 9.1: The unfoldings of the seven elementary catastrophes with codimension $K \leq 4$, with $\boldsymbol{x}=(x_1,x_2,\ldots,x_N)$ and $\boldsymbol{\mu}=(\mu_1,\mu_2,\ldots,\mu_K)$. The normal forms are defined as the unfolding at parameter $\boldsymbol{\mu}=\mathbf{0}$, *i.e.*, $\phi(\boldsymbol{x};\mathbf{0})$.

Catastrophe	Symbol	I_0	β	σ_j
Fold	A_2	1.584	1/6	$\sigma_1=2/3$
Cusp	A_3	2.092	1/4	$\sigma_1=3/4,\sigma_2=1/2$
Swallowtail	A_4	1.848	3/10	$\sigma_1=4/5,\sigma_2=3/5,\sigma_3=2/5$
Butterfly	A_5	1.991	1/3	$\sigma_1=5/6,\sigma_2=2/3,\sigma_3=1/2,\sigma_4=1/3$
Elliptic umbilic	D_4^-	1.096	1/3	$\sigma_1=2/3,\sigma_2=2/3,\sigma_3=1/3$
Hyperbolic umbilic	D_4^+	0.580	1/3	$\sigma_1=2/3,\sigma_2=2/3,\sigma_3=1/3$
Parabolic umbilic	D_5	2.258	3/8	$\sigma_1=5/8,\sigma_2=3/4,\sigma_3=1/2,\sigma_4=1/4$

Table 9.2: The intensity and fringe separation scaling relations for the catastrophes shown shown in Table 9.1. At large ν the maximum intensity (9.14) is given by $I_0\nu^{2\beta}$ (see the discussion following Eq. (9.21)) and the fringe scaling exponents are defined in (9.22).

belong to two families, classified by their co-rank [5]. The A-family is of co-rank $N=1$, while the D-family is of co-rank $N=2$. Critical points with higher co-rank have a co-dimension higher than 4, and for this reason are not included here. The co-rank N and the co-dimension K characterize the critical point. It generally takes N variables to describe the critical point, and it takes K parameters to describe its unfolding. In more prosaic terms, N is the dimension of the space of $\boldsymbol{x}$'s and K is the dimension of the space of $\boldsymbol{\mu}$'s.

[5] The co-rank is the number vanishing eigenvalues of the Hessian matrix.

For each of the normal forms listed in Table 9.1, the normalized amplitude

$$\Psi(\boldsymbol{\mu};\nu) = \left(\frac{\nu}{\pi}\right)^{N/2} \int e^{i\phi(\boldsymbol{x};\boldsymbol{\mu})\nu} \mathrm{d}\boldsymbol{x}\,, \tag{9.20}$$

forms a caustic at the critical point $\boldsymbol{\mu} = 0$. For a detailed analysis including illustrations of the intensities obtained in each case, see chapter 36 of [312]. As ν is increased to large values, the normalized intensity $I(\boldsymbol{\mu};\nu) = |\Psi(\boldsymbol{\mu};\nu)|^2$ diverges and the scale of the associated diffraction fringes shrinks to zero according to scaling laws which are specific for each catastrophe. At large ν, the maximum of the intensity is attained near $\boldsymbol{\mu} = 0$ as illustrated, for example, by the fold singularity shown in Fig. 9.13. The maximum intensity obeys the following scaling law at large ν:

$$I(\mathbf{0},\nu) = I_0 \nu^{2\beta}\,. \tag{9.21}$$

The constant β, termed the *singularity index* by Arnold (Arnold [23] and Varchenko [320]), is universal, being invariant under diffeomorphisms and depending only on the topological class of the catastrophe. It is given, for each case, in the fourth column of Table 9.2. The scaling with ν is easily seen by examining the corresponding normal form. Setting the unfolding parameter to zero, *i.e.*, $\boldsymbol{\mu} = \mathbf{0}$, in the phases listed in Table 9.1, one can render the phase of the integrand independent of ν by rescaling the integration variables $\boldsymbol{x}$. For example, for A_2 we set $x = \nu^{-\frac{1}{3}} y$. Taking into account the ν-dependence arising from the Jacobian in the integration measure as well from the prefactor in (9.20), one infers that the amplitude at the caustic scales as $\nu^{\frac{1}{6}}$ for A_2 and hence that $\beta = \frac{1}{6}$ for this case. For the two dimensional lenses, one has to rescale both x_1 and x_2 in order to remove ν from the exponent but the argument is otherwise the same.

For each of the normal forms of the phase listed in Table 9.1, one may also analytically compute the constant I_0, and we provide its numerical value in Table 9.2. When considering a class of lens models for modeling purposes (such as the localised models we consider later), it may be helpful to notice an additional scaling property. At large ν, the amplitude is determined by the form of the phase near the critical point. Indeed, this is how universality arises. For any lens model which includes a given catastrophe, the leading terms in the Taylor expansion of the phase about the associated critical point will, after coordinate redefinitions, take the form of one of the "normal forms" listed in Table 9.1. Coordinate rescalings are one of the simplest such transformations, which have a simple effect on any model and on its Taylor expansion about any critical point. One can easily derive the scaling behavior of the intensity under such transformations of the lens model. For an A_n catastrophe, for example, we may consider a set of lens models whose phase $\phi(x,0) \sim \frac{a}{n+1} x^{n+1}$ near $x = 0$, with a a constant. By rescaling x we can remove the a dependence

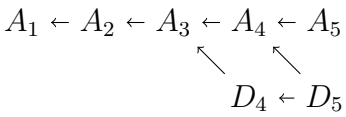


Figure 9.3: The unfolding diagram of the seven elementary catastrophes.

from the phase and hence infer that the intensity scales as $a^{-\frac{2}{n+1}}$. Physically, decreasing a means increasing the size of the lens, so it makes sense that the corresponding intensity grows. For a type D_n catastrophes we may likewise have a phase in the form $ax_1^{n+1} \pm bx_1x_2^2$. Again, the a and b dependence can be removed from the phase by rescaling x_1 and x_2. Hence one can infer that the intensity scales as $a^{-\frac{1}{n-1}}b^{-1}$. (As an aside, note that for D_4^+ the normal form used here differs, from that given in Table 9.1. The two forms may be shown to be equivalent under a linear coordinate transformation $x_i' = A_{ij}x_j$, $i, j = 1, 2$, and hence are in the same equivalence class. In the rest of this chapter we always use the normal forms listed in Table 9.1.)

The pattern of fringes may likewise be shown to scale as

$$\Psi(\boldsymbol{\mu}; \nu) = \nu^{\beta}\Psi\left((\nu^{\sigma_1}\mu_1, \ldots, \nu^{\sigma_K}\mu_K), \nu\right) , \tag{9.22}$$

where the fringe exponents σ_i, defined by Berry [39], are also listed in Table 9.2. The sum of the fringe exponents, $\gamma = \sum_{i=1}^{K} \sigma_i$, represents the scaling exponent for the K-dimensional hypervolume of the diffraction pattern known as the *fringe index*. All of these exponents are invariant under diffeomorphisms, making them topological features.

While these catastrophes provide an exhaustive list, the precise forms in Table 9.1 are unlikely to arise, since apart from $A_1^{\pm}$, they are completely de-localized, with the strength of the lens diverging away from the critical point. Later in this chapter we shall consider more realistic, localized lenses which generate catastrophes within them of the listed form. In the vicinity of such a catastrophe, one can expect the behaviors indicated in Table 9.2 to hold. Note, however, that to compute the maximum intensity at such a catastrophe, one must first redefine the coordinate $\boldsymbol{x}$ to put the exponent locally into the normal form of the catastrophe listed in Table 9.1 (this is guaranteed to be possible by the theorems mentioned above), and take into account the associated Jacobian factor when evaluating the integral.

The seven elementary catastrophes listed above form an intricate hierarchy which unfold under perturbations according to the unfolding diagram (see Fig. 9.3). As we saw before, the fold caustic (A_2) splits under a small perturbation into a maximum and a minimum each corresponding to an A_1. Analogously, butterfly caustic (A_5) unfolds into a swallowtail

caustic (A_4), which in its turn unfolds into the cusp (A_3) and the fold caustic (A_2). The parabolic umbilic caustic (D_5) has a more intricate structure as it can unfold into both the elliptic (D_4^-), hyperbolic umbilic (D_4^+) and the swallowtail caustic (A_4). The elliptic (D_4^-) and hyperbolic umbilic caustic (D_4^+) always unfold into cusp caustics (A_3).

9.3 Example: a one-dimensional lens

Picard-Lefschetz theory, as described for example in [139], enables us to deform the real integration domain $\mathbb{R}^N$ of the Fresnel-Kirchhoff integral (9.13) into a set of Lefschetz thimbles, *i.e.*, steepest descent contours, $\mathcal{J}_i \subset \mathbb{C}^N$ each corresponding to a relevant saddle point $\bar{x}_i$,

$$\Psi(\boldsymbol{\mu};\nu) = \left(\frac{\nu}{\pi}\right)^{N/2} \sum_i \int_{\mathcal{J}_i} e^{i\phi(\boldsymbol{x};\boldsymbol{\mu})\nu} \mathrm{d}\boldsymbol{x}\,. \tag{9.23}$$

The exponent $i\phi(\boldsymbol{x};\boldsymbol{\mu})\nu$ evaluated along a steepest descent contour $\mathcal{J}_i$ has a constant imaginary part while its real part $h = \mathrm{Re}[i\phi(\boldsymbol{x};\boldsymbol{\mu})\nu]$ is monotonically decreasing. As a consequence, the conditionally convergent oscillatory integral is transformed into a sum of convex integrals. This is remarkable, as the originally conditionally convergent integral is generally sensitive to regularization and also, if the integral is performed iteratively, to the order in which the partial integrals are taken (see Appendix 9.A for an instructive example). The integrals over Lefschetz thimbles have none of these ambiguities. It is for this reason that we will interpret the integral over the sum of Lefschetz thimbles as the *definition* of the integral taken over the real integration domain. Once we have identified the correct set of thimbles, we can use conventional numerical methods to evaluate the integral on each thimble.

We shall describe two distinct methods to obtain the sum of Lefschetz thimbles corresponding to the Fresnel-Kirchhoff integral for a one-dimensional lens. In the first method, we follow the techniques explained, for example, in Ref. [139]. We start by computing all the saddle points of the exponent $i\phi(\boldsymbol{x};\boldsymbol{\mu})\nu$ and their corresponding steepest descent and ascent contours . We subsequently study the intersections of the steepest ascent contours with the original integration domain to find the relevant saddle points and associated steepest descent contours $\mathcal{J}_i$. This method is well suited to the Picard-Lefschetz analysis of one-dimensional integrals, for which we can plot the steepest descent and ascent contours in the complex plane $\mathbb{C}$.

In the second method, we instead flow the real integration domain along the downward flow of the real part of the exponent, $h = \mathrm{Re}[i\phi(\boldsymbol{x};\boldsymbol{\mu})\nu]$. We show that this downward flow

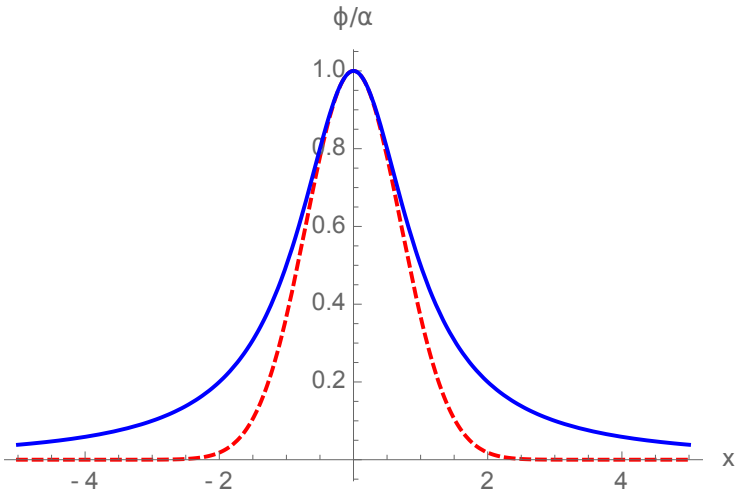


Figure 9.4: A comparison of the Gaussian lens (dashed line) and a corresponding rational lens model (solid line).

terminates on the correct sum of Lefschetz thimbles $\sum_i \mathcal{J}_i$. The relevant saddle points are given by the local maxima of h restricted to this thimble sum $\mathcal{J}$. Note that this second scheme is completely determined by the gradient of the h with respect to the real and imaginary parts of the complexified coordinates $\boldsymbol{x}$. We do not need to find all the saddle points nor evaluate the corresponding steepest ascent and descent contours. We moreover do not need to study the intersections of the steepest ascent contours with the original integration domain. Any Stokes transitions are automatically taken care of. This method is furthermore ideally suited to higher-dimensional oscillatory integrals, where the steepest ascent and descent contours are expensive to evaluate and the intersections are computationally difficult to find.

9.3.1 Geometric optics approximation

In the introduction to this chapter, we discussed a one-parameter family of one-dimensional localized lenses

$$\varphi(x) = \frac{\alpha}{1+x^2}\,, \tag{9.24}$$

with $\alpha \in \mathbb{R}$. For plasma lenses, the parameter α follows the dispersion relation $\alpha \propto \omega^{-2}$ with ω the angular frequency of the source. The longer the wavelength, the stronger the lens. We restrict our analysis to rational lenses for two reasons:

1. their analytic continuation into the complex x-plane is holomorphic and consists of only a finite number of poles,

2. the phase ϕ has only a finite number of saddle points and corresponding steepest-descent contours.

The lens (9.24) is a rational approximation to the Gaussian lens

$$\varphi(x) = \alpha e^{-x^2}\,, \tag{9.25}$$

which has an infinite number of saddle points in the complex plane (as well as an essential singularity at infinity). See Fig. 9.4 for a comparison. It is a wonderful fact that many real-valued functions with intricate structure in the complex plane, can be well-approximated with a Padé approximation, whose analytic continuation possesses only a finite number of poles.

As we derived in Section 9.2.1, the Lagrangian map ξ of the rational lens φ,

$$\xi(x) = x - \frac{\alpha x}{(1+x^2)^2}\,, \tag{9.26}$$

forms caustics at the real roots of the second order derivative of the exponent

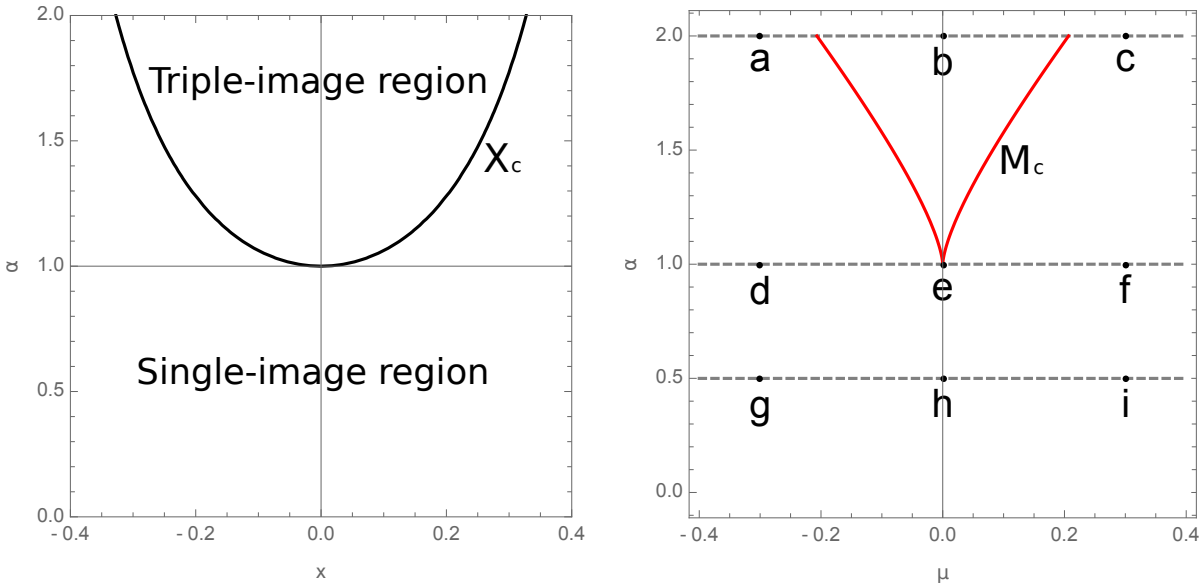
$$\frac{\partial^2 \phi(x)}{\partial x^2} = 2\frac{\partial \xi(x)}{\partial x} = 2 + 2\alpha \frac{3x^2-1}{(1+x^2)^3} = 0\,. \tag{9.27}$$

See Fig. 9.5 for the caustic surface in the x-α and the μ-α planes. For $\alpha < 1$ no such real root exists. The lensed image consists of a single-image region. For $\alpha = 1$ there is a single real-valued root at $x_c = 0$ with multiplicity two and the corresponding point $\mu_c = \xi(x_c) = 0$ in the parameter space M. In the μ-α plane, this point is non-differentiable on the caustic set. This is an example of a cusp caustic. For $\alpha > 1$ there are two symmetric real roots. These are examples of fold caustics. For further reference, for $\alpha = 2$, the two caustics are located at

$$X_c = \{-0.327334\ldots, 0.327334\ldots\}\,, \tag{9.28}$$

in the base space $X = \mathbb{R}$. In the parameter space M the caustic appears at

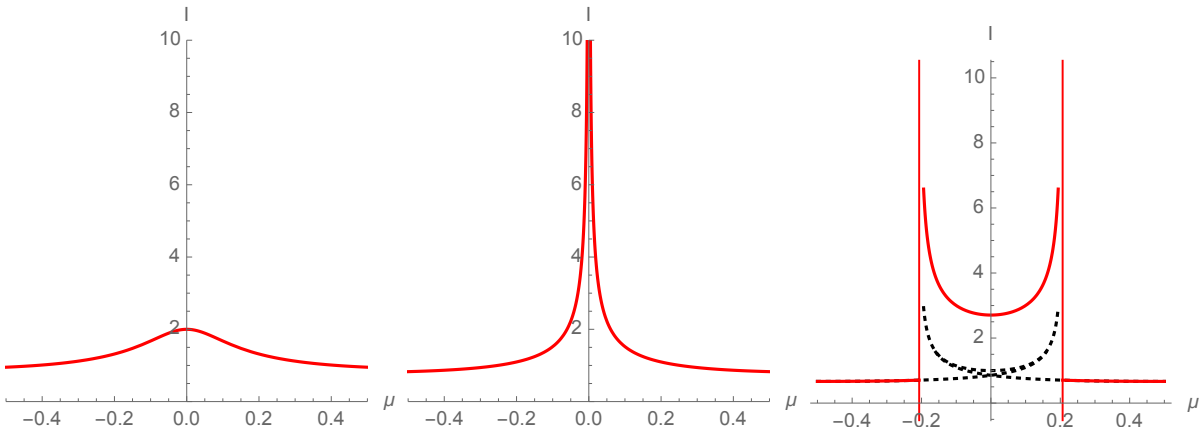
$$M_c = \xi(X_c) = \{+0.206751\ldots, -0.206751\ldots\}\,. \tag{9.29}$$



(a) Caustic in the x-α plane.

(b) Caustic in the μ-α plane.

Figure 9.5: The caustics of the one-dimensional rational lens in the x-α and the μ-α plane. The points in the *left panel* correspond to the Picard-Lefschetz diagrams in Fig. 9.7. The lines correspond to the panels in Fig. 9.11.



(a) $\alpha = 1/2$

(b) $\alpha = 1$

(c) $\alpha = 2$

Figure 9.6: The normalized intensity in geometric optics, $I(\mu;\infty)$, on the three horizontal lines shown in Fig. 9.5, is plotted as a function of μ (solid curves). The contributions from the three distinct images are shown by the dotted curves in (c).

The relative normalized intensity of the lens in the geometric optics limit (see Section 9.2.1) is plotted in Fig. 9.6. For $\alpha = 1/2$, the lens does not form a caustic. The normalized intensity map is finite. For $\alpha = 1$, we see a cusp caustic at $\mu = 0$. For $\alpha = 2$, we observe two fold caustics at M_c enclosing a triple-image region. The black curves in the triple-image region are the three contributions corresponding to the three images. The solid curve is the sum over the multi-image regions depicted by the dashed curves.

9.3.2 Finding the thimbles

We now turn to evaluating the full expression (9.23). First, we need to determine which Lefschetz thimbles contribute. We shall describe two distinct methods, the second of which is more efficient for numerical purposes.

Method 1: following steepest ascent contours

The exponent

$$i\phi(x;\mu)\nu = i\left[(x-\mu)^2 + \frac{\alpha}{1+x^2}\right]\nu \tag{9.30}$$

is imaginary for real μ and x. Its analytic continuation has two poles at $x = \pm i$, and five saddle points in the complex x-plane, satisfying

$$\frac{\partial\phi(x;\mu)}{\partial x} = 2(x-\mu) - \frac{2\alpha x}{(1+x^2)^2} = 0\,. \tag{9.31}$$

For the Picard-Lefschetz analysis we start by writing the analytic continuation of the exponent in terms of its real and imaginary part

$$i\phi(\boldsymbol{x};\boldsymbol{\mu})\nu = h(\boldsymbol{u}+i\boldsymbol{v};\mu) + iH(\boldsymbol{u}+i\boldsymbol{v};\mu)\,, \tag{9.32}$$

with the complex expansion $\boldsymbol{x} = \boldsymbol{u} + i\boldsymbol{v}$ and the real-valued functions h, H. For generality, we describe the flow of the integration contour in N dimensions. The real part h is, in the Picard-Lefschetz analysis, known as the h-function. The downward flow of the h-function $\gamma_\lambda : \mathbb{C}^N \to \mathbb{C}^N$ is defined by

$$\frac{\partial\gamma_\lambda(\boldsymbol{z})}{\partial\lambda} = -\nabla_{\boldsymbol{u}+i\boldsymbol{v}}h[\gamma_\lambda(\boldsymbol{z})] \tag{9.33}$$

with the boundary condition $\gamma_0(\boldsymbol{z}) = \boldsymbol{z} \in \mathbb{C}^N$, the parameter λ in a subset of $\mathbb{R}$, and the complex gradient defined as

$$\nabla_{\boldsymbol{u}+i\boldsymbol{v}} h = \nabla_{\boldsymbol{u}} h + i \nabla_{\boldsymbol{v}} h \,. \tag{9.34}$$

Note that in defining the gradient, we have assumed a corresponding metric on the space $\mathbb{C}^N$. In this chapter we will always associate $\mathbb{C}^N$ with $\mathbb{R}^{2N}$ and us the corresponding Euclidean metric. We are of course free to consider different metrics. Given the saddle points we can compute the steepest ascent and descent contours and intersect the ascent contours with the real axis, to obtain the relevant saddle points.

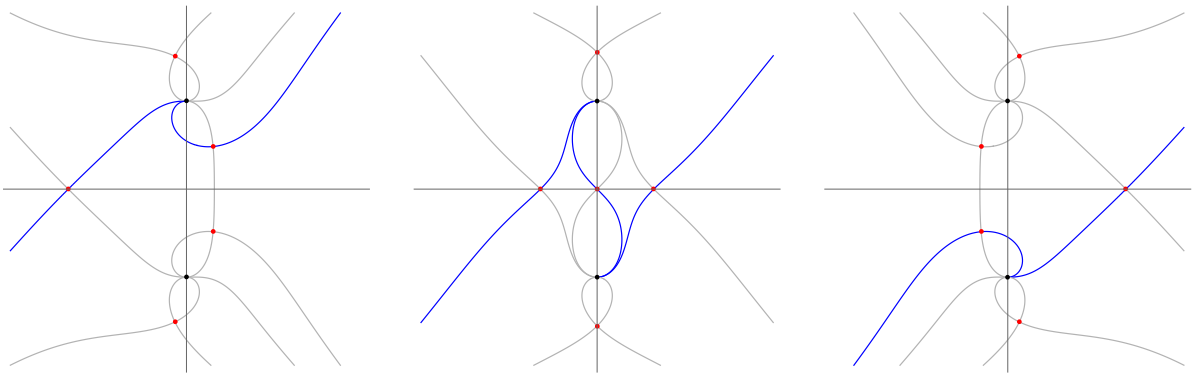
Depending on μ and α either one or three of the saddle points are real-valued. The lens thus has both single- and triple-image regions. See Fig. 9.7 for the five saddle points $\bar{x}_i$ and the corresponding steepest ascent and descent contours. By intersecting the steepest ascent contours with the real line, we obtain the Lefschetz thimble (the heavy solid lines). The thimbles run from $x = -\infty$ to $x = +\infty$, while passing through the poles at $x = \pm i$.

From the caustic structure in Fig. 9.5 we can distinguish three regimes:

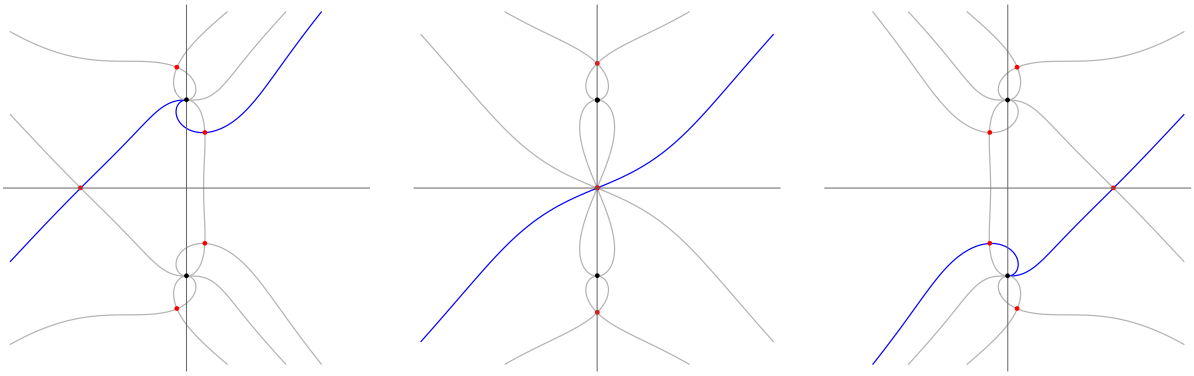
- In the regime $\alpha < 1$, the lens forms a single image. The corresponding Picard-Lefschetz analysis yields a single real-valued saddle point. For large $|\mu|$ there is, in addition, a relevant complex saddle point. When $|\mu|$ decreases to 0, the complex saddle point becomes irrelevant due to a Stokes transition. This phenomenon is discussed in detail in the next section. For $\mu = 0$, only the real saddle point is relevant. Note that the thimble can for all μ be deformed to the original integration domain $\mathbb{R}$. See the lower panels of Fig. 9.5.

- For $\alpha = 1$, the lens contains a cusp caustic at $\mu_c = 0$. For $\mu \neq \mu_c$, the Picard-Lefschetz analysis is similar to the $\alpha < 1$ regime. The thimble passes through one real-valued and one complex-valued saddle point. At the caustic $\mu = \mu_c$, three non-degenerate saddle points merge forming a degenerate saddle point. This is the signature of the cusp caustic, whose normal form is the quartic function x^4. See the middle panels of Fig. 9.5.

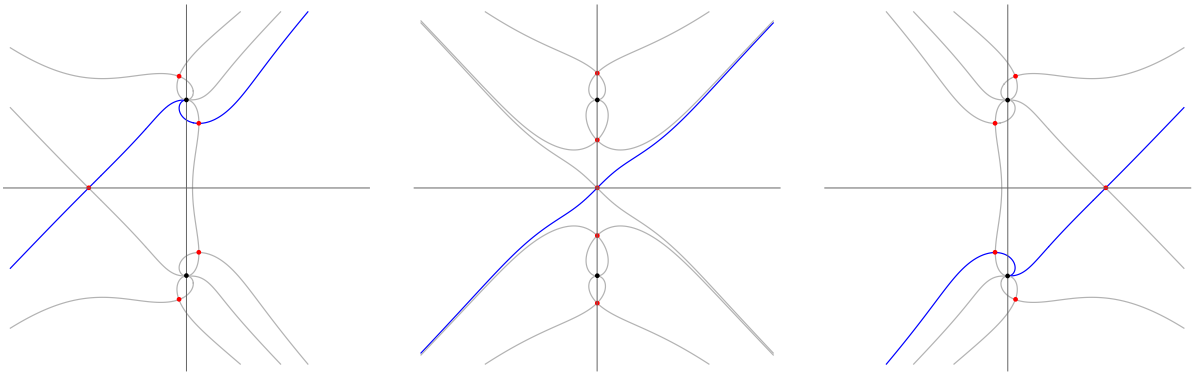
- In the regime $\alpha > 1$, the Picard-Lefschetz analysis splits into three intervals. In the single-image region, *i.e.*, μ in $(-\infty, -\mu_c)$ or (μ_c, ∞), the Picard-Lefschetz analysis consists of two relevant saddle points; one real and one complex. At the caustic, the complex saddle point approaches the real line and merges with its complex conjugate saddle point. This is the signature of the fold caustic. In the triple-image region, *i.e.*, $\mu \in (-\mu_c, \mu_c)$, the analysis consists of three real-valued relevant saddle points. See the upper panels of Fig. 9.5.



(a) $\alpha = 2, \mu < -\mu_c$ (b) $\alpha = 2, -\mu_c < \mu < \mu_c$ (c) $\alpha = 2, \mu > \mu_c$



(d) $\alpha = 1, \mu < -\mu_c$ (e) $\alpha = 1, \mu = \mu_c$ (f) $\alpha = 1, \mu > \mu_c$



(g) $\alpha = 1/2, \mu < 0$ (h) $\alpha = 1/2, \mu = 0$ (i) $\alpha = 1/2, \mu > 0$

Figure 9.7: The Picard-Lefschetz thimbles for $\alpha = 1/2, 1, 2$ as a function of μ. Paths of steepest descent and ascent are shown as curves, with the relevant thimbles shown heavier. Saddle points and poles of the phase are shown as points.

Method 2: flowing the integration domain

We can alternatively obtain the Lefschetz thimble $\mathcal{J}$ by flowing the original integration domain $\mathbb{R}$ along the downward flow of the real part h.

Given the downward flow γ_λ for general points $z \in \mathbb{C}$, we flow the original integration domain X to

$$X_\lambda = \gamma_\lambda(X) \subset \mathbb{C}\,. \tag{9.35}$$

The steepest descent contours $\mathcal{J}_i$ corresponding to the saddle points $\bar{x}_i$ are the fixed points of the flow, *i.e.*,

$$\gamma_\lambda(\mathcal{J}_i) = \mathcal{J}_i \tag{9.36}$$

for all λ. When the h-function has saddle point in the complex plane, it follows from Morse-Smale theory [256, 303] that the flowed contour X_λ will converge to a set of steepest descent contours $\mathcal{J}_i$ as $\lambda \to \infty$. Since X_λ is a continuous deformation of the original integration domain X, it follows that X_λ converges to the Lefschetz contour, *i.e.*,

$$\lim_{\lambda\to\infty} X_\lambda = \mathcal{J}\,. \tag{9.37}$$

When we perform the flow of the original integration domain as a function of the parameter μ, we obtain a family of thimbles. The thimble generally changes smoothly as a function of μ. There are however two ways in which the Picard-Lefschetz structure of the integral can abruptly change its geometry:

1. If for some $\boldsymbol{\mu}$, a few non-degenerate saddle points merge to form a higher order saddle point, the number of relevant critical points will change. At these points, the integral $\Psi(\boldsymbol{\mu};\nu)$ forms a caustic. This phenomenon can be described by catastrophe theory (see sections 9.2.2 and 9.4).

2. When the imaginary part H evaluated in two saddle points coincides for some parameter $\boldsymbol{\mu}_s$, the two corresponding steepest-descent contours can coincide. At such a parameter $\boldsymbol{\mu}_s$, the Lefschetz thimbles flip changing the number of relevant saddle points (see Fig. 9.8 for an illustration). This is known as a Stokes transition. The parameters $\boldsymbol{\mu}_s$ for which this happens form so-called Stokes lines.

We study both phenomena in detail in the Section 9.4.

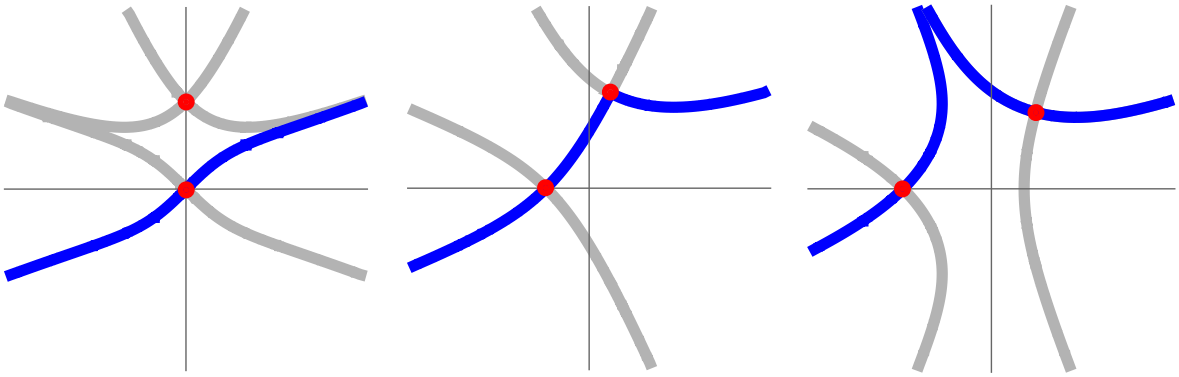


Figure 9.8: The Stokes phenomenon at which the steepest ascent and descent contours coincide and the relevant saddle points change. Steepest ascent and descent paths are shown as curves, with the relevant thimbles shown heavier. In the left panel, we have one real relevant saddle point. In the central panel, we see a Stokes phenomenon. The Lefschetz thimble passes through both a real and a complex saddle point. In the right panel, we have one real and one complex relevant saddle point. The Stokes phenomenon occurs when the steepest descent contour of a saddle point encounters another saddle point.

We numerically evaluate the flow X_λ by approximating X by a set of line-segments and flowing the endpoints. Since the real part of the analytic continuation of an analytic function does not have local extrema (this follows from the Cauchy-Riemann equation), all points $z \in \mathbb{C}$ flow to poles as $\lambda \to \infty$. The limit $\lim_{\lambda\to\infty} X_\lambda$ should not be interpreted as a pointwise limit. We, for this reason, trace the length of the line-segments and add points when neighboring points move too far apart. We moreover remove line-segments in the neighborhoods of the poles of the h-function. The contour X_λ has converged to the thimble when the imaginary part H is approximately constant along the line-segments.

This idea is implemented by the algorithm:

---

with the parameters $a, b, T_1, T_2, T_3 \in \mathbb{R}$, and $n \in \mathbb{Z}_{>0}$.

See Fig. 9.9 for the flow of the original integration domain corresponding the rational lens for $\alpha = 2$ and $\mu = 0$. For $\lambda = 0$ the contour X_λ coincides with the real line. As λ is increased to 1, the original integration domain smoothly flows to the Lefschetz thimble $\mathcal{J}$ consisting of three steepest descent contours $\mathcal{J}_i$ corresponding to three relevant saddle points $\bar{x}_i$. By evaluating the flow for varying α and μ, we obtain the Picard-Lefschetz analysis of the lens.

Require: Represent a subset $[a, b]$ of the original integration domain $X = \mathbb{R}$ by the regular lattice $p_i = a + i\Delta x$ with $\Delta x = \frac{b-a}{n}$ for some $n \in \mathbb{Z}_{>0}$, and the line-segments $(p_0, p_1), (p_1, p_2), \ldots, (p_{n-1}, p_n)$.
  while the variance of the imaginary part H on the points p_i exceeds threshold T_1 **do**
    flow the points: $p_i \mapsto p_i - \nabla h(p_i)\Delta t$
    if the h-function evaluated in the point p_i is smaller than the threshold T_2 **then**
      remove the corresponding line segments
    end if
    if the length of the line-segments (p_i, p_{i+1}) exceeds the threshold T_3 **then**
      split the line segment into the two lines $\left(p_i, \frac{p_i+p_{i+1}}{2}\right), \left(\frac{p_i+p_{i+1}}{2}, p_{i+1}\right)$.
    end if
  end while

Algorithm 1: The flow of the contour of one-dimensional oscillatory integrals.

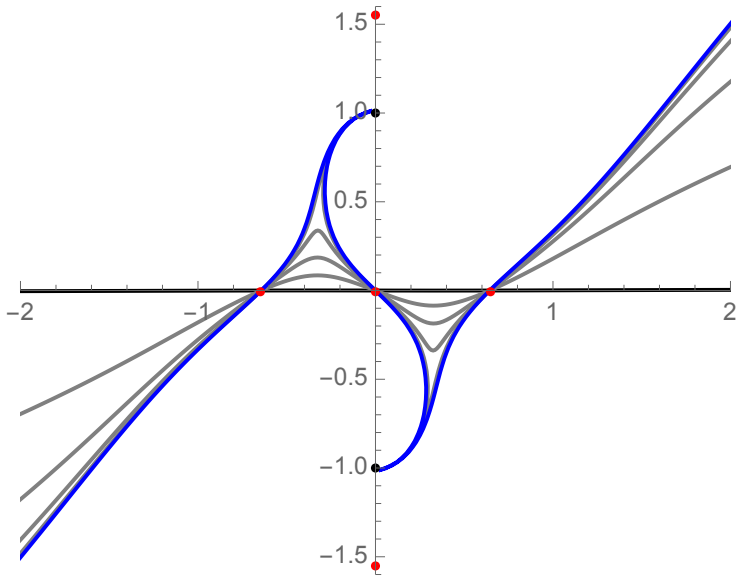


Figure 9.9: The downward flow of the integration domain. The contour X_λ for $\lambda = 0, 0.2, 0.4, 0.6, 0.8, 1$ is plotted as curves, the last being heavier. The five saddle points and the two poles (at $\lambda = \pm i$) are shown as points.

For multi-dimensional oscillatory integrals, the flow algorithm can be generalized by flowing the cells of a tessellation of the original integration domain. In this chapter, we start our calculations with the tessellation of a rectilinear lattice. For an two-dimensional illustration see Fig. 9.10.

Require: Represent a subset of the original integration domain X with a regular tessellation consisting of cells V_i spanned by the points $\boldsymbol{p}_{i,1}, \boldsymbol{p}_{i,2}, \ldots$.
while the variance of the imaginary part H on the points $p_{i,j}$ exceeds threshold T_1 **do**
  flow the points: $\boldsymbol{p}_{i,j} \mapsto \boldsymbol{p}_{i,j} - \nabla h(\boldsymbol{p}_{i,j})\Delta t$
  if the h-function evaluated in the point p_i is smaller than the threshold T_2 **then**
    remove the corresponding cells
  end if
  if the volume of a cell V_i exceeds the threshold T_3 **then**
    subdivide the cell into smaller cells
  end if
end while

Algorithm 2: The flow of the contour of multi-dimensional oscillatory integrals.

There are various possible implementations of this algorithm. However, it follows from Cauchy's theorem that the integral is insensitive to the details of the tesselation employed. For all reasonable tesselations, the algorithm terminates in a polynomial number of steps as it scales roughly linearly with the number of simplices. Remarkably, this cost scaling is no worse than that required by the geometric optics approximation.

9.3.3 Integrating along the thimbles

Given a Lefschetz thimble $\mathcal{J}$ for a range of α and μ, obtained with either one of the above-described methods, we perform the resulting integral along the thimble with the trapezium rule. Given a thimble $\mathcal{J}$ represented as a set of line-segments $l_i = (p_{i,1}, p_{i,2})$, the integral is approximated by

$$\Psi(\mu;\nu) \approx \sum_i \frac{e^{i\phi(p_{i,1};\mu)\nu} + e^{i\phi(p_{i,2};\mu)\nu}}{2}(p_{i,2} - p_{i,1}) \tag{9.38}$$

summed over the line segments. For multi-dimensional oscillatory integrals, we evaluate the integral on a linear approximation of the integrand on the tessellation of the thimble.

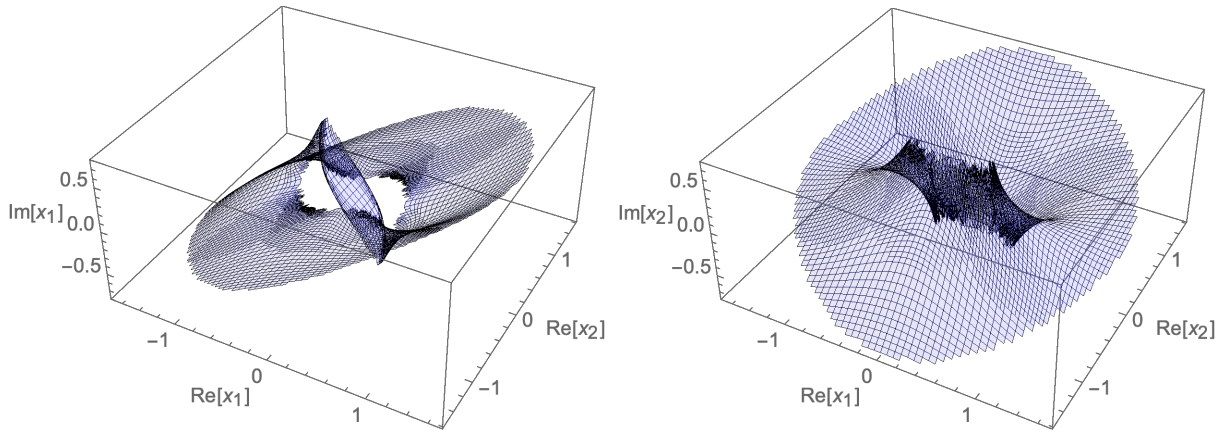


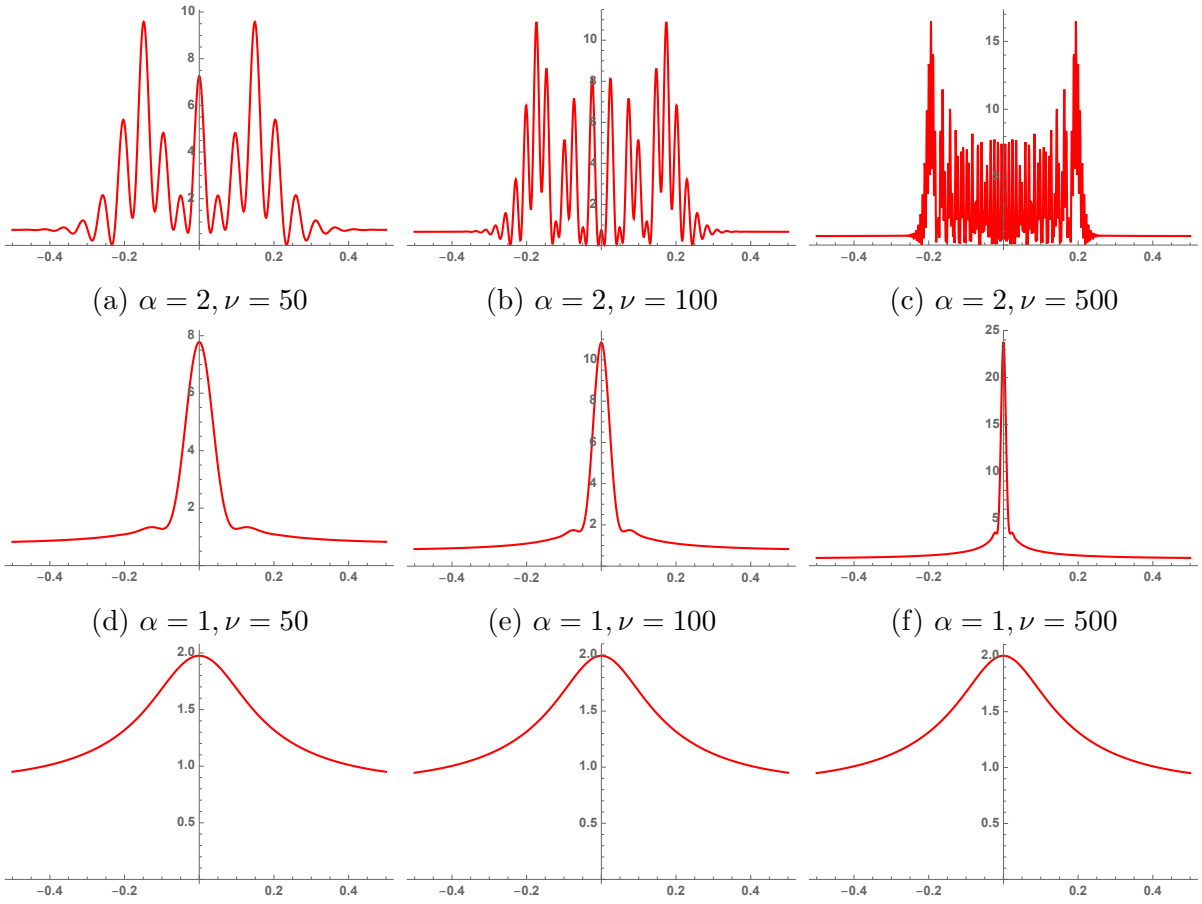
Figure 9.10: Two projections of the numerically obtained two-dimensional thimble $\mathcal{J}$ in $\mathbb{C}^2$ for a two-dimensional oscillatory integral.

Naively, one might expect to have to compute the Lefschetz contour $\mathcal{J}$ for every μ for which one wishes to perform the integral. However, since the thimble is a smooth function of μ, it suffices to compute the thimble for a range of μ. When integrating, we instead evaluate the integral on the thimble corresponding to the closest μ for which we have evaluated the thimble. Finally, it should be noted that for increasing ν, the support of the integral is increasingly concentrated around the relevant saddle points. As a consequence we can, for large ν, restrict the integral to the line segments close to the saddle points. It follows from this that the numerical evaluation of the integral along the thimble becomes more and more efficient as the frequency is increased. This is in sharp contrast with conventional integration techniques which need to trace many oscillations of the integrand along the real line.

See Fig. 9.11 for the normalized intensity profiles of the lens evaluated along the thimble for frequencies $\nu = 50, 100, 500$. We observe the following properties of the normalized intensity profiles:

- In the regime $\alpha < 1$, the lens leads to a single-image region. The normalized intensity profile does not oscillate and is moreover independent of the frequency ν. See the lower panels of Fig. 9.11.

- For $\alpha = 1$, the lens forms a cusp caustic. The caustic corresponds to the peak at $\mu_c = 0$. For increasing frequency, ν, the peak is enhanced and becomes increasingly narrow. In the eikonal limit $\nu \to \infty$, the normalized normalized intensity diverges as



(a) $\alpha = 2, \nu = 50$ (b) $\alpha = 2, \nu = 100$ (c) $\alpha = 2, \nu = 500$

(d) $\alpha = 1, \nu = 50$ (e) $\alpha = 1, \nu = 100$ (f) $\alpha = 1, \nu = 500$

(g) $\alpha = 1/2, \nu = 50$ (h) $\alpha = 1/2, \nu = 100$ (i) $\alpha = 1/2, \nu = 500$

Figure 9.11: The normalized intensity $I(\mu;\nu) = |\Psi(\mu;\nu)|^2$ for $\alpha = 1/2, 1, 2$ as a function of μ for $\nu = 50, 100, 500$.

$\nu^{1/2}$ at the caustic μ_c (see the scaling relations in Table 9.2). See the middle panels of Fig. 9.11.

- In the regime $\alpha > 1$, the lens forms a triple-image region which is bounded by two-fold caustics. We see that the triple-image region $(-\mu_c, \mu_c)$, with $\mu_c = 0.206751\ldots$ for $\alpha = 2$, consists of an interference pattern bounded by two peaks at $\mu = \pm\mu_c$. The interference pattern in the triple-image region is the result of the three real saddle points. The oscillations in the single-image region result from the interplay between the relevant real and the complex saddle point. For increasing ν, the fringes of the interference pattern shrink and spikes corresponding to the fold get sharper and are increasingly enhanced. For the relevant scalings see Table 9.2. See the upper panels of Fig. 9.11.

Note that the normalized intensity in the cusp exceeds the normalized intensity in the fold caustic. This related to the co-dimension of the caustic as described in Section 9.2.2. Moreover remark that the cusp caustic only exists at a single α for the one-dimensional lens, while the fold caustic appears for a range of α. Table 9.2 shows the frequency dependence of the pattern. Furthermore, note that the normalized intensity profiles at frequency $\nu = 500$, for $\alpha = 1/2, 1$, and 2, are close to the normalized intensity maps predicted by geometric optics (see Fig. 9.6).

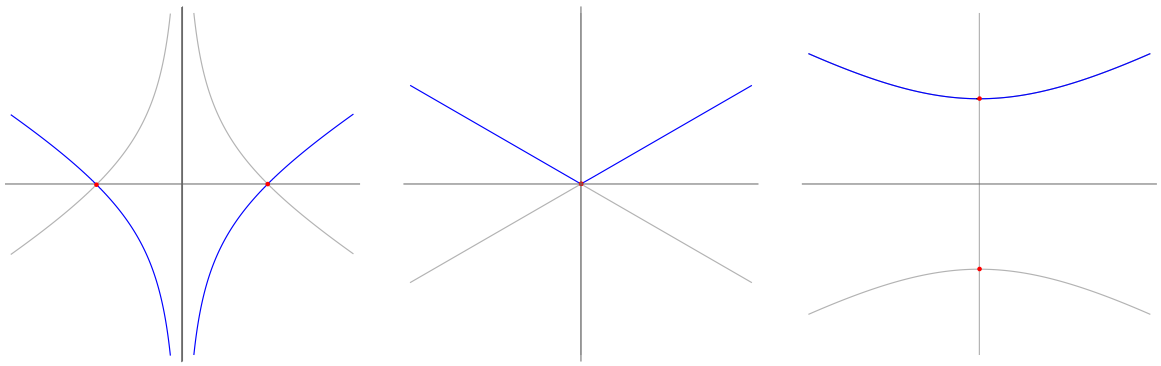
In the context of astronomical radio sources, the signal is dramatically enhanced when the relative position of the observer and the source move through the fold or the cusp caustic of the lens. One would in this context interpret the μ axis as the line traced by the source on the sky, i.e. $\mu = vt + \mu_0$ with μ_0 the initial position, v the speed of the source in parameter space and t the time. This amplification of the signal may be relevant as an selection effect for the recently observed Fast Radio Bursts. Note that if the observed FRBs are indeed the result of caustics in plasma lenses, we expect the peaks to evolve in a characteristic way and satisfy specific scaling relations in frequency space. See Section 9.6 for a more detailed discussion.

9.4 The elementary catastrophes

The unfoldings of the seven *elementary singularities* (see Table 9.1), form a local description of lenses near the caustics. We here study the Picard-Lefschetz analysis of the elementary catastrophes appearing in two-dimensional lenses and evaluate the corresponding normalized intensity maps using the flow algorithm described above. This analysis is complementary to the asymptotic analysis described in chapter 36 of [312].

9.4.1 The fold A_2

The fold singularity is the simplest degenerate critical point and can be viewed as the superposition of two non-degenerate saddle points. The Picard-Lefschetz analysis of the unfolding of the fold singularity is illustrated in Fig. 9.12. For negative μ, there are two relevant real saddle points (see Fig. 9.12a). As μ approaches the caustic at $\mu_c = 0$, the two saddle points merge and form the fold singularity (see Fig. 9.12b). Note that the fold saddle point emanates three steepest ascent and three descent curves. The thimble is non-differentiable at the degenerate saddle point. When μ is increased passed the caustic μ_c, the two saddle points move off the real axis and into the complex plane (see Fig. 9.12c). In this regime only one of them remains relevant.



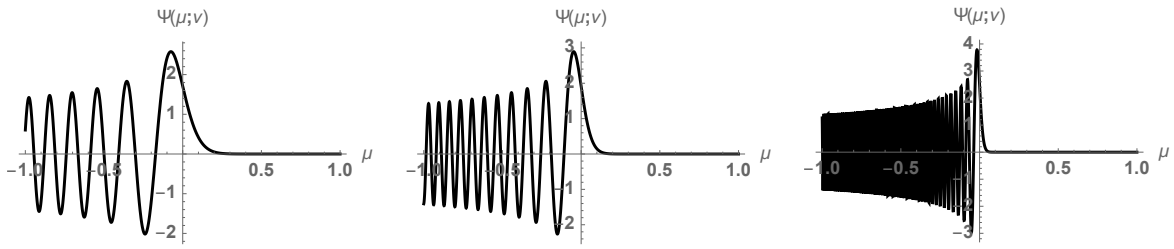
(a) $\mu < 0$ (b) $\mu = 0$ (c) $\mu > 0$

Figure 9.12: The saddle points and Lefschetz thimbles (shown as heavy curves) for the unfolding for the fold singularity A_2. Lighter curves show the remaining steepest ascent and descent contours.

The Fresnel-Kirchhoff integral for the fold singularity can be related to the Airy function

$$\Psi(\mu;\nu) = \sqrt{\frac{\nu}{\pi}} \int_{-\infty}^{\infty} e^{i\left(\frac{x^3}{3}+\mu x\right)\nu} \mathrm{d}x = 2\sqrt{\pi}\nu^{1/6}\mathrm{Ai}[\nu^{2/3}\mu]\,. \tag{9.39}$$

Note the appearance of the singularity and fringe indices 1/6 and 2/3 as listed in Table 9.2. It straightforward to derive the scaling of the amplitude and the fringes, with the change of coordinates $z = \nu^{1/3}x$. The other scaling relations are derived analogously. The Airy function is a good illustration of the interference phenomenon present in multi-image regions (seen in Fig. 9.13). The range $\mu < 0$, for which the two relevant saddle points reside on the real line corresponds to a double-image region, where two saddle points lead



(a) $\nu = 50$ (b) $\nu = 100$ (c) $\nu = 500$

Figure 9.13: The integral $\Psi(\mu, \nu)$ for the fold singularity as a function of μ for $\nu = 50, 100, 500$.

to an interference pattern. The range $\mu > 0$ corresponds to a zero-image region in which the amplitude asymptotes to zero as $\nu \to \infty$.

Asymptotics

Using the Picard-Lefschetz diagrams (Fig. 9.12) we can derive asymptotics for the integral $\Psi(\mu; \nu)$. For $\mu > 0$ the Picard-Lefschetz analysis consists of a single relevant saddle point located at $i\sqrt{\mu}$. The exponent can be approximated around the saddle point by

$$\phi(x; \mu) = i\frac{2}{3}\mu^{3/2} + i\sqrt{\mu}(x - i\sqrt{\mu})^2 + \mathcal{O}\left((x - i\sqrt{\mu})^3\right) . \tag{9.40}$$

The saddle point approximation for this point gives an exponential falloff

$$\Psi(\mu; \nu) \approx \frac{e^{-\frac{2}{3}\mu^{3/2}\nu}}{\sqrt{2}\mu^{1/4}} . \tag{9.41}$$

This matches the the behaviour in Fig. 9.13.

For $\mu < 0$, the Picard-Lefschetz analysis consists of two real relevant saddle points located at $x = \pm\sqrt{|\mu|}$. A saddle point approximation around these points gives us the oscillatory behaviour

$$\Psi(\mu; \nu) \approx \frac{2}{|\mu|^{\frac{1}{4}}} \sin\left(\frac{2|\mu|^{\frac{3}{2}}\nu}{3} + \frac{\pi}{4}\right) \tag{9.42}$$

seen in Fig. 9.13. Observe that wave function becomes increasingly oscillatory and falls off as a power law $\Psi(\mu; \nu) \propto \frac{1}{|\mu|^{1/4}}$ in the geometric limit $\nu \to \infty$.

9.4.2 The cusp A_3

The cusp singularity consists of the superposition of three non-degenerate saddle points. The singularity is of co-dimension $K = 2$ and has two unfolding parameters μ_1 and μ_2, *i.e.*,

$$\Psi(\boldsymbol{\mu};\nu) = \sqrt{\frac{\nu}{\pi}} \int_{-\infty}^{\infty} e^{i\left(\frac{x^4}{4}+\mu_2\frac{x^2}{2}+\mu_1 x\right)\nu} \mathrm{d}x \,. \tag{9.43}$$

See Fig. 9.14 for an illustration of unfolding of the cusp caustic and the Picard-Lefschetz analysis in the (μ_1, μ_2)-plane.

The the exponent $\phi(x;\boldsymbol{\mu})$ has three saddle points $\bar{x}_i$, given by the roots of the cubic equation

$$x^3 + \mu_2 x + \mu_1 = 0 \,. \tag{9.44}$$

Depending on μ, either one or three of the saddle points are real-valued. The complex-valued saddle points always come in conjugate pairs since $\phi(x;\boldsymbol{\mu})$ is a real-valued function, *i.e.* real for real x.

Geometric optics applied to this integral shows that the cusp caustic at $(\mu_1, \mu_2) = (0, 0)$ emanates two fold-lines $A_2 \subset M$, given by cubic root

$$\mu_2 = -\frac{3}{2^{2/3}} |\mu_1|^{2/3} \,. \tag{9.45}$$

The fold lines are non-differentiable at the cusp singularity $(\mu_1, \mu_2) = (0, 0)$.

In the triple-image region enclosed by the two fold-lines, the thimble passes through three real-valued saddle points. When approaching one of the fold lines, we see that two of the real saddle points merge and move in the complex plane. Only one of the two complex saddle points remains relevant to the integral. This is analogous to the behavior observed in the analysis of the fold caustic. At the cusp saddle point at $(\mu_1, \mu_2) = (0, 0)$ all three saddle points merge at the origin. Finally, note that the single-image region consists of three subregions, for which the Picard-Lefschetz analysis either consists of one or two relevant saddle points. These subregions are separated by two Stokes lines (red dashed lines in Fig. 9.14). Along these lines, the Lefschetz thimbles flip while the saddle points remain separated. The Stokes lines can be found by equating the imaginary parts of the exponents evaluated at the saddle points, *i.e.*,

$$\mathrm{Im}[i\phi(\bar{x}_i;\boldsymbol{\mu})\nu] = \mathrm{Im}[i\phi(\bar{x}_j;\boldsymbol{\mu})\nu] \tag{9.46}$$

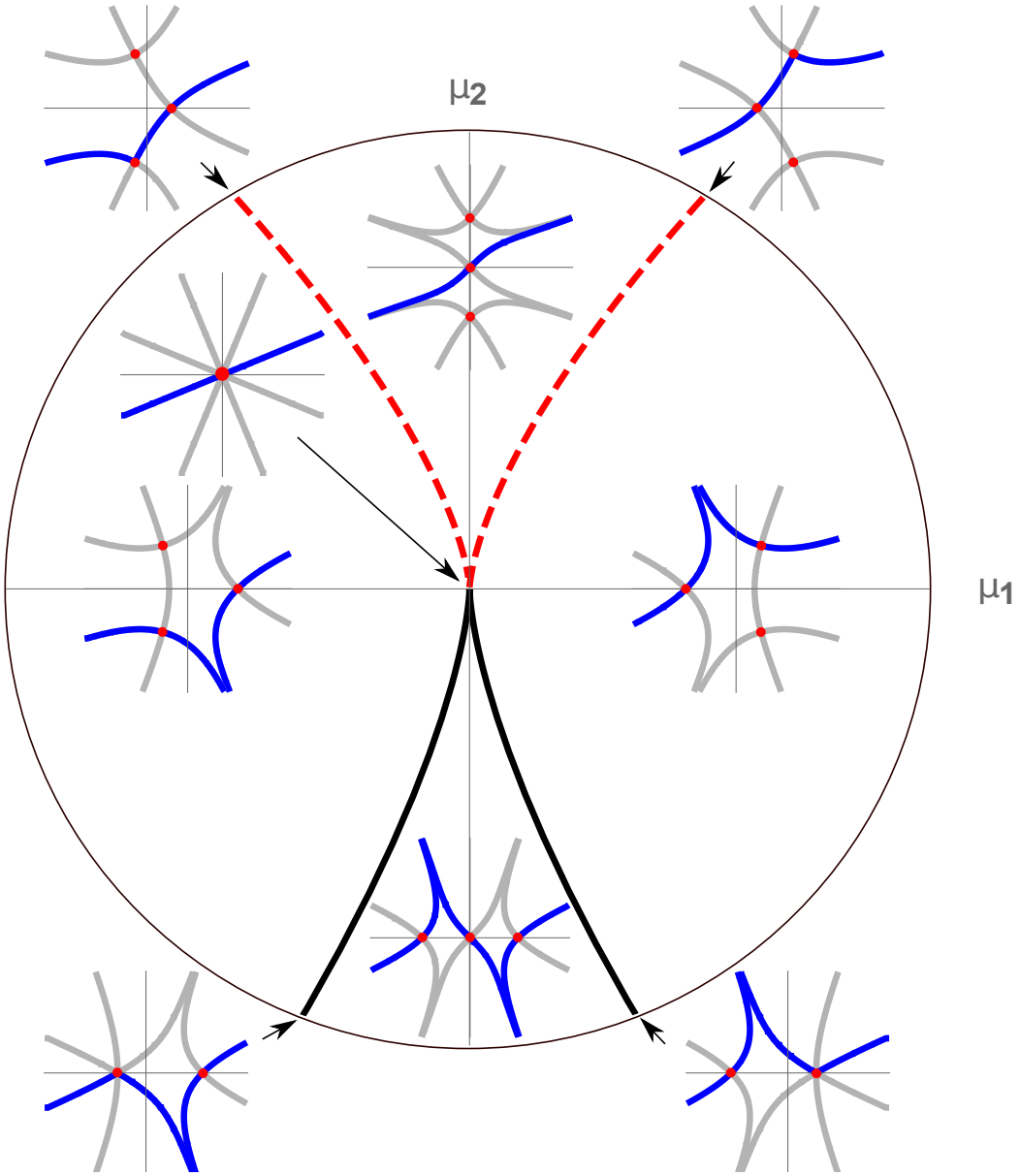
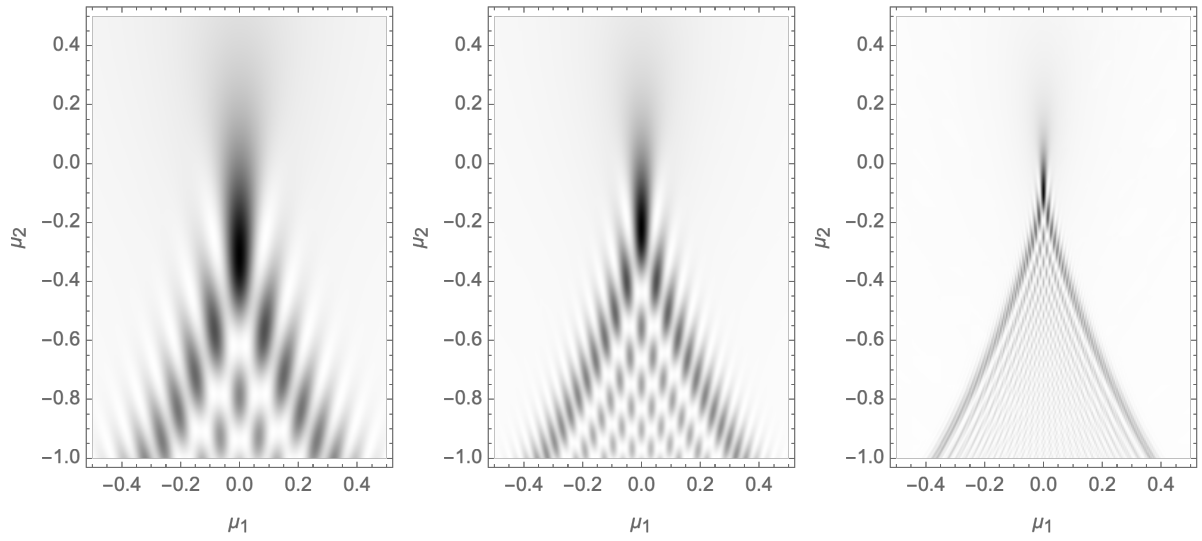


Figure 9.14: The cusp singularity in the unfolding $\boldsymbol{\mu}$-plane. The heavy solid curve is the fold line separating the lower, triple-image region from the remaining single-image region. The dashed curve is the Stokes line. We observe that the Stokes transition corresponds to one complex saddle point becoming (ir)relevant, and that the caustics correspond to the merger of two real, non-degenerate saddle points. There is only one real relevant saddle point in the single image region. There are three real relevant saddles in the lower, triple image region.



(a) $\nu = 50$ (b) $\nu = 100$ (c) $\nu = 500$

Figure 9.15: The normalized intensity $I(\boldsymbol{\mu}; \nu)$ for the cusp caustic for $\nu = 50, 100$ and 500.

for $i \neq j$. For the unfolding of the cusp, we see that the Stokes lines are described by

$$\mu_2 = 3\sqrt[3]{\frac{3\sqrt{3}-5}{2}}|\mu_1|^{2/3}\,, \tag{9.47}$$

for $\mu_1 < 0$ and μ_2. Note that the amplitude across a Stokes line is smooth, even though the saddle point structure changes abruptly. The Stokes lines can be interpreted as the points for which the saddle point approximation of the integral fails.

Numerics

Given the Lefschetz thimble, we can numerically evaluate the amplitude (see Fig. 9.15). In the eikonal limit $\nu \to \infty$ we observe the emergence of a fold-line (A_2) with a sharp exponential falloff in most of the single-image region and the power-law falloff in the triple-image region. We also see the emergence of a cusp caustic at the origin with a power-law falloff along the line $\{\mu_1 = 0\}$.

Asymptotics

The Picard-Lefschetz diagrams (Fig. 9.14) allow us to derive limiting behaviour for $\Psi(\boldsymbol{\mu};\nu)$. From the functional form of $\phi(x;\boldsymbol{\mu})$ along the line $\{\mu_1 = 0\}$, *i.e.*,

$$\phi(x;\boldsymbol{\mu}) = \frac{\mu_2 x^2}{2} + \frac{x^4}{4}\,, \tag{9.48}$$

we observe that one of the relevant saddle points is located at the origin $x = 0$. For $\mu_2 > 0$ this is the only relevant saddle point, whereas for $\mu_2 < 0$ it is one of three real relevant saddle points.

In the case $\mu_2 > 0$, we find that the single saddle leads to a power-law

$$\Psi(x;\boldsymbol{\mu}) \approx \sqrt{\frac{\nu}{\pi}} \int_{-\infty}^{\infty} e^{i\frac{\mu_2\nu}{2}x^2}\mathrm{d}x = \sqrt{2}(-i\mu_2)^{-1/2} \tag{9.49}$$

which in the normalized intensity corresponds to the falloff

$$I(x;\mu) = |\Psi(x;\boldsymbol{\mu})|^2 \approx \frac{2}{\mu_2}\,, \tag{9.50}$$

independent of the frequency. This feature is absent in the unfolding of the fold caustic.

For $\mu_2 < 0$, the Picard-Lefschetz analysis consists of three real relevant saddle points located at $\pm\sqrt{\mu_2}$ and 0. The exponent at the saddle point $\pm\sqrt{-\mu_2}$ can be approximated by

$$\phi(x;\boldsymbol{\mu}) = -\frac{\mu_2^2}{4} - \mu_2(x \pm \sqrt{-\mu_2})^2 + \mathcal{O}\left((x \pm \sqrt{-\mu_2})^3\right)\,. \tag{9.51}$$

In the saddle point approximation,

$$\Psi(\boldsymbol{\mu};\nu) \approx \sqrt{2}\frac{-(-1)^{3/4} + (1+i)e^{-\frac{i}{4}\mu_2^2\nu}}{\sqrt{-\mu_2}}\,. \tag{9.52}$$

The normalized intensity $I(\mu;\nu)$ thus oscillates in μ_2 with increasing frequency with power-law suppression

$$I(\boldsymbol{\mu};\nu) \propto \frac{2}{\mu_2}\,. \tag{9.53}$$

Along the line $\{\mu_2 = 0\}$, the Picard-Lefschetz analysis consists of two relevant saddle points, one real and one complex. The real relevant saddle point is located at $\bar{x} = -\mu_1^{1/3}$, giving the oscillatory behaviour

$$\Psi(\boldsymbol{\mu};\nu) \approx \sqrt{2}\frac{(-1)^{1/4}e^{-i\mu_1^{4/3}\nu}}{-\sqrt{3}\mu_1^{1/3}}\,, \tag{9.54}$$

so that the normalized intensity again falls off as a power-law

$$I(\boldsymbol{\mu};\nu) = |\Psi(\boldsymbol{\mu};\nu)|^2 \approx 2\mu_1^{-2/3}\,. \tag{9.55}$$

9.4.3 The swallowtail A_4

The swallowtail singularity is more complicated, as it consists of the superposition of four non-degenerate saddle points. The singularity is of co-dimension $K = 3$ and has three unfolding parameters μ_1, μ_2 and μ_3, *i.e.*,

$$\Psi(\boldsymbol{\mu};\nu) = \sqrt{\frac{\nu}{\pi}}\int_{-\infty}^{\infty} e^{i\left(\frac{x^5}{5}+\mu_3\frac{x^3}{3}+\mu_2\frac{x^2}{2}+\mu_1 x\right)\nu}\mathrm{d}x\,. \tag{9.56}$$

See figures 9.17 and 9.18 for an lustration of unfolding of the swallowtail caustic and the Picard-Lefschetz analysis in the (μ_1, μ_2, μ_3)-space.

The analytic continuation of the exponent $i\phi(x;\boldsymbol{\mu})\nu$ has four saddle points $\bar{x}_i$, given by the roots of the quartic equation

$$x^4 + \mu_3 x^2 + \mu_2 x + \mu_1 = 0\,. \tag{9.57}$$

Depending on $\boldsymbol{\mu}$, either zero, two or four of the saddle points are real-valued. The complex-valued saddle points always come in conjugate pairs since $\phi(x;\boldsymbol{\mu})$ is a real-valued function.

Geometric optics applied to this integral shows that the swallowtail caustic at $\boldsymbol{\mu} = (0,0,0)$ emanates a cusp-line and a fold-surface (see Fig. 9.16). The fold-surface (the yellow surface in Fig. 9.16) is given by

$$A_2 = \left\{(3u^4 + u^2 v, -4u^3 - 2uv, v) | (u,v) \in \mathbb{R}^2\right\} \subset M \tag{9.58}$$

satisfying the two constraints

$$\frac{\mathrm{d}\phi(x;\boldsymbol{\mu})}{\mathrm{d}x} = 0\,, \quad \frac{\mathrm{d}^2\phi(x;\boldsymbol{\mu})}{\mathrm{d}x^2} = 0\,. \tag{9.59}$$

The cusp-line (the black curve in Fig. 9.16) lays on the fold-surface and is given by

$$A_3 = \{(-3t^4, 8t^3, -6t^2) | t \in \mathbb{R}\} \subset M \tag{9.60}$$

satisfying the three constraints

$$\frac{\mathrm{d}\phi(x;\boldsymbol{\mu})}{\mathrm{d}x} = 0\,, \quad \frac{\mathrm{d}^2\phi(x;\boldsymbol{\mu})}{\mathrm{d}x^2} = 0\,, \quad \frac{\mathrm{d}^3\phi(x;\boldsymbol{\mu})}{\mathrm{d}x^3} = 0\,. \tag{9.61}$$

Note that the caustics are symmetric in the (μ_1, μ_3)-plane and that caustics only appear for negative μ_1. This aids our analysis, since we can consider the three-dimensional swallowtail unfolding as a one parameter family of unfoldings in the $\{\mu_3 = \text{const}\}$ planes.

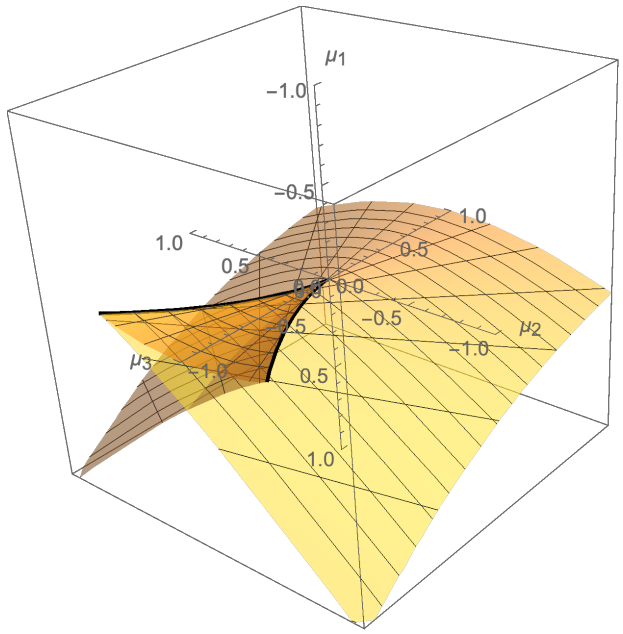


Figure 9.16: The swallowtail singularity in the unfolding space (μ_1, μ_2, μ_3). The curved surface is the fold surface separating the single- and triple-image regions. The heavy line is the cusp line, along which we find cusp saddle points.

In figures 9.17 and 9.18 we plot three slices of the fold-surface and cusp-line for $\mu_3 = -1, 0$ and $+1$. For $\mu_3 = -1$ we obtain the characteristic swallowtail shape in the fold-surface with the cusp-line intersecting at the tips, which gives the singularity its name. For $\mu_3 = 0$ we see the actual swallow caustic. The slice $\mu_3 = +1$, is simpler as it does not

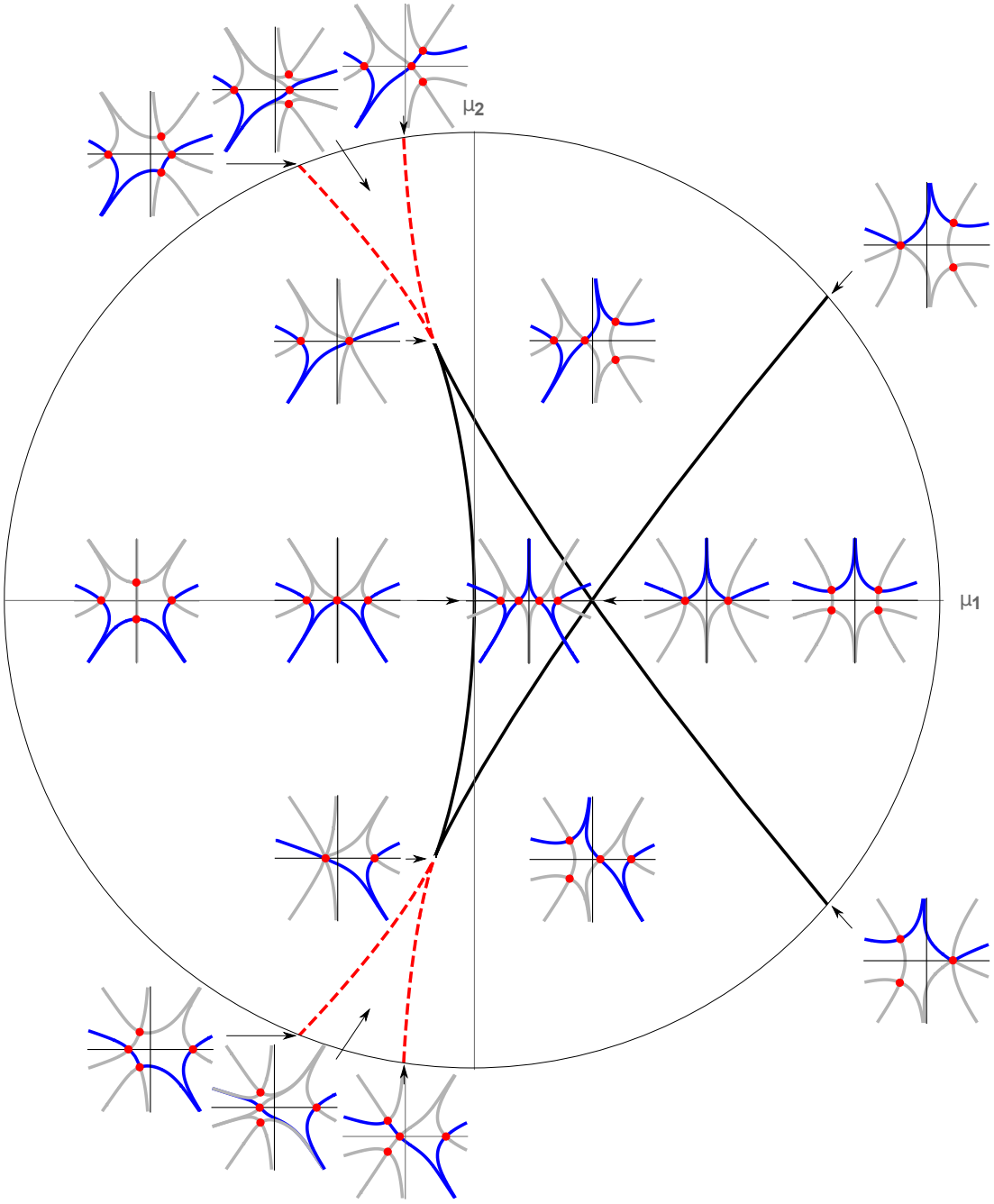


Figure 9.17: Picard-Lefschetz analysis of the unfolding of the swallowtail (A_4) singularity at $\mu_3 = -1$. The number of real saddles gives the number of images in geometric optics.

contain intersections with the cusp-line and only consists of the fold-surface separating two regions.

Given the caustics of geometric optics, we can evaluate the Lefschetz thimble. It again suffices to study the three cases $\mu_3 < 0, \mu_3 = 0$ and $\mu_3 > 0$:

- We start by analysing the saddle points in the $\mu_3 = -1$ plane (Fig. 9.17). The Picard-Lefschetz analysis for the enclosed region in the middle of the circle consists of four relevant real saddle points. This is a quadruple-image region. Note that multiple-image regions for localized lenses always consist of an odd number of images. In such lenses, the swallowtail will in practice always appear near another caustic such as a fold.

  Starting from the quadruple-image region and moving through the fold-line on the left, we observe that the two central saddle points merge to form a fold saddle point. The two saddle points subsequently move in the complex plane, one remaining relevant. Since this region corresponds to two real saddle points it is a double-image region.

  Again, starting from the quadruple-image region and moving in the vertical direction, we observe that two of the outer saddle points merge to form a fold saddle point and subsequently move into the complex plane. The resulting Picard-Lefschetz analysis again consists of three relevant saddle points; two real and one complex. This again is a double-image region. If we, however, move from this double-image region to the double-image region on the left of the quadruple-image region, we pass through two Stokes lines, at which the complex saddle point switches from relevant to irrelevant. The Stokes lines are defined by

  $$\mathrm{Im}[i\phi(\bar{x}_i;\boldsymbol{\mu})\nu] = \mathrm{Im}[i\phi(\bar{x}_j;\boldsymbol{\mu})\nu] \tag{9.62}$$

  for $i \neq j$. Note that the Stokes lines can be associated with the cusp caustic at the tips of the fold-line. Note that the three relevant saddle points merge at these tips, to form a cusp saddle point.

  Finally, if we move from the quadruple-image region along the line $\mu_2 = 0$ to the right, we pass through the intersection of the fold lines. At this point, both the left and right two real saddle points merge to form a 'double' fold caustic. After passing this point, the four saddle points move in the complex plane. The Picard-Lefschetz analysis consists here of two relevant complex saddle points. This is a zero-image region (which will not be realized in localized lenses). If we pass from the double-image region to the zero-image region, we again observe a fold caustic in which two

relevant real saddle points merge and move in the complex plane. This completes the analysis of the unfolding of the swallowtail caustic at $\mu_3 = -1$.

- For $\mu_3 = 0$, the geometry of the fold-line is simpler as the quadruple-image region has merged into the swallowtail caustic at the origin (see Fig. 9.18a). The Picard-Lefschetz analysis of this slice is largely similar to the one at $\mu_3 = -1$. The double-image region (including the Stokes lines) has been deformed but is otherwise the same. The zero-image region is also unchanged. However, the intersection of the two fold-lines is replaced by the swallowtail saddle point at the origin of in the (μ_1, μ_2)-plane. Since this saddle point is the superposition of four non-degenerate saddle points, the amplitude integral is enhanced.

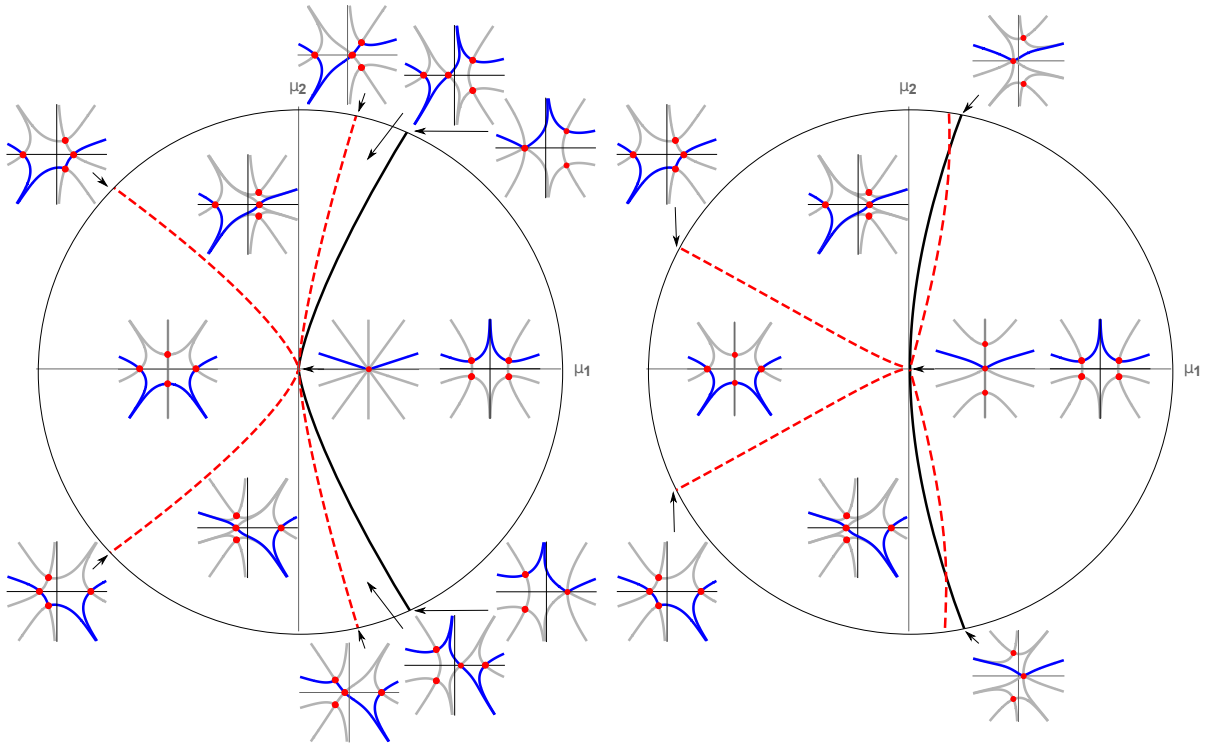
- For $\mu_3 = +1$, the geometry of the caustics is depicted in Fig. 9.18b. The fold-line separates the zero-image region on the right from the double-image region on the left. Since the Picard-Lefschetz diagram in the zero-image region consists of four complex saddle points – two of them being relevant – there exist two distinct ways in which we can pass to the double-image region; either by merging the two saddle points on the left or on the right (see upper and lower diagram). The transition between these to takes place at the origin, where the four saddle points are located on the imaginary axis. The double-image region consists of three subregions. The rightmost Stokes lines at $\mu_3 = 0$ (see Fig. 9.18a) have partly moved into the zero-image region.

By patching the Picard-Lefschetz analysis at $\mu_3 = -1, 0$ and $+1$ together, we obtain a complete description of the unfolding of the swallowtail singularity in the (μ_1, μ_2, μ_3)-space. Note that the Stokes lines obtained in figures 9.17 and 9.18 are intersections of Stokes-surfaces, which together with the fold-surface partition the μ-space.

Numerics

Given the Lefschetz thimble, we can numerically compute the normalized intensity map of the lens (see Fig. 9.19). The left, central and right panels depict the normalized intensity $I(\boldsymbol{\mu}; \nu)$ for $\mu_3 = -1, 0$ and $+1$. The upper, middle and lower panels depict the different frequencies $\nu = 50, 100$ and 500.

We observe that for $\nu = 50$, interference is a dominant feature of the geometry of the caustic. The images are blurry and the geometry of the swallowtail is not resolved (Fig. 9.19a). We do observe the power-law falloff associated with the cusp singularities, which contrasts with the exponential falloff of the fold singularities.



(a) $\mu_3 = 0$ (b) $\mu_3 = +1$

Figure 9.18: The Picard-Lefschetz analysis of the unfolding of the swallowtail (A_4) singularity at $\mu_3 = 0$ and $+1$. The number of real saddles gives the number of images in the geometric optics approximation.

In the eikonal limit $\nu \to \infty$ we observe the emergence of a fold-line (A_2) with cusps (A_3). For $\nu = 500$ the swallowtail structure at $\mu_3 = -1$ is fully resolved. Note the difference in normalized intensity between the double- and quadruple-image regions. As the frequency ν is increased we observe that the enhanced flares, in the double-image regions, corresponding to the cusp caustics get thinner. However, note that they are independent of the frequency ν.

9.4.4 The elliptic umbilic D_4^-

The caustics described above were part of the A-family. They are of co-rank 1 and can be described by a one-dimensional integral. This should be contrasted by the D family which is of co-rank 2 and can only be studied in two-dimensional integrals.

The elliptic umbilic D_4^- is a singularity with co-rank 2 and co-dimension $K = 3$. The unfolding is described in terms of the three unfolding parameters (μ_1, μ_2, μ_3). We consider the interference pattern emerging from the integral

$$\Psi(\boldsymbol{\mu}; \nu) = \frac{\nu}{\pi} \int_{\mathbb{R}^2} e^{i\left(x_1^3 - 3x_1x_2^2 - \mu_3(x_1^2+x_2^2) - \mu_2 x_2 - \mu_1 x_1\right)\nu} \mathrm{d}x_1 \mathrm{d}x_2 \,. \tag{9.63}$$

The analytic continuation of the exponent $i\phi(\boldsymbol{x}; \boldsymbol{\mu})\nu$ has four saddle points $\bar{x}_i$, given by the roots of the two quadratic equations

$$3x_1^2 - 3x_2^2 - 2\mu_3 x_1 - \mu_1 = 0 \tag{9.64}$$

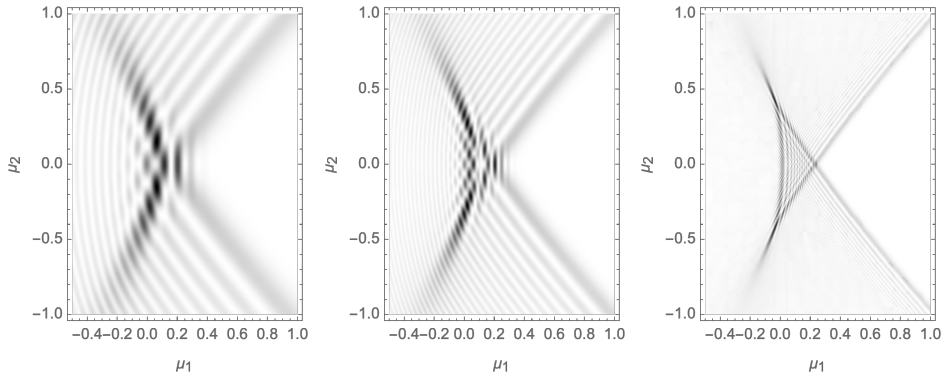
$$-6x_1x_2 - 2\mu_3 x_2 - \mu_2 = 0 \,. \tag{9.65}$$

Depending on $\boldsymbol{\mu}$, either two or four of the saddle points are real-valued. The complex-valued saddle points always come in conjugate pairs since $\phi(\boldsymbol{x}; \boldsymbol{\mu})$ is real-valued for real x. Solving this set of equations for μ_1 and μ_2 we obtain the Lagrangian map as a function of μ_3,

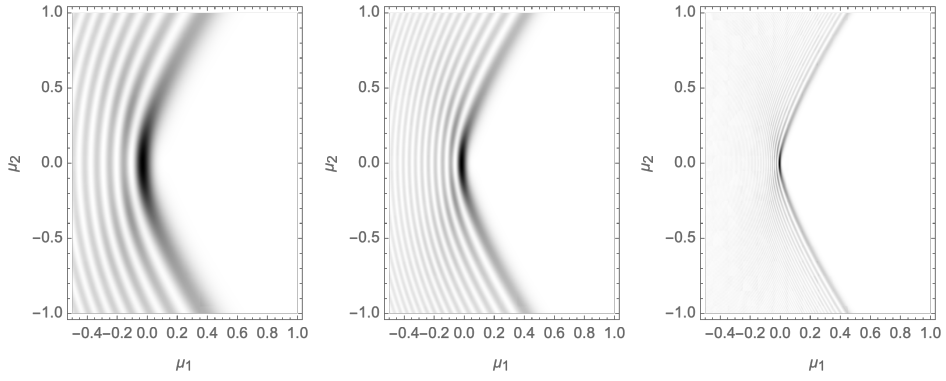
$$\xi_{\mu_3}(x_1, x_2) = (3x_1^2 - 3x_2^2 - 2x_1\mu_3, -2x_2(3x_1 + \mu_3), \mu_3) \,. \tag{9.66}$$

In the geometric limit, we form a fold-surface and three cusp lines. The fold-surface in base space $X = \mathbb{R}^2$ is given by

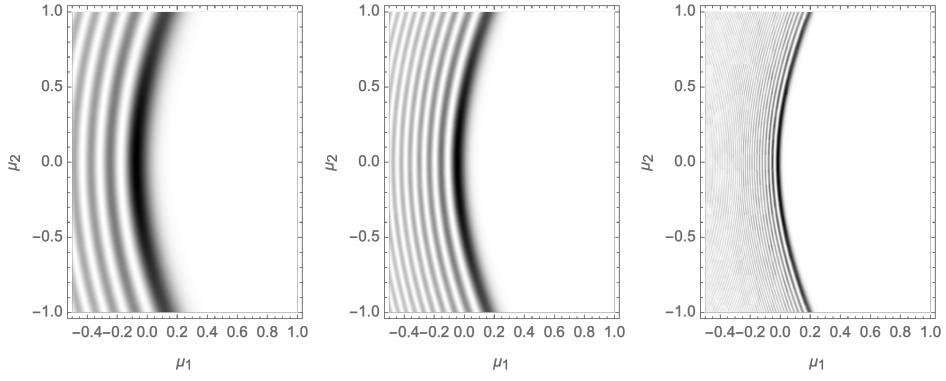
$$A_2^X(\mu_3) = \left\{ \left(\frac{\mu_3}{3} \cos\theta, \frac{\mu_3}{3} \sin\theta \right) | \theta \in [0, 2\pi) \right\} \tag{9.67}$$



(a) $\mu_3 = -1, \nu = 50$ (b) $\mu_3 = -1, \nu = 100$ (c) $\mu_3 = -1, \nu = 500$



(d) $\mu_3 = 0, \nu = 50$ (e) $\mu_3 = 0, \nu = 100$ (f) $\mu_3 = 0, \nu = 500$



(g) $\mu_3 = +1, \nu = 50$ (h) $\mu_3 = +1, \nu = 100$ (i) $\mu_3 = +1, \nu = 500$

Figure 9.19: The normalized intensity, $I(\boldsymbol{\mu}; \nu)$, of the unfolding of the swallowtail caustic (A_4) sliced by the surfaces $\{\mu_3 = -1\}, \{\mu_3 = 0\}, \{\mu_3 = +1\}$ (respectively the left, central and right panels) for the frequencies $\nu = 50, 100$ and 500 (respectively the upper, the middle and lower panels).

which is a cylinder with radius $\frac{\mu_3}{3}$, satisfying the equation

$$|\mathcal{M}| = 0\,, \tag{9.68}$$

where the deformation tensor is given by

$$\mathcal{M} = \left[\frac{\partial^2 \phi(\boldsymbol{x};\boldsymbol{\mu})}{\partial x_i \partial x_j}\right]_{i,j=1,2} \tag{9.69}$$

$$= \begin{pmatrix} 6x_1 - 2\mu_3 & -6x_2 \\ -6x_2 & -6x_1 - 2\mu_3 \end{pmatrix}. \tag{9.70}$$

The three cusp-lines are straight lines and lay on the fold-surface,

$$A_3^X(\mu_3) = \left\{ \left(\frac{\mu_3}{3}, 0\right), \left(\frac{\mu_3}{3}\cos\frac{2\pi}{3}, \frac{\mu_3}{3}\sin\frac{2\pi}{3}\right), \left(\frac{\mu_3}{3}\cos\frac{4\pi}{3}, \frac{\mu_3}{3}\sin\frac{4\pi}{3}\right)\right\} \tag{9.71}$$

in the X space.

In M space, after being mapped by ξ_{μ_3}, the elliptic umbilic point is located at the origin. The fold-surface is given by

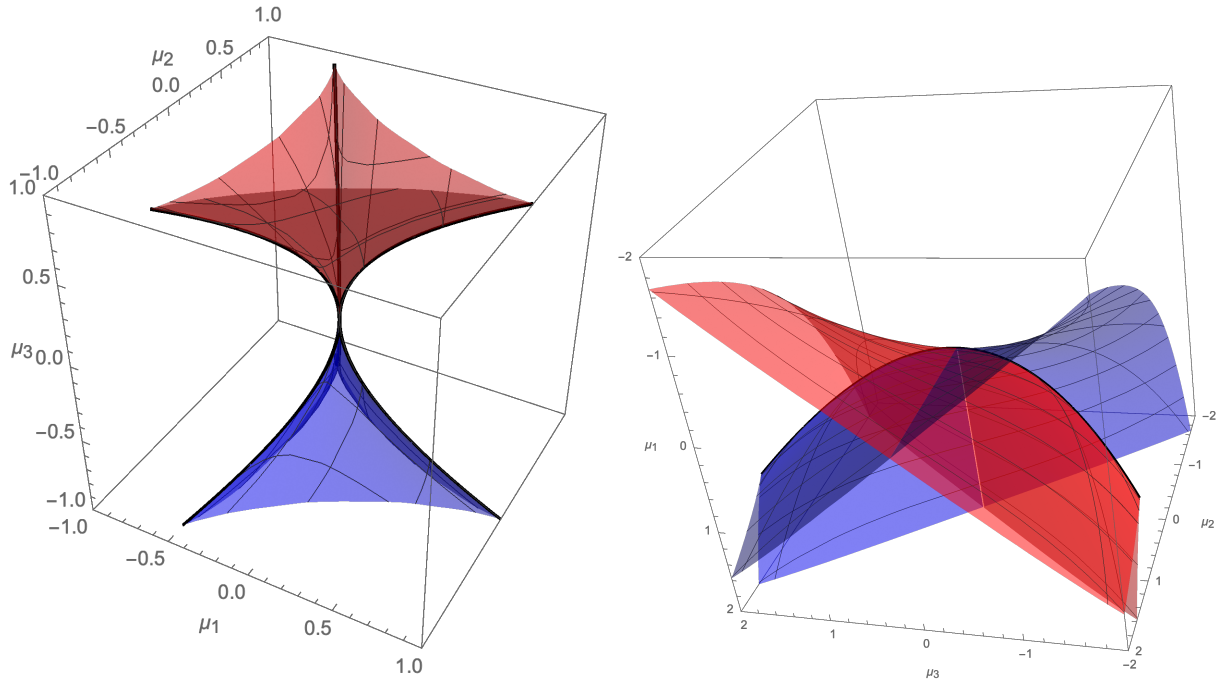
$$A_2 = \left\{ \left(\frac{\mu_3^2}{3}(\mp 2\cos\theta + \cos(2\theta), -\frac{2\mu_3^2}{3}(\pm 1 + \cos(\theta))\sin(\theta), \pm\mu_3 \right) \middle| \theta \in [0, 2\pi), \mu_3 \in \mathbb{R} \right\} \tag{9.72}$$

where the two branches corresponding to $\pm$ correspond to two disconnected pieces corresponding to the two eigenvalue fields of $\mathcal{M}$. The cusp lines are given by

$$A_3 = \left\{ (t^2, 0, t), (-t^2/2, \sqrt{3}t^2/2, t), (-t^2/2, -\sqrt{3}t^2/2, t) | t \in \mathbb{R} \right\}. \tag{9.73}$$

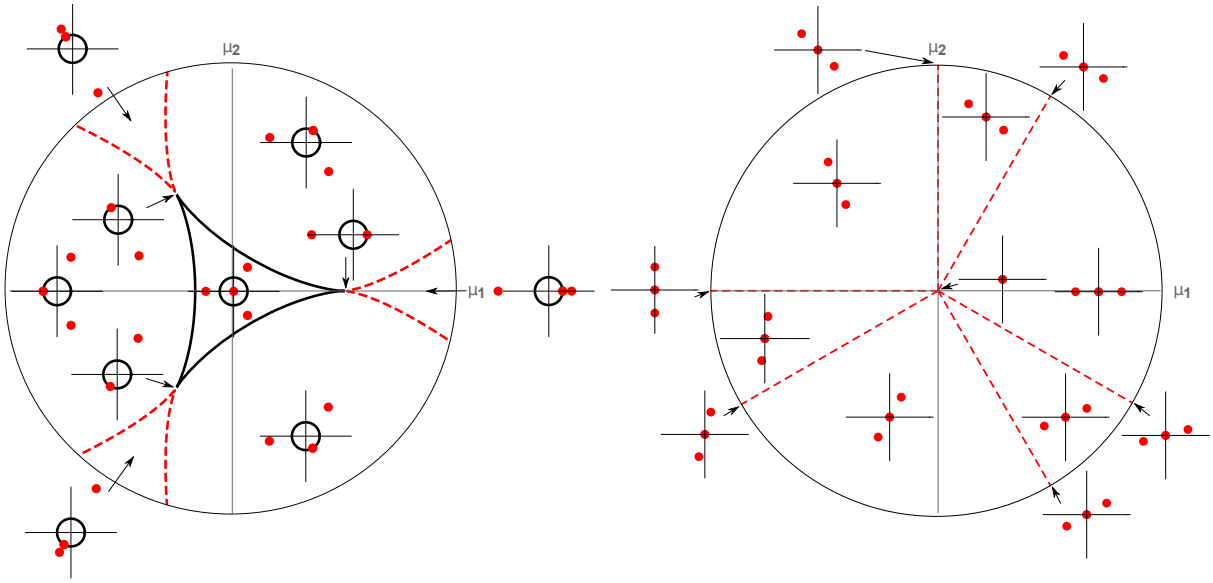
The fold-surface and cusp lines are illustrated in Fig. 9.20a. The upper and the lower surfaces denote the fold surfaces corresponding to the eigenvalue fields λ_1 and λ_2. The fold surface has a harp edge at the cusp lines (in black).

Note the symmetry of the triangular singularity and point symmetry of the caustic. By performing the Picard-Lefschetz analysis for the two slices $\mu_3 = \pm 1$ and $\mu_3 = 0$ we can obtain the Picard-Lefschetz diagram of the unfolding of the singularity. See Fig. 9.21 for the Picard-Lefschetz analysis of the two slices. The small diagrams are the real parts of the four saddle points in the (x_1, x_2)-plane. The black circle is the caustic in the base space at the corresponding μ_3.



(a) Elliptic umbilic D_4^- (b) Hyperbolic umbilic D_4^+

Figure 9.20: The elliptic (D_4^-) and hyperbolic umbilic (D_4^+) singularities in the unfolding space (μ_1, μ_2, μ_3). For D_4^-, the upper and lower curved surfaces are the fold-surfaces corresponding to the eigenvalue fields λ_1 and λ_2, separating the single-and triple-image regions. The heavy lines are cusp-lines, along which we find the cusp saddle points. For D_4^+, the fold surfaces for λ_1 and λ_2 extend from upper left to lower right and from upper right to lower left respectively.



(a) $\mu_3 = \pm 1$ (b) $\mu_3 = 0$

Figure 9.21: Intersection of the elliptic umbilic D_4^- by the surface $\{\mu_3 = -1\}$. The heavy curve is the fold-line and the dashed curve is the Stokes line. The points represent the four saddle points. Note that the saddle points away from the black lines are real-valued. The saddle points on the black curves are complex-valued and come in conjugate pairs. The number of real saddles gives the number of images in the geometric optics approximation.

- We first consider the case $\mu_3 \neq 0$. At the origin, the four saddle points are real-valued (see Fig. 9.21a). As a consequence, we conclude that they are all relevant. This is a quadruple-image region. One of the four saddle points is located inside the black circle. The other three are symmetrically distributed around the circle.

  When crossing the fold-line, the saddle point in the circle merges with one of the outer saddle points on the circle. After passing the fold-line, the two saddle points become complex. The saddle point with the smallest real part of the exponent $i\phi(\bar{\boldsymbol{x}}; \boldsymbol{\mu})$ will remain relevant whereas the other saddle point becomes irrelevant. The outside of the triangle is a double-image region. Note that the real parts of the two complex saddle points always coincides with the black circle.

  When approaching one of the three the cusp points, three of the four saddle points merge at a single point on the circle. Note that the four saddle points are collinear in the cusps.

  Finally, note that the double-image region consists of six subregions divided by six Stokes lines. In the regions on the left, the upper right and the lower right, the Lefschetz thimble passes through two real and one complex saddle point. In the regions to the right, upper left and lower right, the Picard-Lefschetz analysis consist of only two real saddle points (the ones outside the circle. The Stokes lines are again associated with the three cusps.

- In the case, $\mu_3 = 0$, the central region is replaced by the elliptic umbilic saddle point (see Fig. 9.21b). The rest of the μ_1-μ_2-plane is divided into six distinct regions by the six Stokes lines. The upper left, lower left, and the upper right regions consist of two relevant real saddle points. These regions correspond to the upper left, lower left, and right region in Fig. 9.21a. The three remaining regions consist of two real and one complex relevant saddle points.

These slices form a complete description of the Lefschetz thimble of the unfolding of the elliptic umbilic in the μ-space.

Numerics

Given the Lefschetz thimble, we can numerically evaluate the normalized intensity map of the lens (Fig. 9.22). The upper and lower panels depict the normalized intensity $I(\boldsymbol{\mu}; \nu)$ for $\mu_3 = \pm 1$ and 0. The left, central and right panels depict the frequencies $\nu = 50, 100$ and 500.

The normalized intensity map corresponding to the unfolding of the elliptic umbilic (D_4^-) has a triangular symmetry. As the frequency increases, the normalized intensity profile steepest and increases in amplitude. In the plane $\mu_3 = \pm 1$ we observe a fold-line in a triangular configuration with three cusp caustics at the corners. For the frequency, $\nu = 50$ the fold-line is relatively blurry. We again observe outward stripes emanating from the cusp caustics. These again follow a power-law falloff independent of the frequency. As the frequency is raised to $\nu = 100$ and $\nu = 500$ we observe that the fold lines become sharper and the fringes in the quadruple image region shrink. The normalized intensity at frequency $\nu = 500$ is very close to the normalized intensity map predicted by geometric optics.

9.4.5 The hyperbolic umbilic D_4^+

The hyperbolic umbilic D_4^+ completes the set of caustics appearing in two-dimensional lenses. It is again a singularity with co-rank 2 and co-dimension $K = 3$. The unfolding is described in terms of the three unfolding parameters (μ_1, μ_2, μ_3). We consider the integral

$$\Psi(\boldsymbol{\mu}; \nu) = \frac{\nu}{\pi} \int_{\mathbb{R}^2} e^{i\left(x_1^3 + x_2^3 - \mu_3 x_1 x_2 - \mu_2 x_2 - \mu_1 x_1\right)\nu} \mathrm{d}x_1 \mathrm{d}x_2 \,. \tag{9.74}$$

The analytic continuation of the exponent $i\phi(\boldsymbol{x}; \boldsymbol{\mu})\nu$ has four saddle points $\bar{x}_i$, given by the roots of the two quadratic equations

$$3x_1^2 - \mu_3 x_2 - \mu_1 = 0 \,, \tag{9.75}$$

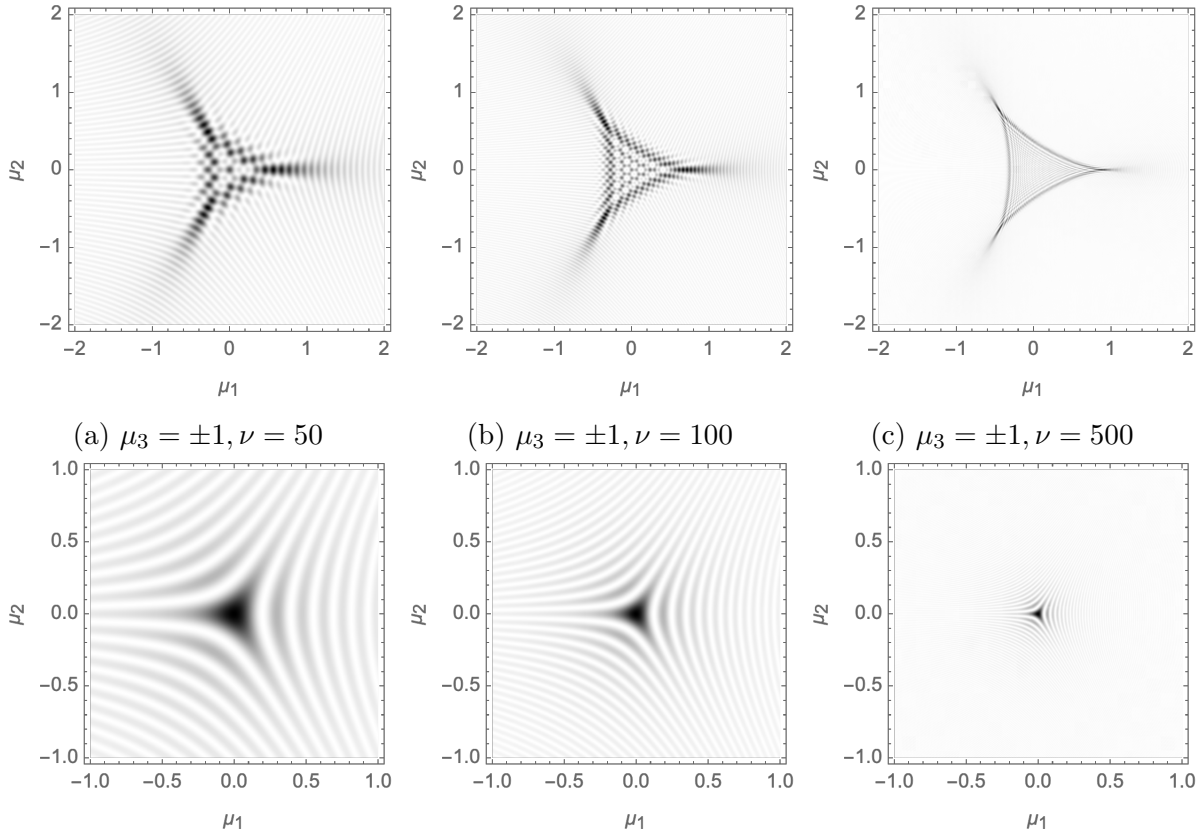
$$3x_2^2 - \mu_3 x_1 - \mu_2 = 0 \,. \tag{9.76}$$

Depending on μ, either zero, two or four of the saddle points are real-valued. The complex-valued saddle points always come in conjugate pairs since $\phi(\boldsymbol{x}; \boldsymbol{\mu})$ is real-valued for real x. Solving this set of equations for μ_1 and μ_2 we obtain the Lagrangian map as a function of μ_3,

$$\xi_{\mu_3}(x_1, x_2) = (3x_1^2 - x_2\mu_3, 3x_2^2 - x_1\mu_3) \,. \tag{9.77}$$

In the geometric limit, we form a fold-surface and a cusp lines. The fold-surface in X space is given by

$$A_2^X(\mu_3) = \left\{ \left(\pm \frac{\mu_3^2}{36t}, t \right) | t \in \mathbb{R} \right\} \tag{9.78}$$



(a) $\mu_3 = \pm 1, \nu = 50$ (b) $\mu_3 = \pm 1, \nu = 100$ (c) $\mu_3 = \pm 1, \nu = 500$

(d) $\mu_3 = 0, \nu = 50$ (e) $\mu_3 = 0, \nu = 100$ (f) $\mu_3 = 0, \nu = 500$

Figure 9.22: The normalized intensity, $I(\boldsymbol{\mu}; \nu)$, of the unfolding of the elliptic umbilic caustic (D_4^-) sliced by the surfaces $\{\mu_3 = \pm 1\}, \{\mu_3 = 0\}$ (respectively the upper and lower panels) for the frequencies $\nu = 50, 100$ and 500 (respectively the left, the centre and right panels).

which is a cylinder with radius $\frac{\mu_3}{3}$, satisfying the equation

$$|\mathcal{M}| = 0 \tag{9.79}$$

where the deformation tensor is given by

$$\mathcal{M} = \left[\frac{\partial^2 \phi(\boldsymbol{x};\boldsymbol{\mu})}{\partial x_i \partial x_j}\right]_{i,j=1,2} \tag{9.80}$$

$$= \begin{pmatrix} 6x_1 & -\mu_3 \\ -\mu_3 & 6x_2 \end{pmatrix} . \tag{9.81}$$

The three cusp-lines are linear lines laying on the fold-surface,

$$A_3^X(\mu_3) = \{(-\mu_3/6, -\mu_3/6)\} \tag{9.82}$$

in the $X = \mathbb{R}^2$ space.

In the parameter space M, the elliptic umbilic point is located at the origin. The fold-surface is given by

$$A_2 = \left\{\left(3u^4 \pm 6uv^3, \pm 6u^3 v + 3v^4, \mp 6uv\right) | u, t \in \mathbb{R}\right\} \tag{9.83}$$

$$A_2 = \left\{\left(3u^4 \mp 6uv^3, \mp 6u^3 v + 3v^4, \mp 6uv\right) | u, t \in \mathbb{R}\right\} \tag{9.84}$$

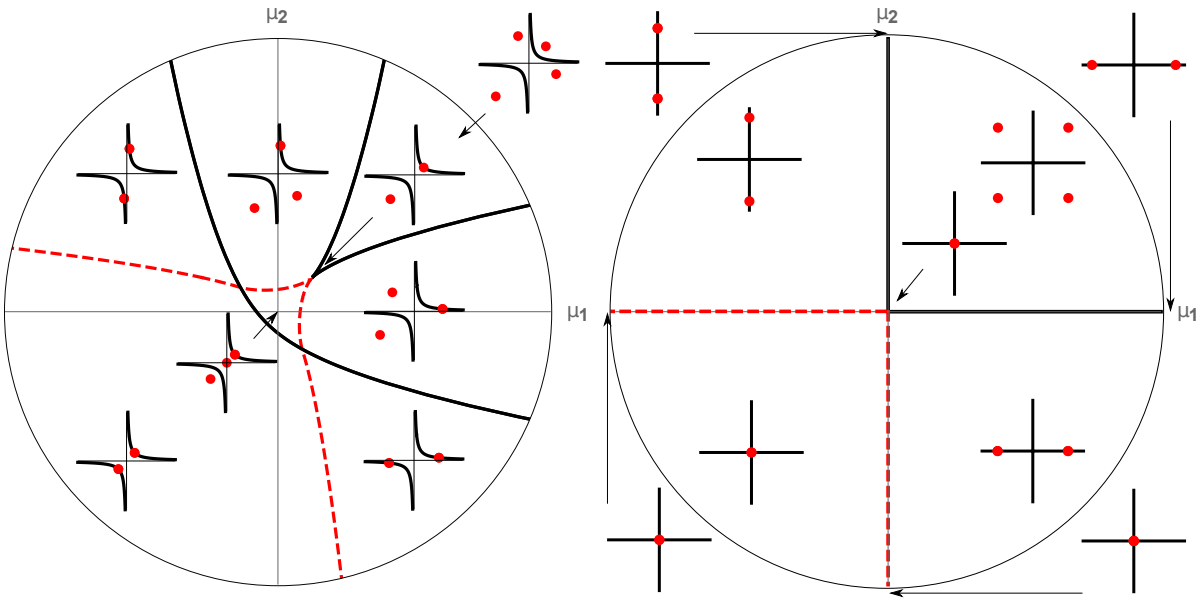
where the two solutions correspond to two disconnected pieces corresponding to the two eigenvalue fields of $\mathcal{M}$. The cusp line in the parameter space is given by

$$A_3 = \left\{(t^2/4, t^2/4, t) | t \in \mathbb{R}\right\} . \tag{9.85}$$

The fold-surface and cusp-line are illustrated in Fig. 9.20b. The surfaces stretching from the upper left to lower right and upper right to lower left denote the fold surfaces corresponding to the eigenvalue fields λ_1 and λ_2 of the deformation tensor $\mathcal{M}$. The fold surface has a harp edge at the cusp lines (in black).

Note the symmetry of the triangular singularity and point symmetry of the caustic. By performing the Picard-Lefschetz analysis for the two slices $\mu_3 = \pm 1$ and $\mu_3 = 0$ we can obtain an understanding of the relevant saddle points.

See Fig. 9.23 for the Picard-Lefschetz analysis of the two slices. The small diagrams are the real parts of the four saddle points in the (x_1, x_2) plane. The black circle is caustic in X space at $\mu_3 = \pm 1$ and 0.



(a) $\mu_3 = \pm 1$ (b) $\mu_3 = 0$

Figure 9.23: Intersection of the hyperbolic umbilic D_4^+ with the surface $\{\mu_3 = -1\}$. The heavy line is the fold-line and the dashed line is the Stokes line. The points represent the four saddle points. Note that the saddle points away from the black lines are real-valued. The saddle points on the black curves are complex-valued and come in conjugate pairs. The number of real saddles gives the number of images in the geometric optics approximation.

- Consider the slice $\mu_3 = \pm 1$ (see Fig. 9.23a). In the upper right corner, the four saddle points are real-valued. In the corresponding Picard-Lefschetz analysis, they are all relevant. This is a quadruple-image region. When we pass the left or lower fold-line, two of the four saddle points merge at the hyperbola in X space, to form a fold singularity. Afterward, both saddle points become complex. The one with the smallest real part of the exponent $i\phi(\bar{\boldsymbol{x}}; \boldsymbol{\mu})\nu$ remains relevant whereas the other saddle point becomes irrelevant. Just like in the elliptic umbilic, the real part of the complex saddle points remains on the hyperboloid. This is a double-image region. Depending on whether we cross the fold line to the left or below the quadruple region, two different saddle points merge.

  If we move from the quadruple-image region to the cusp, we obtain a singularity due to the merger of three saddle points. After passing through the cusp, only the two real saddle points will be relevant. The two complex saddle points are irrelevant.

From the double-image region, we can pass the second fold-line. At this fold-line, the two remaining real saddle points merge to form a fold saddle point after which they move in the complex plane. Note that the real parts of these two saddle points remain on the second branch of the hyperbolic. Since the Picard-Lefschetz analysis does not contain any real-valued saddle points after passing the second fold-line, this is a zero-image region. The zero-image region is again subdivided into three subregions. In the upper left and lower right regions, the Picard-Lefschetz analysis consists of two relevant complex saddle points. In the lower left region, the Picard Lefschetz analysis consists of one relevant complex saddle point.

- In the case, $\mu_3 = 0$ the analysis is similar to the one obtained for $\mu_3 = \pm 1$, since the regions are trivially deformed (see Fig. 9.23b). In the upper right region, again four saddle points are real. All of them are relevant. This is still a quadruple-image region. The fold line along the positive μ_1 and μ_2 axis is double fold lines, as the two fold lines at $\mu_3 = \pm 1$ have merged. The left and lower right regions are zero-image regions. In the upper left and lower right regions, the Picard-Lefschetz analysis consists of two relevant complex saddle points. In the lower left region, the Picard-Lefschetz analysis again consists of one relevant complex saddle point. This concludes the Picard-Lefschetz analysis.

Numerics

Given the Lefschetz thimble, we can numerically compute the normalized intensity map (see Fig. 9.24). The upper and lower panels depict the normalized intensity $I(\boldsymbol{\mu}; \nu)$ for $\mu_3 = \pm 1$ and 0. The left, central and right panels depict the different frequencies $\nu = 50, 100$ and 500.

For both unfoldings at $\mu_3 = \pm 1$ and $\mu_3 = 0$, the normalized intensity map closely follows the caustics structure represented in Fig. 9.23. In the zero-image region, the normalized intensity vanishes. In the double-image regions, for $\mu_3 = \pm 1$, the normalized intensity oscillates forming lines of equal normalized intensity as should be expected from the presence of the left fold line. In the quadruple-image regions, the normalized intensity oscillates in two directions, for $\mu_3 = \pm 1$ forming the structure we observed for the cusp caustic, and for $\mu_3 = 0$ forming an interference pattern with rectangular symmetry.

In the eikonal limit $\nu \to \infty$, the normalized intensity becomes sharper and the caustics become more pronounced. It should, in particular, be noted that the normalized intensity

at the hyperbolic umbilic (in the origin in the plots corresponding to $\mu_3 = 0$), the normalized intensity rises rapidly with ν. This is in correspondence with the scaling relations we found above.

9.5 Two dimensional localized lenses

The seven elementary singularities form a dictionary of the local behavior of the lens integral

$$\Psi(\boldsymbol{\mu};\nu) = \left(\frac{\nu}{\pi}\right)^{N/2} \int_{\mathbb{R}^N} e^{i\phi(\boldsymbol{x};\boldsymbol{\mu})\nu} \mathrm{d}\boldsymbol{x}\,, \tag{9.86}$$

$$\phi(\boldsymbol{x};\boldsymbol{\mu}) = (\boldsymbol{x}-\boldsymbol{\mu})^2 + \varphi(\boldsymbol{x})\,, \tag{9.87}$$

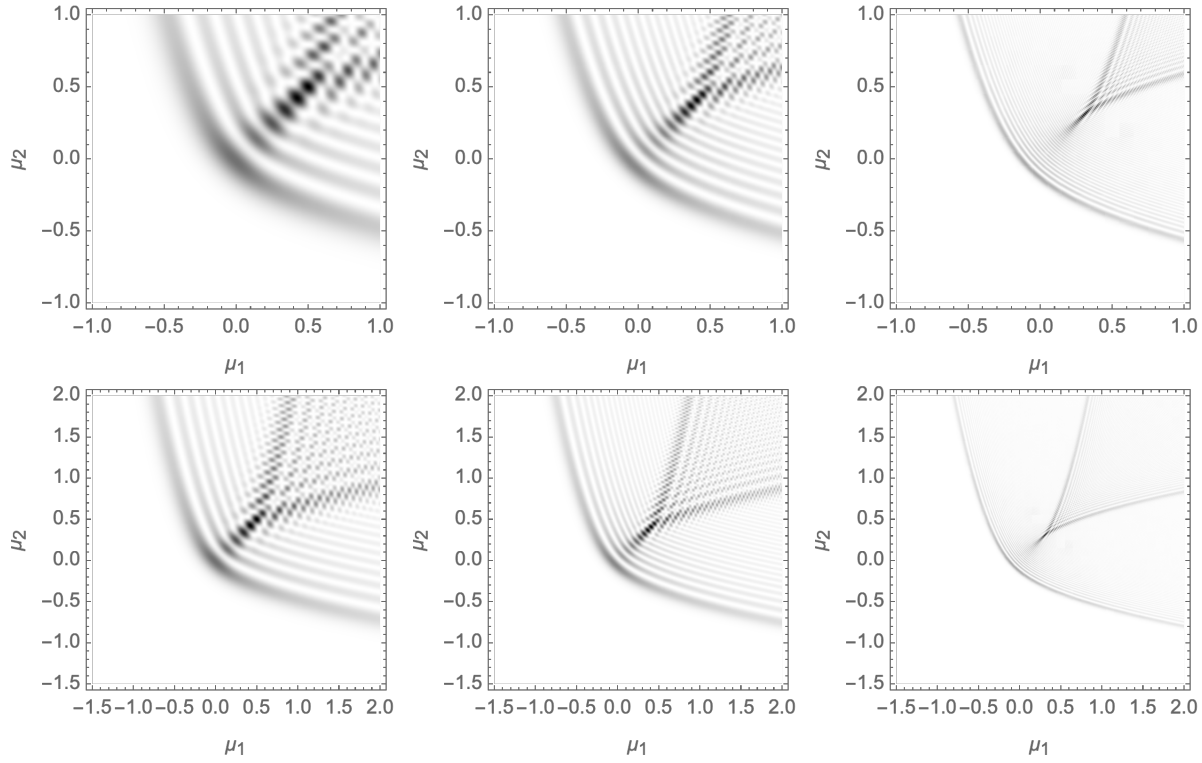
near caustics. Their corresponding normalized intensity map completely describes the local properties of lensed images. However, the global structure of the caustic is in general different. Since the normal forms of the elementary singularities are polynomials, the corresponding phase φ has support throughout the base space $X = \mathbb{R}^N$. The catastrophes with an even co-dimension K, lead to an image with an even number of images. In contrast, localized lenses lead to n-image regions with n an odd integer. We now turn to the study of interference patterns appearing in localized lenses near caustics. We evaluate three two-dimensional lenses, which simulate the behavior of a localized lens and include the five elementary catastrophes appearing in two-dimensional lenses. In the process, we also demonstrate the accuracy of the integration scheme along the Lefschetz thimble.

9.5.1 A generic peak

In general, lensing effects are strongest near the extrema of the variation of the phase φ. It is for this reason natural to study the effect of an asymmetric peak in the phase variation φ, with

$$\varphi(\boldsymbol{x}) = \frac{\alpha}{1 + x_1^2 + 2x_2^2}\,, \tag{9.88}$$

the two-dimensional generalization of the one-dimensional lens studied in Section 9.3. For astrophysical plasma lenses, the parameter α scales according to the dispersion relation $\alpha \propto \nu^{-2}$.



(a) $\mu_3 = \pm 1, \nu = 50$ (b) $\mu_3 = \pm 1, \nu = 100$ (c) $\mu_3 = \pm 1, \nu = 500$

(d) $\mu_3 = 0, \nu = 50$ (e) $\mu_3 = 0, \nu = 100$ (f) $\mu_3 = 0, \nu = 500$

Figure 9.24: The normalized intensity, $I(\boldsymbol{\mu}; \nu)$, of the unfolding of the hyperbolic umbilic caustic (D_4^+) sliced by the surfaces $\{\mu_3 = \pm 1\}, \{\mu_3 = 0\}$ (respectively the upper and lower panels) for the frequencies $\nu = 50, 100$ and 500 (respectively the left, the centre and right panels).

The Lagrangian map is given by

$$\xi(\boldsymbol{x}) = \boldsymbol{x} + \frac{1}{2}\nabla\varphi(\boldsymbol{x}) \tag{9.89}$$

$$= \boldsymbol{x} - \frac{\alpha}{(1+x_1^2+2x_2^2)^2}(x_1, 2x_2)\,. \tag{9.90}$$

The map forms a caustic where the deformation tensor

$$\mathcal{M}_{ij} = \frac{\partial^2\phi(\boldsymbol{x};\boldsymbol{\mu})}{\partial x_i \partial x_j}\,, \tag{9.91}$$

with the eigenvalue and eigenvector fields $\lambda_i(x), v_i(x)$, is singular, *i.e.*,

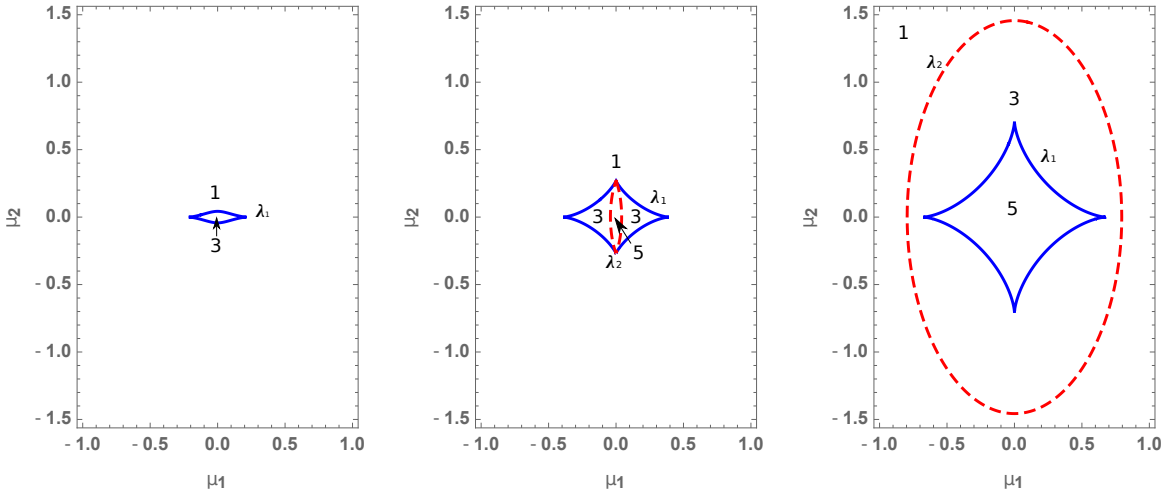
$$|\mathcal{M}(\boldsymbol{x})| = \lambda_1(\boldsymbol{x})\lambda_2(\boldsymbol{x}) = 0\,. \tag{9.92}$$

For convenience, we order the eigenvalue and eigenvector fields by $\lambda_1(\boldsymbol{x}) \leq \lambda_2(\boldsymbol{x})$.

The first caustic forms at the origin $(\mu_1, \mu_2) = (0, 0)$ for the parameter $\alpha = \frac{1}{2}$ (see Fig. 9.25). This is a cusp singularity. Note that by construction this caustic corresponds to the eigenvalue field λ_1. For $\frac{1}{2} < \alpha < \frac{64}{49}$ the A_3 point forms an outgoing fold-line (A_2) with two cusps (A_3) on the left and the right. At $\alpha = 1$, a new A_3 point is created, this time corresponding to the second eigenvalue field λ_2. For $1 < \alpha < \frac{64}{49}$ the A_3 point forms a fold-line (A_2) with two cusps (A_3) at the top and the bottom. At $\alpha = \frac{64}{49}$ the two fold lines merge in a hyperbolic umbilic (D_4^+) at $(\mu_1, \mu_2) = (0, \pm 1/\sqrt{14})$. For $\alpha > \frac{64}{49}$ the two fold lines continue to move outwards, where the fold-line corresponding to λ_1 has four cusps while the fold-line corresponding to λ_2 does not contain a cusp. Outside the fold-line of the caustics, the image consists of a single-image region. Inside the solid fold line, we find a triple-image and a five-image region enclosed by the dashed fold line.

The analytic continuation of the exponent, $\phi(\boldsymbol{x};\boldsymbol{\mu})$, possesses a pole on the two-dimensional surface $x_1^2+2x_2^2+1 = 0$. Note that poles are never isolated in multi-dimensional complex analysis [282]. The exponent has nine saddle points $\bar{\boldsymbol{x}}_i$. By evaluating the gradient of the h-function and flowing the original integration domain, we obtain a numerical representation of the thimble $\mathcal{J} \subset \mathbb{C}^2$.

Given the two-dimensional thimble $\mathcal{J}$, we numerically evaluate the integral $\Psi(\boldsymbol{\mu};\nu)$. In Fig. 9.26, we plotted the normalized intensity of the sensed signal for $\alpha = 7/10, 64/49$ and 4 as a function of the frequency. Observe that when the wavelength is comparable to the size of the caustic structure, the normalized intensity is blurred. The caustics emerge when the wavelength becomes shorter. At the frequency $\nu = 500$, we accurately recover



(a) $\alpha = 0.7$ (b) $\alpha = 64/49$ (c) $\alpha = 4$

Figure 9.25: The caustics corresponding to λ_1 (heavy line) and λ_2 (dashed line) as a function of α. We observe the formation of a triple- and a five-image region.

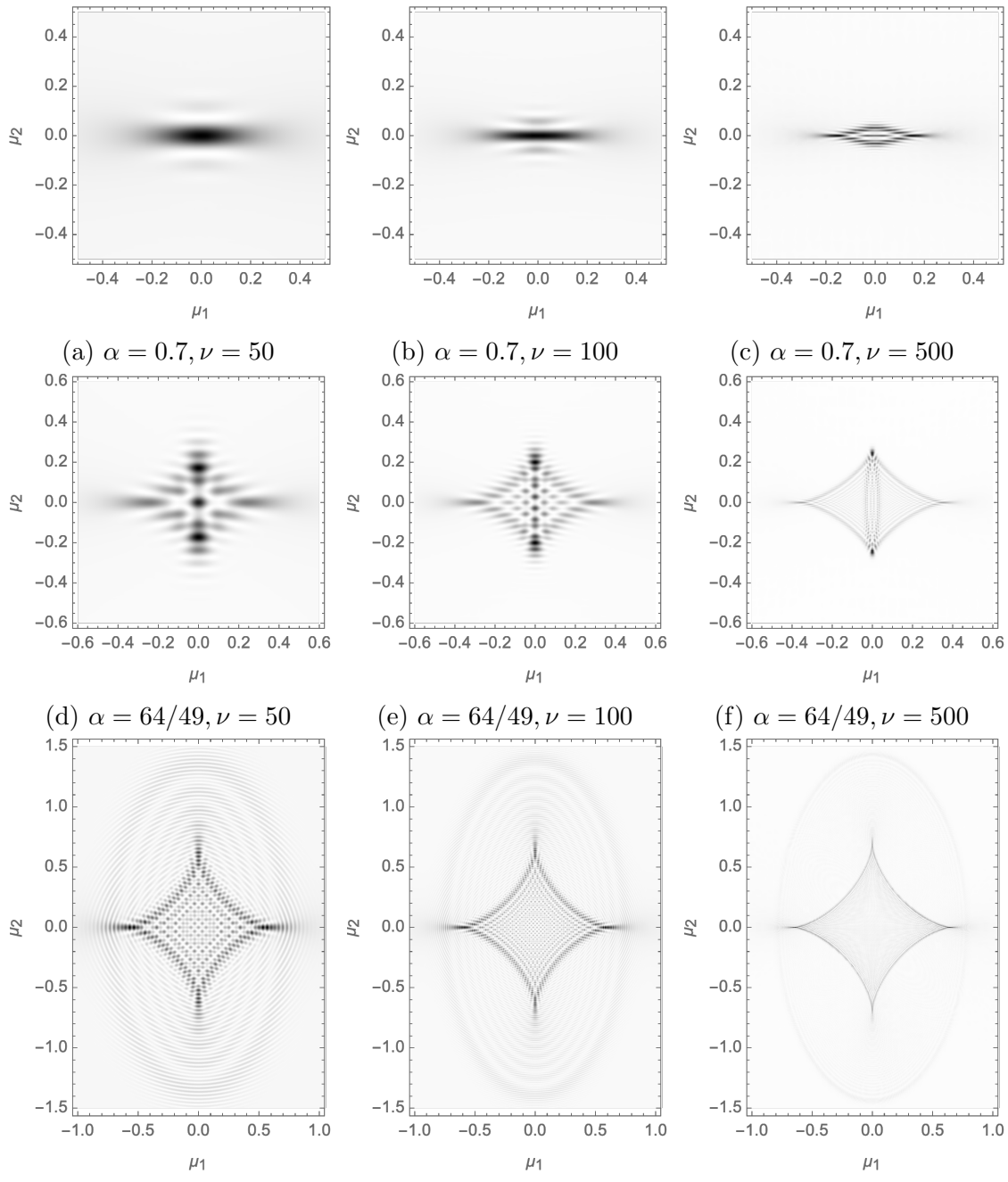
the image corresponding to geometric optics. Remark the stripes emanating from the cusp singularities. This is the frequency independent power-law falloff we observed in the elementary singularities.

In Fig. 9.27, we plot the cross-section of the normalized intensity map along the diagonal $\mu_1 = \mu_2$ for the lens with $\alpha = 4$ for $\nu = 50, 100$ and 500. Observer the four spikes while passing through the fold catastrophe. Note that the spikes increase in magnitude as ν is raised. In the astronomical context, these spikes correspond to amplification in the light-curve of the lensed source.

9.5.2 A degenerate peak

A more intricate structure arises for the lens corresponding to the degenerate peak in the phase,

$$\varphi(\boldsymbol{x}) = \frac{\alpha}{1 + x_1^4 + x_2^2}\,, \tag{9.93}$$

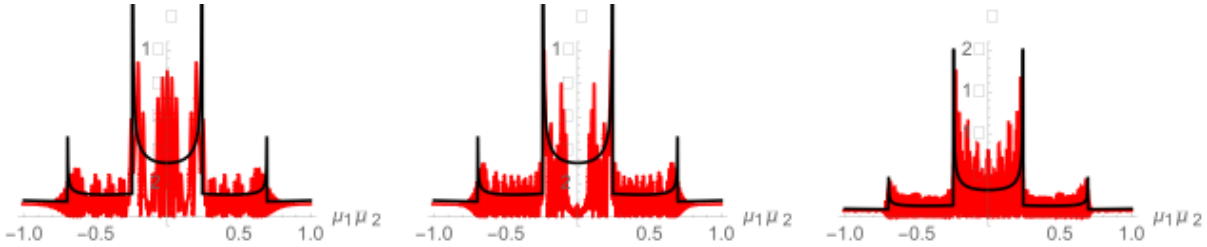


(a) $\alpha = 0.7, \nu = 50$ (b) $\alpha = 0.7, \nu = 100$ (c) $\alpha = 0.7, \nu = 500$

(d) $\alpha = 64/49, \nu = 50$ (e) $\alpha = 64/49, \nu = 100$ (f) $\alpha = 64/49, \nu = 500$

(g) $\alpha = 4, \nu = 50$ (h) $\alpha = 4, \nu = 100$ (i) $\alpha = 4, \nu = 500$

Figure 9.26: Intensity $I(\boldsymbol{\mu}; \nu)$ of the local lens at $\alpha = 0.7, 64/49, 4$ for $\nu = 50, 100$ and 500.



(a) $\nu = 50$ (b) $\nu = 100$ (c) $\nu = 500$

Figure 9.27: The normalized intensity evaluated along the diagonal in the (μ_1, μ_2)-plane for $\alpha = 4$ for $\nu = 100$ and 500. The heavy curve is the envelope predicted by geometric optics (see Eq. (9.18)).

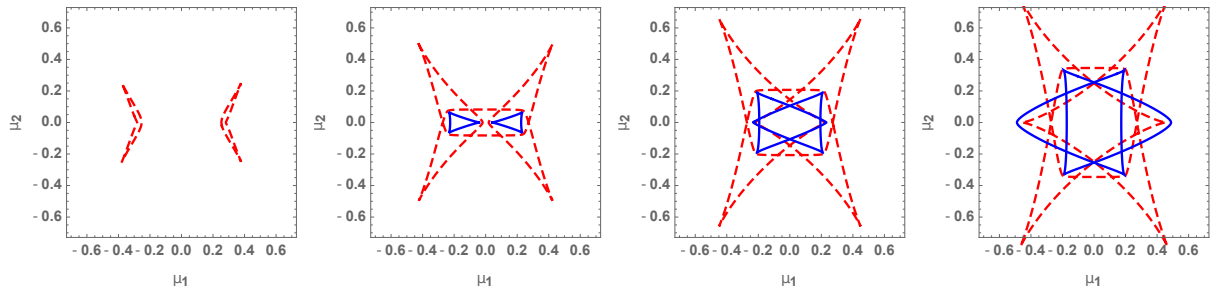
with the Lagrangian map

$$\xi(\boldsymbol{x}) = \boldsymbol{x} + \frac{1}{2}\nabla\varphi(\boldsymbol{x}) \tag{9.94}$$

$$= \boldsymbol{x} - \frac{\alpha}{(1 + x_1^4 + x_2^2)^2}(2x_1^3, x_2)\,. \tag{9.95}$$

The caustics structure of the Lagrangian map for varying α is plotted in Fig. 9.28. For $\alpha = 1$ we find two disconnected components, which are joined at $\alpha = 1.5$ and form an intricate pattern at $\alpha = 2$ and $\alpha = 2.5$. At $\alpha = 2$ we again find a hyperbolic umbilic caustic (D_4^+) at the two points where the cusps corresponding to the first and second eigenvalue fields λ_1, λ_2 coincide. We thus see that not only the structure at the peak but also the falloff of the variation in the phase φ is important in the study of caustics in lensed images. The caustic structure is generally sensitive to the Hessian of the phase φ, *i.e.*, the second order derivatives.

After flowing the original integration contour to the Lefschetz thimble $\mathcal{J}$, we numerically evaluate the amplitude $\Psi(\boldsymbol{\mu}; \nu)$ and the corresponding normalized intensity $I(\boldsymbol{\mu}; \nu)$. The resulting normalized intensity maps are plotted in figures 9.29 and 9.30. For the frequency $\nu = 50$, the image is again rather blurry. We can see the general shape, but cannot distinguish the detailed line structure. For the frequency $\nu = 100$, the lines are better resolved. However, the length scale of the caustics is comparable to the length scales of the interference patterns in the multi-image regions. For $\nu = 500$, we see the complete geometric structure of the caustics. The oscillations in the multi-image regions are now very fine. For this frequency, we are very close to the geometric optics approximation.



(a) $\alpha = 1$ (b) $\alpha = 1.5$ (c) $\alpha = 2$ (d) $\alpha = 2.5$

Figure 9.28: The caustics of the Lagrangian map for various α. The caustics corresponding to the first and second eigenvalue fields λ_1 and λ_2 are shown as dashed and solid lines respectively.

9.5.3 The swallowtail caustic

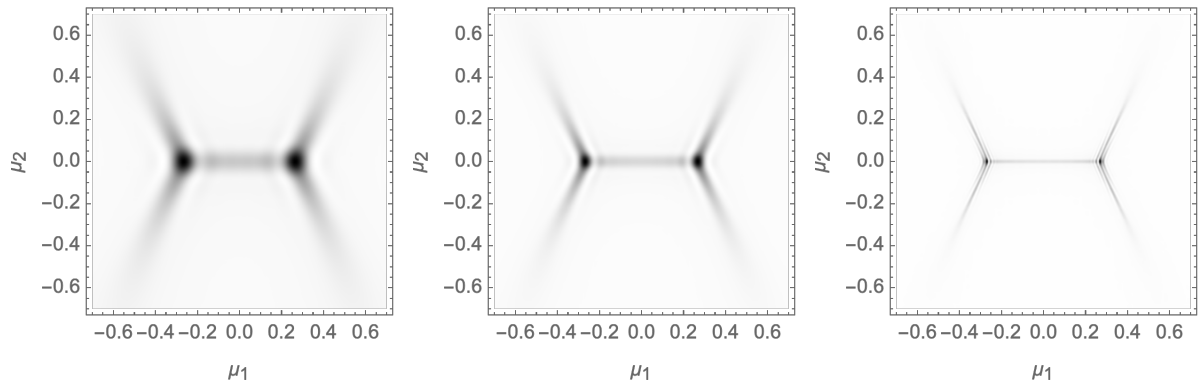
In the previous two examples of lenses corresponding to the simple peaks, we found both fold (A_2) and cusp caustics (A_3) corresponding to a single eigenvalue field, and the interaction between two eigenvalue fields via the hyperbolic umbilic (D_4^+). The two remaining caustics, *i.e.*, the swallowtail (A_4) and the elliptic umbilic (D_4^-), appear in slightly more involved lenses. For the swallowtail caustic, consider the lens

$$\varphi(\boldsymbol{x}) = \frac{\alpha x_1}{1 + x_1^4 + x_2^2}\,. \tag{9.96}$$

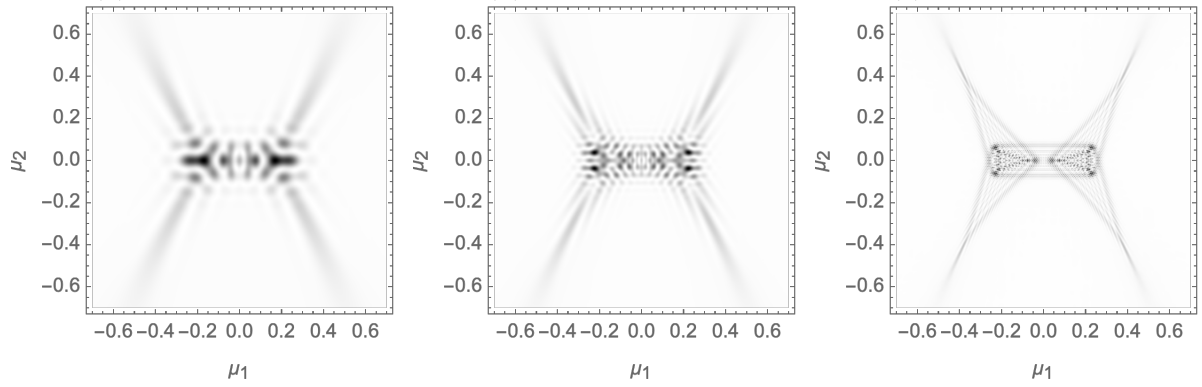
Again, in the astrophysical context, α follows the dispersion relation $\alpha \propto \nu^{-2}$.

The corresponding integrand $i\phi(\boldsymbol{x};\boldsymbol{\mu})\nu$, consists of 23 saddle point in the complex plane. By deforming the integration domain to the thimble, we evaluate the two-dimensional lens integral numerically. See figures 9.31 and 9.32 for the caustics obtained from geometric optics and the corresponding normalized intensity maps for the frequency $\nu = 50, 100, 500$.

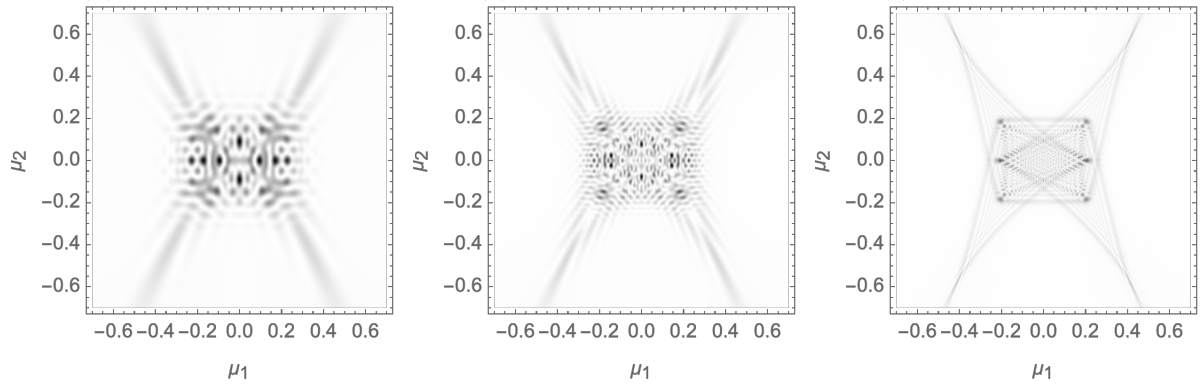
- For $\alpha = 2$, the lens forms a caustic corresponding to a single eigenvalue field (see the upper panels of Fig. 9.31.). The profile a pancake with two cusps at the tips. In the corresponding normalized intensity field, we see an interference pattern in the triple-image region, two stripes emanating from the cusps and more strikingly two diagonal stripes going to the left in the single image region. These stripes are a precursor of the swallowtail caustic emerging at later α.



(a) $\alpha = 1, \nu = 50$ (b) $\alpha = 1, \nu = 100$ (c) $\alpha = 1, \nu = 500$

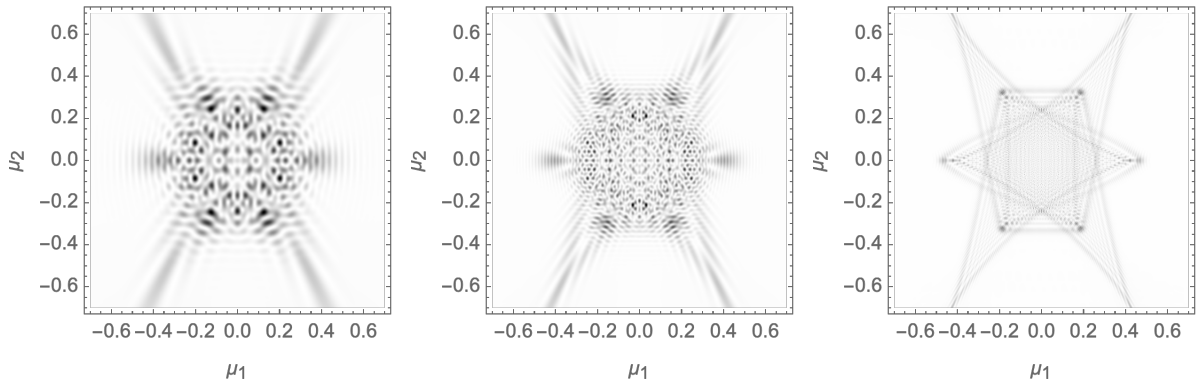


(d) $\alpha = 1.5, \nu = 50$ (e) $\alpha = 1.5, \nu = 100$ (f) $\alpha = 1.5, \nu = 500$



(g) $\alpha = 1, \nu = 100$ (h) $\alpha = 0.7, \nu = 500$ (i) $\alpha = 2, \nu = 500$

Figure 9.29: The normalized intensity map, $I(\boldsymbol{\mu}; \nu)$, for different frequencies.



(a) $\alpha = 2.5, \nu = 50$ (b) $\alpha = 2.5, \nu = 100$ (c) $\alpha = 2.5, \nu = 500$

Figure 9.30: The normalized intensity map, $I(\boldsymbol{\mu}; \nu)$, for different frequencies.

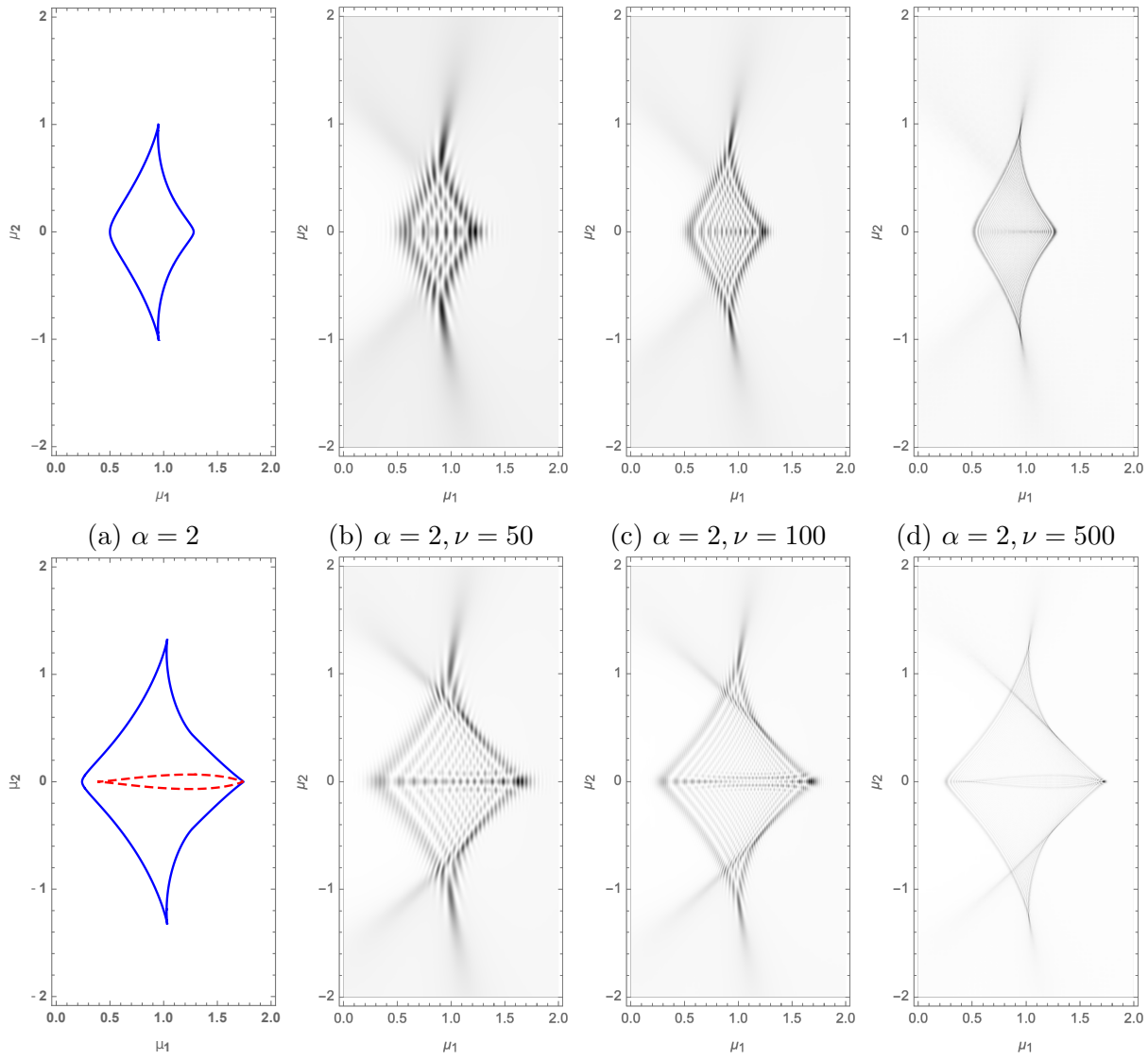
- As α is raised to 3, a second caustic emerges in the triple-image region (see the lower panels of Fig. 9.31). This caustic corresponds to the second eigenvalue field of the deformation tensor. At $\alpha = 3$ one of the two cusps of the second fold line merges with the outer fold line and transfers the cusp singularity via an elliptic umbilic caustic (D_4^-). For larger α, the solid line will thus have three cusps whereas the dashed line has only one.

  However, more importantly, the lens forms a swallowtail caustic (A_4) in the solid line at $\alpha = 2$. This phenomenon cannot be observed in the solid fold-line but is apparent in the normalized intensity map. The two stripes already visible for $\alpha = 2$ are amplified. At the location where the swallowtail stripe coincides with the fold-line, we see an amplification of the normalized intensity in the swallowtail point.

  In the normalized intensity map, we see that the geometry becomes sharper and sharper as we increase the frequency and approach the geometric optics limit. Note that the normalized intensity of the hyperbolic umbilic (D_4^+) outshines the other caustics at frequency $\nu = 500$.

- Finally, for $\alpha = 4$, we see that the swallowtail caustic has unfolded into its characteristic shape in the solid fold-line (see Fig. 9.32). We see the same structure emerge in the normalized intensity map. However, in addition, we how to obtain a large number of stripes emanating from the cusp caustics.

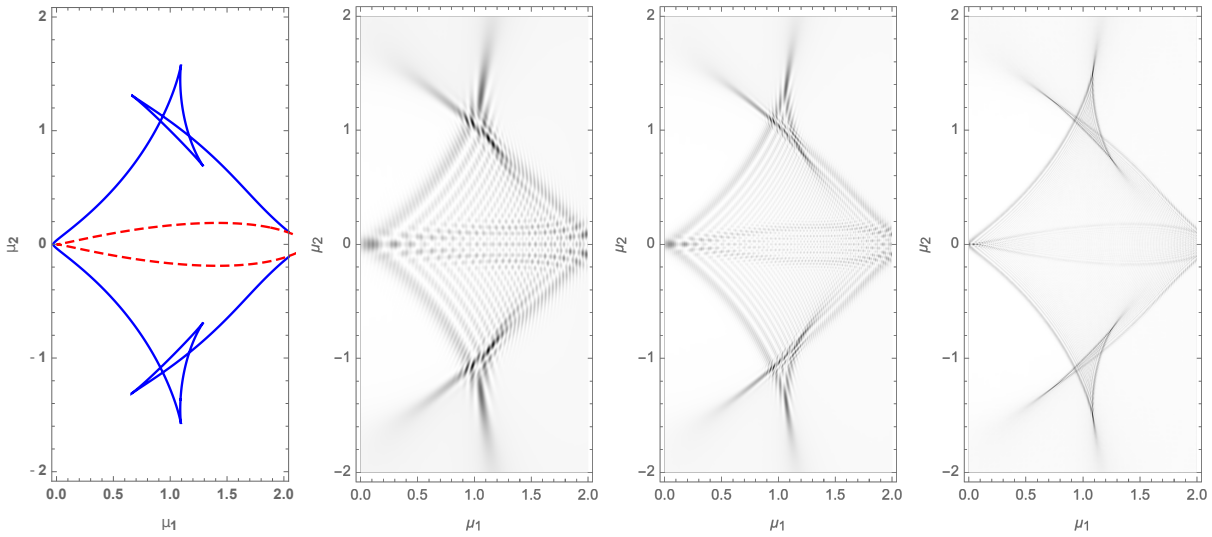
  We also see that the lens at $\alpha = 4$, consists of a second hyperbolic umbilic (D_4^+) appearing at the origin, where the solid and the dashed fold-lines meet. As the frequency is raised, we again see that the normalized intensity spikes for this caustic.



(a) $\alpha = 2$ (b) $\alpha = 2, \nu = 50$ (c) $\alpha = 2, \nu = 100$ (d) $\alpha = 2, \nu = 500$

(e) $\alpha = 3$ (f) $\alpha = 3, \nu = 50$ (g) $\alpha = 3, \nu = 100$ (h) $\alpha = 3, \nu = 500$

Figure 9.31: The normalized intensity map, $I(\boldsymbol{\mu}; \nu)$, for $\alpha = 2, 3$ and frequencies $\nu = 50, 100, 500$.



(a) $\alpha = 4$ (b) $\alpha = 4, \nu = 50$ (c) $\alpha = 4, \nu = 100$ (d) $\alpha = 4, \nu = 500$

Figure 9.32: The normalized intensity map, $I(\boldsymbol{\mu}; \nu)$, for $\alpha = 4$ and frequencies $\nu = 50, 100, 500$.

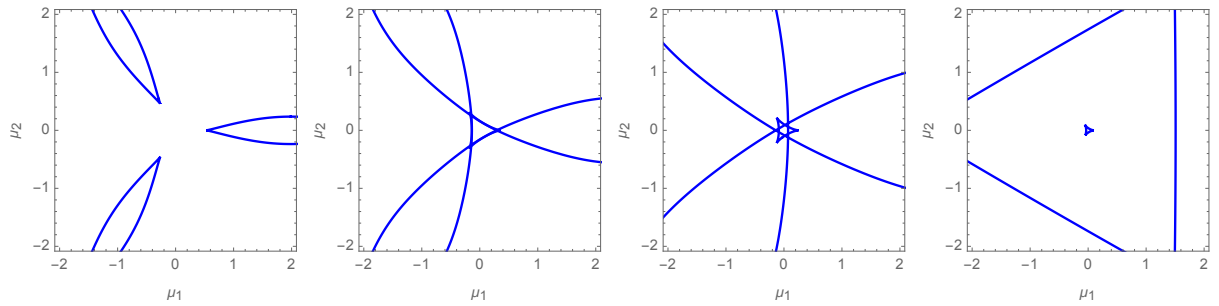
9.5.4 The elliptic umbilic caustic

We conclude this section by studying the elliptic umbilici (D_4^-) caustic in a localized lens. The elliptic umbilic forms when the deformation tensor is singular due to two eigenvalues vanishing simultaneously. The geometry of the caustic however differs from the hyperbolic umbilic (D_4^+), in that it includes the merger of three cusp caustics. We here study the localized lens

$$\varphi(\boldsymbol{x}) = \frac{\alpha(x_1^3 - 3x_1x_2^2)}{1 + x_1^2 + x_2^2}\,. \tag{9.97}$$

From geometric optics, we observe the caustic structure of the lens (see Fig. 9.33).

- For small, $\alpha < 1.4$, the lens consists of three Zel'dovich pancakes with a triangular symmetry. Three of the cusp caustic point to the origin of the parameter space.
- At $\alpha = 1.4$, we observe that the three Zel'dovich pancakes are joined by three fold-lines forming a triangular structure.



(a) $\alpha = 1$ (b) $\alpha = 1.4$ (c) $\alpha = 2$ (d) $\alpha = 5$

Figure 9.33: The caustics of the Lagrangian map for varying α.

- As $\alpha > 1.4$, the triangle decouples from the three Zel'dovich pancakes. The three resulting fold lines move away from the origin and the triangle shrinks to a point. The point a which the triangle is contracted to a point is the elliptic umbilic caustic. The region enclosed by the large triangle is a 5-image region. The region enclosed by the small triangle is a 7-image region.

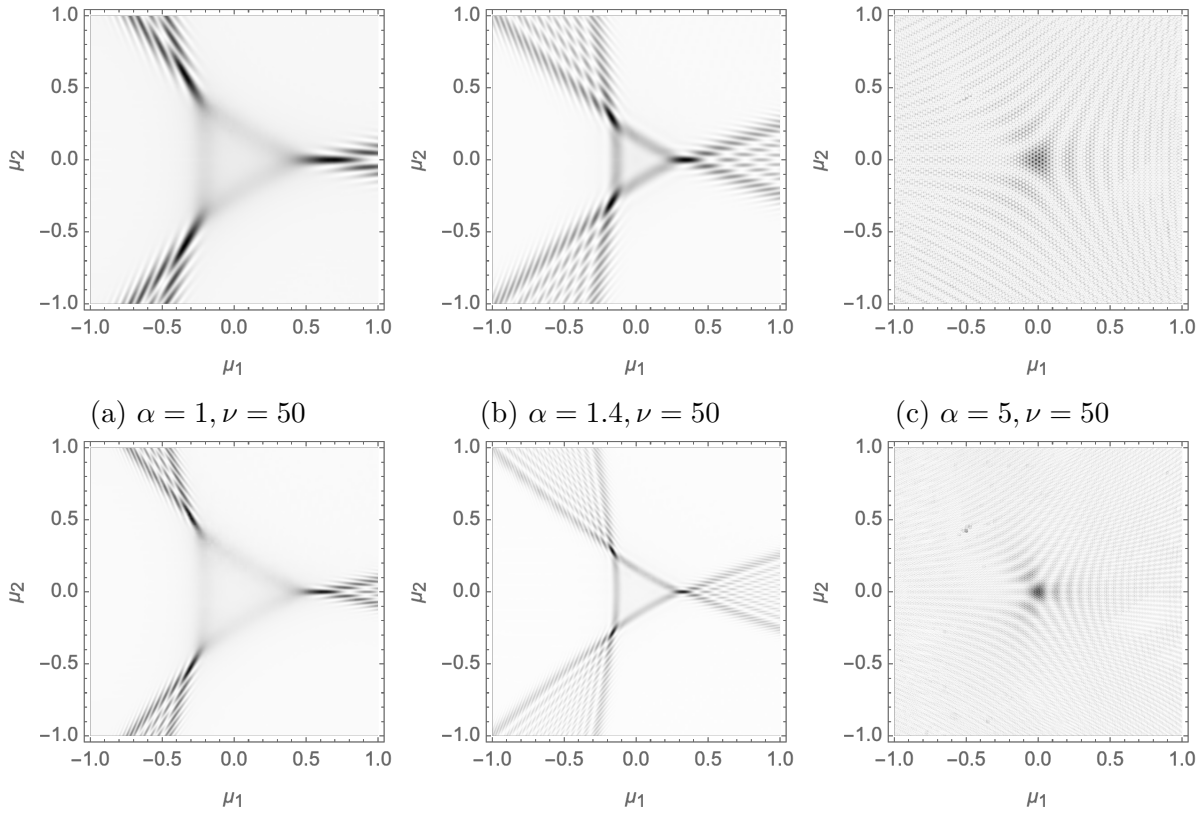
Note that since the elliptic umbilic caustic only forms after three cusp caustics have formed a triangular fold line, the caustic will be rare in simple simple lenses. It is nonetheless a stable configuration, as a small deformation of the lens preserves the structure.

Using the Picard-Lefschetz analysis, we evaluate the normalized intensity map for the configurations $\alpha = 1, 1.4$, and 5 for the frequencies $\nu = 50, 100$ (see Fig. 9.34).

- For $\alpha = 1$, we observe that even though the triangular structure is not yet present in the geometric optics analysis, it is present in the normalized intensity map at finite frequency (see the left panels of Fig. 9.34). That is to say, the normalized intensity is enhanced at the triangle, however as $\nu \to \infty$ the normalized intensity at the triangle will remain finite.

- At $\alpha = 1.4$, the triangle has formed in the geometric optics analysis (see the central panels of Fig. 9.34). In the normalized intensity maps, the triangle is enhanced. The normalized intensity will now diverge in the geometric optics limit.

- As α is further increased to $\alpha = 5$, the triangle shrinks to a point and interference effects between the different fold lines start to appear (see the right panels of Fig. 9.34). At $\alpha = 5$, we do no longer observe the fold lines but rather observe a triangular blob at the origin of the parameter space. This closely resembles the



(a) $\alpha = 1, \nu = 50$ (b) $\alpha = 1.4, \nu = 50$ (c) $\alpha = 5, \nu = 50$

(d) $\alpha = 1, \nu = 100$ (e) $\alpha = 1.4, \nu = 100$ (f) $\alpha = 5, \nu = 100$

Figure 9.34: The normalized intensity map, $I(\boldsymbol{\mu}; \nu)$, for $\alpha = 1, 1.4, 5$ and frequencies $\nu = 50, 100$.

normalized intensity map of the elementary elliptic umbilic catastrophe. It is however a bit more intricate as a close inspection demonstrates that caustic structure oscillates at a high frequency due to the interference of the elliptic umbilic with the surrounding multi-image region.

Note that there are a few small numerical artefacts present in the normalized intensity map for the lens at $\alpha = 5$. The lens outside of the triangle, is a 5-image region in which some of the real saddle points are located far away from the origin in the lens plane. The inside of the triangular region is a 7-image region. The Lefschetz thimble has a complicated shape and the tessellation of the thimble can occasionally miss a few points.

9.6 Signatures of caustics in fast radio bursts

A Fast Radio Burst (FRB) is a millisecond transient radio pulse, caused by some yet to be identified high-energy astrophysical process. The first burst was found by Duncan Lorimer and his student David Narkevic in 2007 while scanning through archival pulsar survey data [242]. The burst in question had been detected in 2001 by the Parkes Observatory in Australia. In subsequent years, several other bursts were observed, among which the first repeating source (named FRB 121102) [301] was detected in 2012 by the Arecibo Observatory in Puerto Rico. In the last few months, several new detections have been announced by the Canadian Hydrogen Intensity Mapping Experiment (CHIME) collaboration, including the second repeating FRB source (named FRB 180814) [75]. FRBs are now known to be relatively common, with approximately $10,000$ bright fast radio bursts occurring per day over the entire sky. Telescopes capable of detecting a significant fraction of these bursts should become possible in coming decades, an exciting prospect indeed.

The source of fast radio bursts is yet to be identified. Many different models have been proposed but none is yet compelling. They range from rapidly spinning neutron stars or black holes and regions of very high electromagnetic fields, to more exotic sources [266, 275]. It seems likely that the bursts are extragalactic in origin, as the first observed repeater, FRB 121102, has been identified with a galaxy at a distance of approximately 3 billion light years [72, 73, 251]. As mentioned in the introduction, it is likely that the phenomenology of fast radio bursts is strongly affected by astrophysical plasma lensing. They have a characteristic time-frequency profile, their frequency typically falling during the pulse, or series of pulses. This profile is probably due to the fact that lower frequencies are more strongly lensed and thus follow longer geometrical paths, and also because they propagate more slowly.

The methods and results we have reported here should be helpful in modeling the effects of plasma lensing on observed FRBs. The lensing may take place in a variety of places – near the source, near the observer or in between. If the line of sight encounters a caustic due to a plasma lens, the FRB may be amplified, enhancing the chances of detection. For reasons we have explained, caustics are likely to be localized in frequency, leading to the observed spectral shape. The "marching down" features could also be due to asymmetric structures in the lens, leading to angled caustics. This requires a preferred time asymmetry, which could in turn provide hints about the structure of the lens itself. In the lensing example of B1957+20 [244], the lens is due to a companion wind. In this specimen, the time-frequency caustics march both up and down. This symmetry could be broken if the wind contained shock waves, which could preferentially move retrograde in the rotating frame. Quantitative lens modeling can be tested on the pulsar binary system, and then applied to FRB data. This could be the scope of a future paper.

Since the observed radio waves have a relatively long wavelength, the corresponding diffraction catastrophes are likely to fill a significant volume in the parameter space of the normalized intensity maps. Therefore it is important to study the complete interference pattern. It follows from Table 9.2 that the elliptic (D_4^-), the hyperbolic (D_4^+) umbilic and to a lesser extent the swallowtail (A_4) caustic lead to the largest spikes in the normalized intensity map. Of these three caustics, the swallowtail (A_4) and the hyperbolic (D_4^+) umbilic caustics are most likely to be realized in simple lenses, of which the hyperbolic caustic gives the greatest amplification. However, these caustics will not generically occur in time-frequency data, as they are formed at point in three-dimensional functions. The line of sight, is, however, reasonably likely to pass close to them, as they fill a finite volume of the parameter space. In principle, we do expect to see the cusp (A_3) points and the fold (A_2) lines caustics, in the data. However, note that these caustics lead to a lesser amplification of the source.

As we observed in the previous sections, caustics due to multi-dimensional lenses never occur as isolated events. The caustics of co-dimension four, *i.e.*, the umbilics $D_4^{\pm}$, and the swallowtail A_4 caustics, are always accompanied by cusp (A_3) points and fold (A_2) lines. It thus follows that when a fast radio burst is indeed amplified by a lens, that the corresponding peak in time-frequency space will be of characteristic shape. More concretely, after identifying the time and the frequency with the two of the unfolding parameters μ, we expect the peak to resample the normalized intensity map of the corresponding elementary catastrophe computed in Section 9.4. That is to say, the peak corresponding to elliptic (D_4^-) umbilic caustic should exhibit a triangular symmetry and the peak corresponding to the swallow (A_4) caustic will exhibit the characteristic swallowtail geometry in the fold-line and two cusps caustics.

Further investigation is required to estimate the number density of the different caustics for generic two-dimensional lenses and the most likely normalized intensity profiles along the line of sight.

9.7 Conclusions

Conditionally convergent oscillatory integrals play a central role in modern physics. However, these integrals are often difficult to define as their definition, in the multi-dimensional case, can depend on the order of integration or the regularization scheme. They are, moreover, generically impossible to evaluate analytically and too expensive to evaluate with conventional numerical methods. In this chapter we have brought Picard-Lefschetz theory to bear. We have shown how in a multi-dimensional oscillatory integral, the integrand generically defines a set of relevant Lefschetz thimbles in the complexified integration domain, along which the integral is absolutely convergent. These thimbles can be thought of as an 'integrand-dependent Wick rotation'. The integral evaluated along the set of relevant thimbles in fact provides an unambiguous definition of the original integral itself. We moreover have presented a new, efficient numerical scheme both to find the thimbles and to efficiently evaluate the integral along them in polynomial time. The virtue of this new method that the efficiency actually *increases* as the integrand becomes more oscillatory.

In particular, we have studied the Lefschetz thimbles for caustic catastrophes and the Stokes phenomenon occurring in two-dimensional lenses. Given the thimbles, we numerically evaluate the normalized intensity maps over all frequencies study the resulting interference patterns. We have shown that the normalized intensity maps smoothly converge to the caustics predicted by geometric optics, without introducing numerical artifacts.

Our method renders feasible the calculation of interference patterns in a wide variety of interesting astrophysical contexts, in particular to model the effect of plasma lenses on radio sources. So far, such modeling has been restricted to the simplest examples of fold and cusp singularities, produced by one dimensional lenses. More realistic, two-dimensional models, including the swallowtail, elliptic umbilic and hyperbolic umbilic caustics are now accessible. We have computed the normalized intensity maps for a few representative examples, and commented briefly on likely observational signatures. A statistical analysis of the normalized intensity profiles for the diffraction catastrophes generated by a realistic plasma lens ensembles will be the subject of further investigations.

Finally, we analyzed a simple model of Young's double slit experiment, representing an initial exploration of the use of these methods for describing interference in quantum mechanics.

9.A Defining oscillatory integrals

Oscillatory integrals, which do not converge absolutely, are sometimes claimed to be ill-defined since the key theorems of measure theory, *e.g.* the dominated convergence theorem and Fubini's theorem, do not apply [115]. We here study conditionally convergent oscillatory integrals for the one- and multi-dimensional case and propose a definition using Picard-Lefschetz theory in terms of absolutely convergent ones.

9.A.1 One-dimensional integral

The Fresnel integral

$$F(\infty) = \int_{-\infty}^{\infty} e^{ix^2} \mathrm{d}x = (1+i)\sqrt{\frac{\pi}{2}} \tag{9.98}$$

exists, even though the integral is only conditionally convergent. The integral is usually defined as a limit of the partial integral

$$F(R) = \int_{-R}^{R} e^{ix^2} \mathrm{d}x, \tag{9.99}$$

i.e., $\lim_{R\to\infty} F(R) = (1+i)\sqrt{\frac{\pi}{2}}$ following the Euler or Cornu spiral (see Fig. 9.35). This definition is as important to the integral as the integrand, as different regularization schemes – which do not approach the real line by adding points incrementally – lead to different answers.

The definition of the conditionally convergent integral in terms of the limit $R \to \infty$ is equivalent to the assumption of analyticity, since Cauchy's integral theorem only applies to integrals over $\mathbb{R}$ defined this way. We can alternatively define the integral by deforming the integration contour $\mathbb{R}$ in the complex plane $\mathbb{C}$ to the Lefschetz thimble

$$\mathcal{J} = \{(1+i)u | u \in \mathbb{R}\} = (1+i)\mathbb{R}\,, \tag{9.100}$$

for which the integrand is convex and the integral is absolutely convergent, *i.e.*,

$$\int_{\mathrm{R}} e^{ix^2} \mathrm{d}x = \int_{\mathcal{J}} e^{ix^2} \mathrm{d}x \tag{9.101}$$

$$= (1+i) \int_{\mathrm{R}} e^{-2u^2} \mathrm{d}u \tag{9.102}$$

$$= (1+i)\sqrt{\frac{\pi}{2}}\,. \tag{9.103}$$

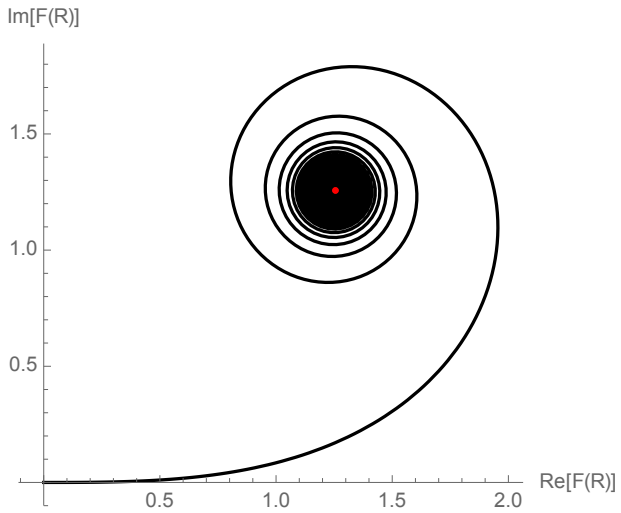


Figure 9.35: The Euler or Cornu spiral, showing the value of $F(R)$ as a function of R in the complex plane. The spiral focuses on the point $\lim_{R\to\infty} F(R) = (1+i)\sqrt{\frac{\pi}{2}}$.

Note that this definition does not depend on a limit. Defining the integral in terms of the analytic properites of the integrand has the consequence of removing any ambiguity. There is, in particular, no need for any regularization.

9.A.2 Multi-dimensional integrals

Multi-dimensional conditionally convergent oscillatory integrals such as

$$\int_{\mathbb{R}^N} e^{if(x_1,\ldots,x_N)} \mathrm{d}x_1 \ldots \mathrm{d}x_N \tag{9.104}$$

for $N \in \mathbb{N}$ and appropriate functions f, play an important role in optics but cannot be uniquely defined using an extension of the regularization scheme described above for the one-dimensional case. To show this, lets consider the two-dimensional analogue of the Fresnel integral

$$\int_{\mathbb{R}^2} e^{i(x^2+y^2)} \mathrm{d}x\mathrm{d}y\,. \tag{9.105}$$

Since this integral factorizes, it is reasonable to require the integral to converge to

$$F(\infty)^2 = \left((1+i)\sqrt{\frac{\pi}{2}}\right)^2 = i\pi\,. \tag{9.106}$$

However, for general $f(x_1, \ldots, x_n)$ we are not able to write the integral as a product of one-dimensional integrals. This thus should not be considered as a desirable definition of the integral.

To see the dependence on the regularization scheme, consider the integral in polar coordinates. We write

$$I(R) = \int_{\mathbb{D}_R} e^{i(x^2+y^2)} \mathrm{d}x\mathrm{d}y \tag{9.107}$$

$$= 2\pi \int_0^R r e^{ir^2} \mathrm{d}r \tag{9.108}$$

$$= i\pi \left(1 - e^{iR^2}\right) , \tag{9.109}$$

with $\mathbb{D}_R$ the disk of radius R centred at the origin. We thus find that the limit $\lim_{R\to\infty} I(R)$ does not exist! The function $I(R)$ instead circles the 'correct answer' $i\pi$ with increasing angular velocity.

It is instead appropriate to define the integral in terms of the Lefschetz thimble

$$\mathcal{J} = \{(1+i)(u,v) | (u,v) \in \mathbb{R}^2\} \tag{9.110}$$

$$= (1+i)\mathbb{R}^2 . \tag{9.111}$$

Along the thimble, the integral is absolutely convergent

$$\int_{\mathbb{R}^2} e^{i(x^2+y^2)} \mathrm{d}x\mathrm{d}y = \int_{\mathcal{J}} e^{i(x^2+y^2)} \mathrm{d}x\mathrm{d}y \tag{9.112}$$

$$= (1+i)^2 \int_{\mathbb{R}^2} e^{-2(u^2+v^2)} \mathrm{d}u\mathrm{d}v \tag{9.113}$$

$$= 2i \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-2(u^2+v^2)} \mathrm{d}u\mathrm{d}v \tag{9.114}$$

$$= i\pi . \tag{9.115}$$

On the thimble we can safely convert the integral over the real plane $\mathbb{R}^2$ into the iterative integral using Fubini's theorem, since the integral in u and v over $\mathbb{R}^2$ is absolutely convergent. This definition straightforwardly generalizes to general multi-dimensional conditionally convergent integrals.

9.B Young's double-slit experiment

In this appendix, we generalize our treatment of interference in order to tackle Young's famous double slit experiment. In spite of the extreme simplicity of this example, and its centrality to introductory discussions of quantum physics, detailed interference patterns are surprisingly hard to compute. By generalizing our treatment of the Fresnel-Kirchhoff integral we shall be able to efficiently study the pattern created by a pair of smooth, finite size slits in detail. In particular, we shall see how quantum interference effects disappear in the classical limit, as $\hbar$ is taken to zero.

The generalization required is to make the interference "phase" complex in order to damp out the amplitude away from two narrow slits. Modeling this complex phase with a simple rational function, our numerical techniques allow us to efficiently find the relevant Lefschetz thimbles and compute the detailed interference pattern at all values of the parameters.

Consider a distant point source emitting particles towards a screen, with a thin barrier separating the screen from the source. The barrier is opaque to the particles except in the neighbourhood of two slits. In dimensionless coordinates (which we shall define below), the transmission amplitude takes the form

$$T(x) \propto \exp\left[\frac{\epsilon}{\epsilon^2+(x-s_1)^2}+\frac{\epsilon}{\epsilon^2+(x-s_2)^2}-\frac{1}{\epsilon}\right], \tag{9.116}$$

consisting of two peaks each of strength unity, centered respectively at $x = s_1$ and $x = s_2$. Here, $\epsilon > 0$ is a small number representing both the width of the slits (see Fig. 9.36) and the opacity of the barrier: away from the slits, the latter is given by $T \sim \exp(-1/\epsilon)$.

Assuming the incident amplitude for the particles to be coherent and constant across the slits, we may then compute the path integral amplitude just as in Section 9.1. Here, however, we deal with a particle of fixed mass m, energy E and momentum $p = \sqrt{2mE}$. The last formula in (9.8), in the same small displacement-approximations made in Eq. (9.12) above, yields a Pythagorean contribution to the phase, $p(x-\mu)^2/(2d\hbar)$ where d is the distance from the slits to the screen. Setting $x \to xa$ where a is the characteristic dimension of the slits and x is dimensionless, we take the quantity $\hbar 2d/(pa^2)$ to be our new, dimensionless $\hbar$. In terms of these dimensionless quantities, the amplitude for the particle to arrive at position μ on the screen is therefore given by the oscillatory integral

$$\Psi(\mu) = \mathcal{N}\int e^{\frac{i}{\hbar}(x-\mu)^2}T(x)\mathrm{d}x \tag{9.117}$$

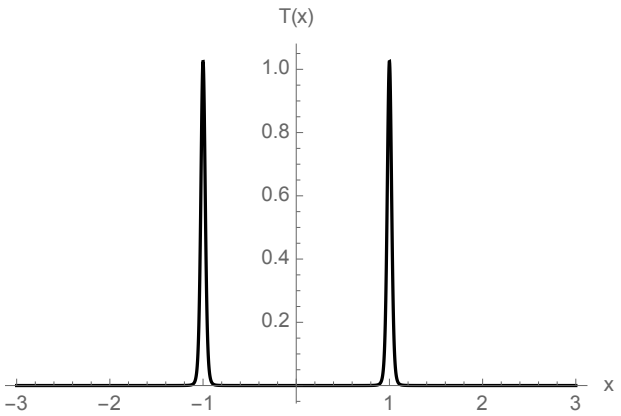


Figure 9.36: The real transmission amplitude of the wall $T(x)$ with two slits at $s_1 = -1$, $s_2 = 1$ with width $\epsilon = 0.1$.

with the normalization constant $\mathcal{N}$, ensuring unitarity $\int |\Psi(\mu)|^2 \mathrm{d}\mu = 1$. The probability for the particle to arrive at μ on the screen is given by the absolute square of the wavefunction

$$I(\mu) = |\Psi(\mu)|^2 \,. \tag{9.118}$$

Note that the dimensionless version of Planck's constant $\hbar$ appears in this nonrelativistic problem, whereas it cancelled out of our earlier formulae for a massless particle, as a result of the latter's scale covariance.

We evaluate the wavefunction (equation (9.117)), by analytically continuing the exponent

$$\phi(x;\mu) = \frac{i}{\hbar}(x-\mu)^2 + \frac{\epsilon}{\epsilon^2 + (x-s_1)^2} + \frac{\epsilon}{\epsilon^2 + (x-s_2)^2} - \frac{1}{\epsilon}\,, \tag{9.119}$$

in the complex plane and evaluating the Lefschetz thimble. The exponent, ϕ, has four poles and nine saddle points. The poles at $x = s_i \pm i\epsilon$ correspond to the slit centereed at s_i. The saddle points are roots of a nineth order polynomial. We can associate four saddle points to each slit. The remaining saddle point is shared and moves between the the poles corresponding to the two slits as a function of the position on the screen μ.

Fig. 9.37 shows the corresponding Picard-Lefschetz diagrams for various positions μ for $\hbar = 1$. In the description we will for simplicity assume the left slit to be at s_1 and the right slit to be at s_2, *i.e.*, $s_1 < s_2$:

- For positions on the screen far to the left of the slits, $\mu \ll s_1$, the thimble consists of five steepest descent contours. The thimble runs from the lower left to the upper right

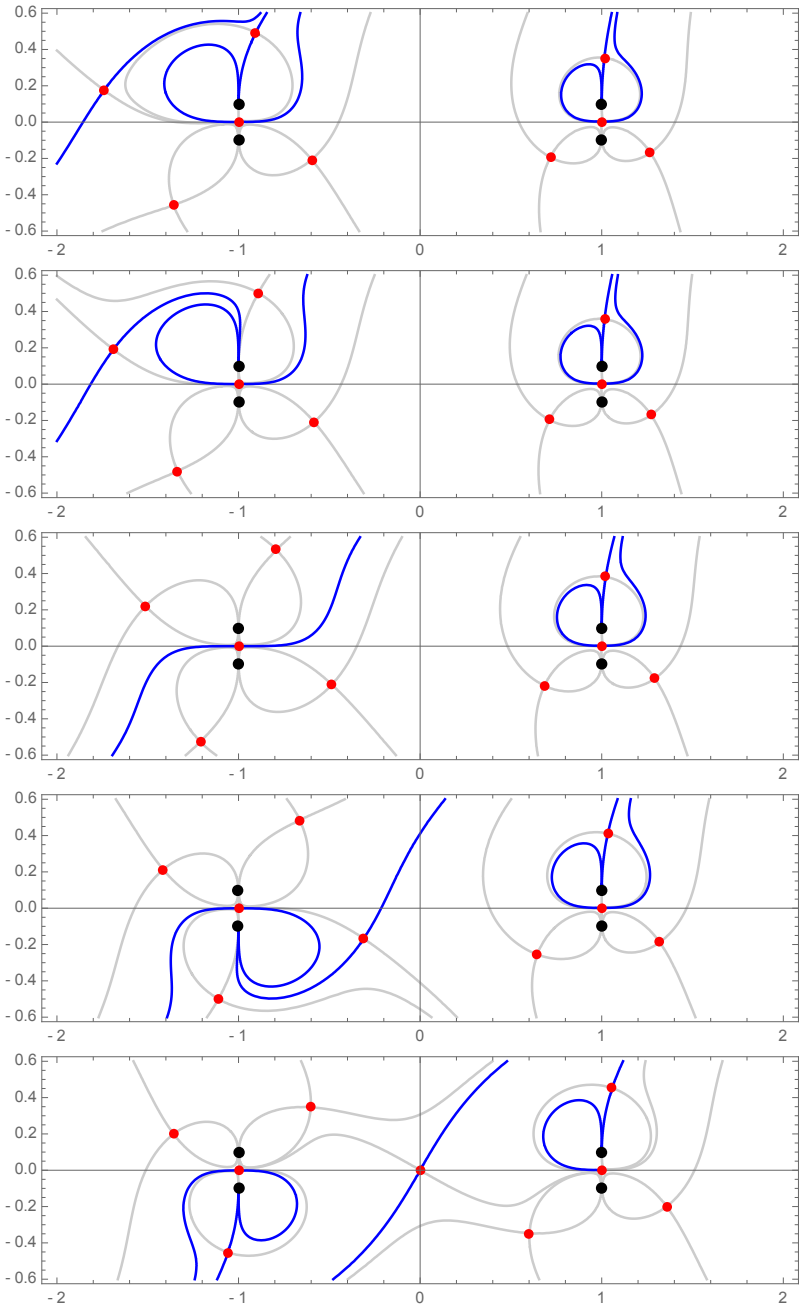


Figure 9.37: The Picard-Lefschetz diagram for Young's double slit experiment for $\epsilon = 0.1$ from $\mu = -1.6, -1.5, -1, -0.5$, and $\mu = 0$. The steepest ascent and descent contours are shown as curves, with the relevant Lefschetz thimbles shown as heavier curves. Saddle points and singularities are shown as points. Plot shown for $\hbar = 1$.

via a complex saddle point. The thimble subsequently loops around the upper left and the upper right poles. For positions $\mu \leq \frac{s_1+s_2}{2}$, the wavefunction is dominated by the left slit. It is for this reason not surprising to see that the thimble corresponding to the right slit is representative in this regime.

- As μ approaches s_1, we observe a Stokes transition after which only four saddle points are relevant. The thimble moves from the lower-left via a saddle point to the upper left pole, after which it passes through the saddle point between the two left poles. The right part of the thimble is largely unchanged.

- For μ near s_1, we observe yet another Stokes transition after which only three saddle points remain relevant. The thimble runs from the lower left via the saddle point between the to left poles to the upper right.

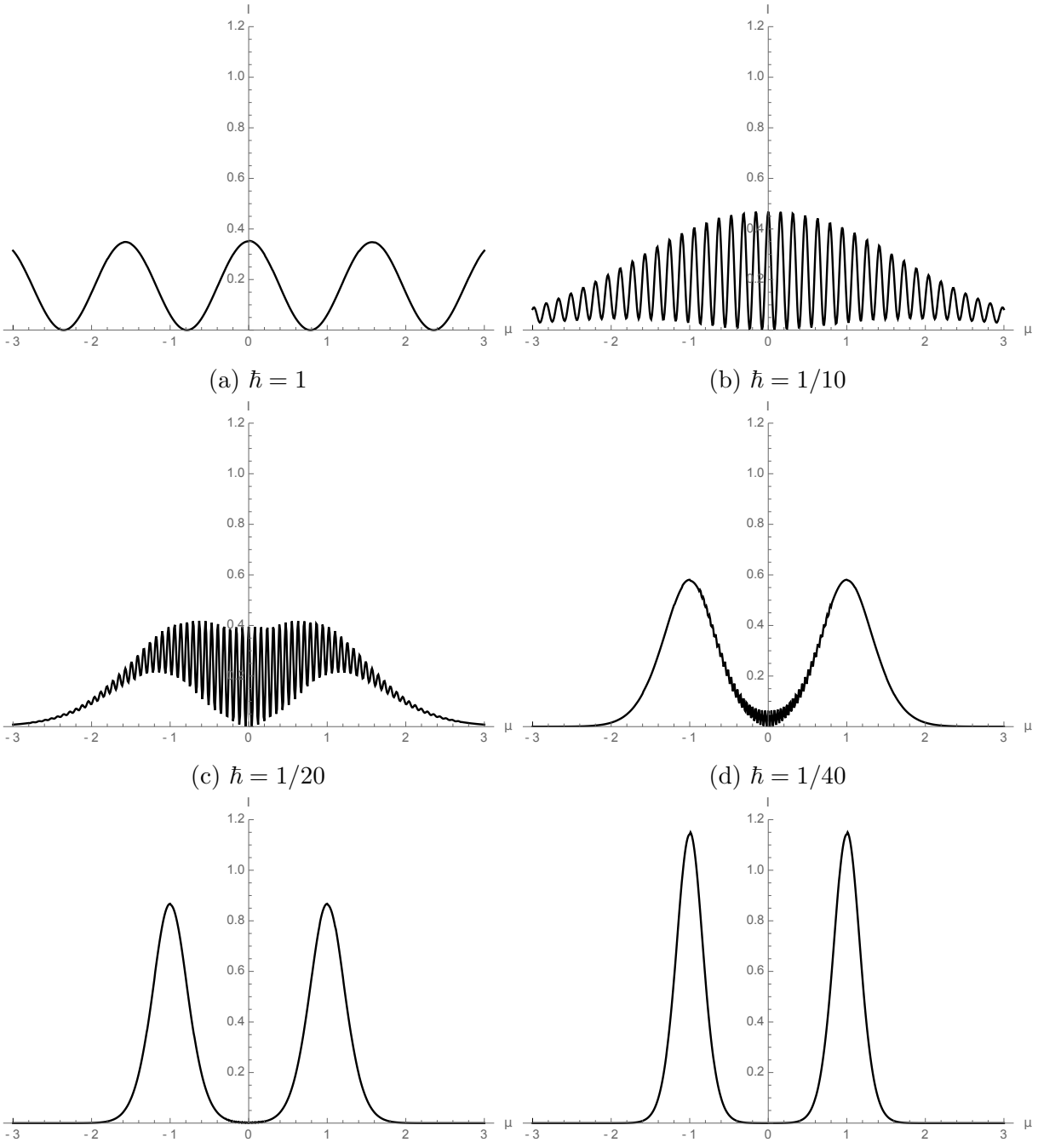
- When μ approaches the mid-point $\frac{\mu_1+\mu_2}{2}$, we observe that a complex saddle point becomes relevant after a Stokes phenomenon. The thimble now consists of four steepest descent contours.

- For μ near the mid-point $\mu = \frac{\mu_1+\mu_2}{2}$, we observe that after yet another Stokes transition, we obtain a thimble consisting of five steepest descent contours. Note that the middle saddle point has moved to the origin $x = 0$. When the position μ is increased further, this saddle point will move to the poles corresponding to the right slit. The corresponding Picard-Lefschetz diagrams are mirror images of the ones discussed above.

In the semi-classical limit $\hbar \to 0$, the geometry of the Lefschetz thimble is to an increasing extent determined by the Pythagorean term in equation (9.119). As a consequence, after a few Stokes transitions, the eight saddle points which can be associated to the poles corresponding to the two slits become tighter bound to the poles representing the geometry of the right part of the thimble in Fig. 9.37. The remaining saddle point still moves between the poles corresponding to the two slits. However, note that the integral is increasingly dominated by the two saddle points between the four poles. These two saddle points approach the real line at $x = s_1$ and $x = s_2$ in this limit.

Given the thimble, we can efficiently evaluate the oscillatory integral for various $\hbar$ (see Fig. 9.38). For relatively large $\hbar$, the intensity on the screen is dominated by interference effects. For both $\hbar = 1$ and $\hbar = 1/2$ we do not observe the classical intensity peaks corresponding to the two slits. In the semi-classical limit, $\hbar \to 0$, we the interference pattern is slowly replaced by the classical peaks. Note that this transition from the quantum to the classical regime cannot be studied in the traditional thin slit approximation.

Observe that, while the behavior of strong lenses is dominated by caustics where the saddle points become degenerate, the qualitative behavior of the double-slit experiment is completely determined by the Stokes transitions. The saddle points are everywhere non-degenerate and the h-function is a Morse function. In both instances, the saddle point approximation fails and the integral should be evaluated along the complete Lefschetz thimble. We expect this to be a generic feature in quantum mechanical interference phenomena.



(a) $\hbar = 1$

(b) $\hbar = 1/10$

(c) $\hbar = 1/20$

(d) $\hbar = 1/40$

(e) $\hbar = 1/60$

(f) $\hbar = 1/80$

Figure 9.38: The intensity I as a function of position μ for various $\hbar$

Chapter 10

Caustic skeleton

Scientific progress is the discovery of a more and more comprehensive simplicity.
Georges Lemaître

Abstract

We present a general formalism for identifying the caustic structure of a dynamically evolving mass distribution, in an arbitrary dimensional space. The identification of caustics in fluids with Hamiltonian dynamics, viewed in Lagrangian space, corresponds to the classification of singularities in Lagrangian catastrophe theory. On the basis of this formalism we develop a theoretical framework for the dynamics of the formation of the cosmic-web, and specifically those aspects that characterize its unique nature: its complex topological connectivity and multiscale spinal structure of sheetlike membranes, elongated filaments and compact cluster nodes. Given the collisionless nature of the gravitationally dominant dark matter component in the universe, the presented formalism entails an accurate description of the spatial organization of matter resulting from the gravitationally driven formation of cosmic structure.

The present work represents a significant extension of the work by Arnol'd et al. [20], who classified the caustics that develop in one- and two-dimensional systems that evolve according to the Zel'dovich approximation. His seminal work established the defining role of emerging singularities in the formation of nonlinear structures in the universe. At the transition from the linear to nonlinear structure evolution, the first complex features emerge

at locations where different fluid elements cross to establish multistream regions. Involving a complex folding of the 6-D sheetlike phase-space distribution, it manifests itself in the appearance of infinite density caustic features. The classification and characterization of these mass element foldings can be encapsulated in *caustic conditions* on the eigenvalue and eigenvector fields of the deformation tensor field.

In this study we introduce an alternative and transparent proof for Lagrangian catastrophe theory. This facilitates the derivation of the caustic conditions for general Lagrangian fluids, with arbitrary dynamics. Most important in the present context is that it allows us to follow and describe the full three-dimensional geometric and topological complexity of the purely gravitationally evolving nonlinear cosmic matter field. While generic and statistical results can be based on the eigenvalue characteristics, one of our key findings is that of the significance of the *eigenvector field* of the deformation field for outlining the entire spatial structure of the *caustic skeleton* emerging from a primordial density field.

In this chapter we explicitly consider the caustic conditions for the three-dimensional Zel'dovich approximation, extending earlier work on those for one- and two-dimensional fluids towards the full spatial richness of the cosmic-web. In an accompanying publication, we apply this towards a full three-dimensional study of caustics in the formation of the cosmic-web and evaluate in how far it manages to outline and identify the intricate skeletal features in the corresponding N-body simulations.

10.1 Introduction

Caustics[1] that emerge in fluid flows are best studied in a Lagrangian space. They are important features, marking the positions where fluid elements cross and multi-stream regions form. These caustics can be associated to the regions with infinite density, corresponding to locations where shell-crossing occurs. In the present study, we concentrate specifically on the role of caustics in the formation of the cosmic-web. The gravitationally driven formation of structure in the universe is dominated by the dark matter component. Given its collisionless nature, the formalism that we present in this study entails an accurate description of the spatial structure that emanates as a result of its dynamical evolution. The emerging caustics even have a direct physical impact on the baryonic matter, given its accretion into the gravitational potential wells delineated by the evolving dark matter distribution. Notwithstanding this cosmological focus, the caustic conditions and mathematical formalism that we have derived for this are of a more generic nature, with a validity

[1]In singularity theory, a caustic is the curve of critical values the Lagrangian mapping $q \rightarrow x_t(q)$.

that extends to all systems which allow for a Lagrangian description.

The cosmic-web is the complex network of interconnected filaments and walls into which galaxies and matter have aggregated on Megaparsec scales. It contains structures from a few megaparsecs up to tens and even hundreds of megaparsecs of size. The weblike spatial arrangement is marked by highly elongated filamentary and flattened planar structures, connecting in dense compact cluster nodes surrounding large near-empty void regions. As borne out by a large sequence of N-body computer experiments of cosmic structure formation (*e.g.* [302, 331, 290]), these web-like patterns in the overall cosmic matter distribution do represent a universal but possibly transient phase in the gravitationally driven emergence and evolution of cosmic structure (see *e.g.* [12, 71]).

According to the *gravitational instability scenario* [265], cosmic structure grows from tiny primordial density and velocity perturbations. Once the gravitational clustering process has progressed beyond the initial linear growth phase, we see the emergence of complex patterns and structures in the density field. The resulting web-like patterns, outlined by prominent anisotropic filamentary and planar features surrounding characteristic large underdense void regions, are therefore a natural manifestation of the gravitational cosmic structure formation process.

The recognition of the *cosmic-web* as a key aspect of the emergence of structure in the Universe came with early analytical studies and approximations concerning the emergence of structure out of a nearly featureless primordial Universe. In this respect the Zel'dovich formalism [344] played a seminal role. The emphasis on anisotropic collapse as agent for forming and shaping structure in the Zel'dovich "pancake" picture [344, 210] was seen as the rival view to the purely hierarchical clustering view of structure formation. The successful synthesis of both elements in the *cosmic-web* theory of Bond et al. [49] appears to provide a succesful description of large scale structure formation in ΛCDM cosmology. The cosmic-web theory emphasizes the intimate dynamical relationship between the prominent filamentary patterns and the compact dense clusters that stand out as the nodes within the cosmic matter distribution [49, 77, 319]. It also implies that a full understanding of the cosmic-web's dynamical evolution is necessary to clarify how its structural features are connected in the intricate network of the cosmic-web. To answer this question we need to turn to a full phase-space description of the evolving matter distribution and mass flows.

The Zel'dovich formalism [344] already underlined the importance of a full phase-space description for understanding cosmic structure formation, however, with the exception of a few prominent studies [20], the wealth of information in the full 6-D phase-space escaped attention. This changed with the publication of a number of recent publications [2, 134, 262, 296, 281] (for an early study on this observation see [64]) in which it was

realized that the morphology of components in the evolving matter distribution is closely related to its multistream character. This realization is based on the recognition that the emergence of nonlinear structures occurs at locations where different streams of the corresponding flow field cross each other.

Looking at the appearance of the evolving spatial mass distribution as a 3D *phase space sheet* folding itself in 6D phase space, a connection is established between the structure formation process and the morphological classification of the emerging structure. Caustics, which are the subject of this study, mark the regions where the cosmic-web begins to form. Based on recent advances and insights, in this study we discuss the role of caustics in the formation of the cosmic-web. By tracing the caustics during the formation of the cosmic-web we obtain a skeleton of the current three-dimensional large scale structure.

Caustics in fluids with Hamiltonian dynamics, viewed in Lagrangian space, are classified by Lagrangian catastrophe theory [15, 17, 228, 343]. Following this, these results were soon extended to fluids with generic dynamics [60]. For the classification of caustics emerging in the context of a one- and two-dimensional description of cosmic structure formation by the Zel'dovich approximation, Arnol'd et al. [20] translated this into conditions on the displacement field. Following up on this seminal work, Hidding et al. [203] analyzed the overall morphology and connectivity of caustics that emerge in a displacement field described by the one- and two-dimensional Zel'dovich approximation. The visual illustration of the emerging structure, for a field of initially Gaussian random density and potential fluctuations, revealed how the caustics spatially outline the spine of the cosmic-web. Feldbrugge et al. [136] elaborated this into an analytical evaluation of the statistical properties of caustics, assuming a random Gaussian initial density field.

In the current study we assess the caustics emerging in a one-parameter family of sufficiently differentiable maps $x_t : \mathbb{R}^3 \mapsto \mathbb{R}^3$, mapping the initial mass distribution to the final mass distribution at time t. For practical considerations we consider the evolution of a collisionless medium of matter in 6-dimensional phase-space. The collisionless Boltzmann equation, known as the Vlasov equation, describes the development of the phase-space density $f(x, v)$ of the medium. In a gravitational field Φ, the phase-space density of mass elements with velocity v at location x evolves according to

$$\frac{\partial f}{\partial t} + v_k \frac{\partial f}{\partial x_k} - \frac{\partial \Phi}{\partial x_k} \frac{\partial f}{\partial v_k} = 0 \,. \tag{10.1}$$

While the medium strictly speaking cannot be considered as a physical fluid, in the sense of a medium characterized by continuously varying one-valued quantities in Eulerian space, we might use the term "Lagrangian fluid" or "Vlasov fluid" for the dark matter medium.

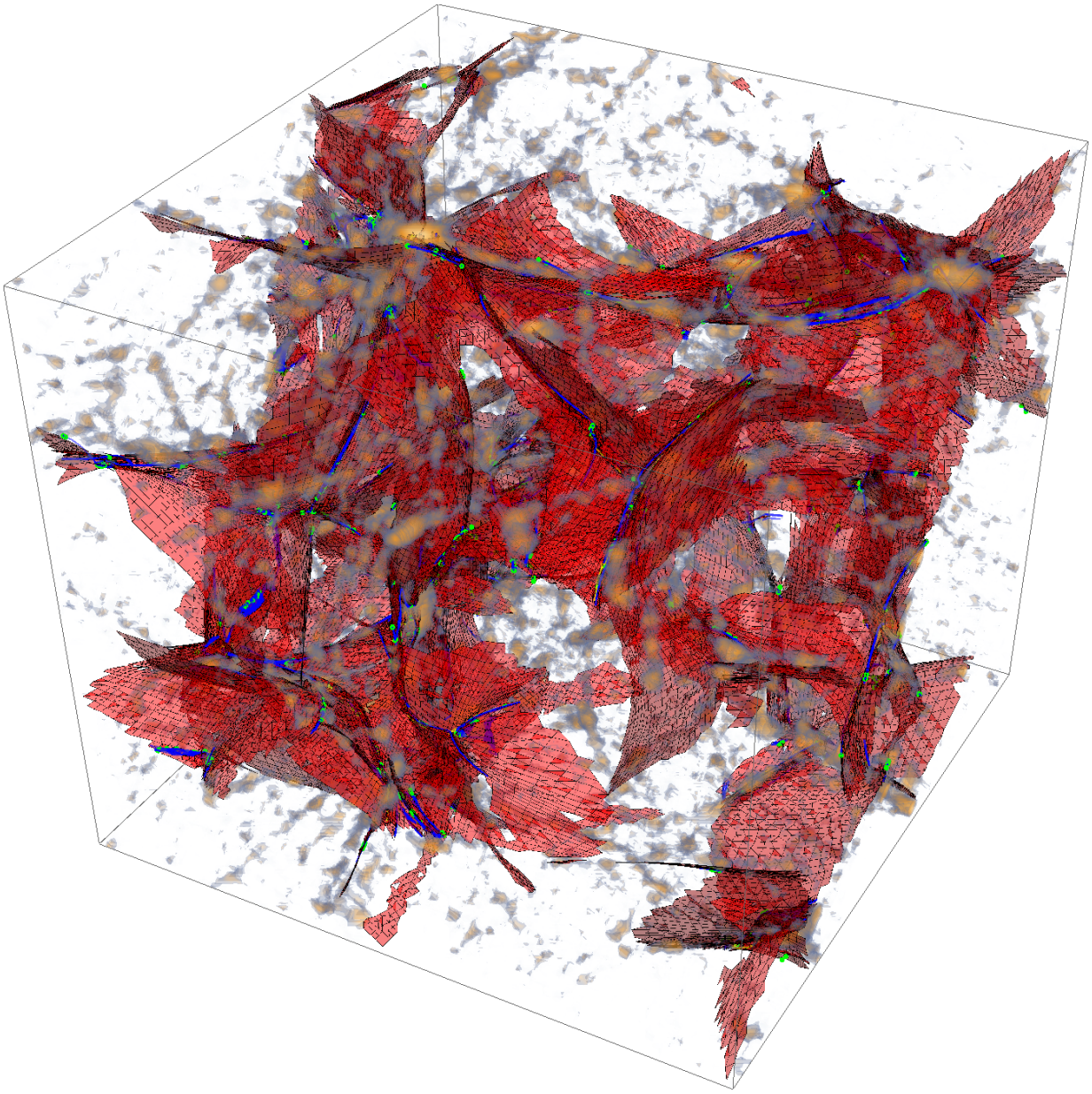


Figure 10.1: Illustration of caustic features in the cosmic-web, and its relation with the corresponding density field. The N-body simulation is a CDM simulation in an Einstein-de Sitter Universe. On the basis of the initial flow deformation field, the caustics in the matter distribution have been identified. The detailed description of these follows in section 4. The red sheets represent the cusps (A_3) singularities which correspond to the walls or membranes of the cosmic-web. The blue lines and the green points are the swallowtail (A_4) and butterfly (A_5) singularities corresponding to the filaments and clusters of the large scale structure. The dark matter distribution in the N-body simulation is represented by a log density colour scheme.

For reasons of lucidity, in the remainder of this study we denote a "Vlasov fluid" shortly as "fluid".

Within this context, we give a novel proof of Lagrangian catastrophe theory and the corresponding *caustic conditions* for three-dimensional Hamiltonian fluids. These conditions are expressed in both the eigenvalue and the eigenvector fields of the mass flow deformation tensor. Moreover, our scheme allows us to extend these caustic conditions to fluids with non-Hamiltonian dynamics. Applied to the three-dimensional Zel'dovich approximation, these conditions on the initial density field lead to a *caustic skeleton* of the cosmic-web. In this skeleton the walls, filaments and clusters of the large scale structure are directly related to the A_3, A_4, A_5, D_4 and D_5 caustics of Lagrangian catastrophe theory. See Figure 10.1 for an illustration of the caustic skeleton of the Zel'dovich approximation and a dark matter N-body simulation. A detailed analysis of the caustic skeleton of the Zel'dovich approximation and a comparison with N-body simulations is the subject of a follow-up paper [145].

It should be emphasized that the eigenvalue fields of the mass flow deformation tensor have, for a long time, been successfully used in Lagrangian studies of the cosmic-web [76, 338, 234]. In these studies, the clusters, filaments and walls are related to the number of eigenvalues exceeding a threshold. The caustic skeleton here proposed complements their work in that it include the information of the eigenvector fields, which so far has been largely neglected.

The chapter begins section 10.2 with a concise description of Lagrangian fluid dynamics. The formation of caustics and derivation of the shell-crossing conditions for the occurrence of multistream regions in a flow field is studied in section 10.3. These conditions are among the main results presented here. In section 10.4 we apply these shell-crossing conditions to the classification of catastrophes, described in section 10.5, to derive the caustic conditions. Section 10.6 discusses the relevance and significance of the caustic structure in the context of the evolving cosmic mass distribution, and in particular the emergence and morphological structure of the cosmic-web. Also, it discusses the further application and development of the caustic formalism in a cosmological context, outlining the main elements of our project. In section 10.7 we describe the dynamical framework resulting from the considerations above. Finally, in section 10.8 we summarize the results and discuss possible applications.

10.2 Lagrangian fluid dynamics

There exist multiple approaches to fluid dynamics. In the Eulerian approach, the evolution of the smoothed density and velocity fields is analyzed. The equations of motion of Eulerian fluids are relatively concise and give a reasonably accurate description of the mean flow in a fluid element at a given location in the fluid. The Lagrangian view of particle flows is the appropriate tool for following the complex dynamical evolution of fluid elements, including the evolution of multi-stream regions and the emergence of caustics, where the caustics are the critical values of the Lagrangian map.

In Lagrangian fluid dynamics, we assume every point in space to consist of a mass element that is moving with the fluid. Their motion is described by a Lagrangian map $x_t : L \to E$, mapping the initial position q in the Lagrangian manifold L to the position $x_t(q)$ of the mass element in the Eulerian manifold E at time t.[2] In the context of Lagrangian fluid dynamics, it is most convenient to describe the evolving fluid in terms of the displacement map s_t defined by,

$$s_t(q) = x_t(q) - q\,, \tag{10.2}$$

for all $q \in L$. For the Zel'dovich approximation [344] of cosmic structure formation the displacement field is given by

$$s_t(q) = -b_+(t)\nabla_q \Psi(q)\,, \tag{10.3}$$

with the growing mode b_+ and the displacement potential Ψ (appendix 10.A). The displacement potential is proportional to the linearly extrapolated gravitational potential to the current epoch ϕ_0, *i.e.*,

$$\Psi(q) = \frac{2}{3\Omega_0 H_0^2}\phi_0(q)\,, \tag{10.4}$$

with H_0 the current Hubble parameter and Ω_0 the current total energy density. In this chapter we always assume the maps x_t and s_t to be continuous and sufficiently differentiable. While in the Lagrangian description a mass element has a constant mass, it may contract, expand, deform and even rotate. This is described in terms of the deformation tensor $\mathcal{M}$, the gradient of the displacement field with respect to the Lagrangian coordinates of a mass element,

$$\mathcal{M} \;=\; \frac{\partial s_t}{\partial q} \;=\; \begin{pmatrix} M_{1,1} & M_{2,1} & M_{3,1} \\ M_{1,2} & M_{2,2} & M_{3,2} \\ M_{1,3} & M_{2,3} & M_{3,3} \end{pmatrix}. \tag{10.5}$$

[2]Note that here we do not explicitly use a distinct notation for vector quantities: q and x_t are vectors which in conventional cosmology notation are usually written as $\vec{q}$ and $\vec{x}_t$. Throughout this chapter we use the notation familiar to the mathematics literature.

While mass elements in a Lagrangian fluid are characterized by a few fundamental quantities, which characterize them and remain constant throughout their evolution, most physical properties are basically derived quantities. A good example and illustration of a derived quantity is the density field. The density in a point $x' \in E$ is defined as the initial mass in the mass element times the ratio of the initial and final volume of the mass element. Formally, this is expressed as a change of coordinates involving the Jacobian of the map x_t,

$$\begin{aligned} \rho(x',t) &= \sum_{q \in A_t(x')} \rho_i(q) \left| \frac{\partial x_t(q)}{\partial q} \right|^{-1} \\ &= \sum_{q \in A_t(x')} \rho_i(q) \left| I + \frac{\partial s_t(q)}{\partial q} \right|^{-1} . \end{aligned} \tag{10.6}$$

This can be written as

$$\rho(x',t) = \sum_{q \in A_t(x')} \frac{\rho_i(q)}{|1+\mu_{t1}(q)||1+\mu_{t2}(q)||1+\mu_{t3}(q)|} , \tag{10.7}$$

with $A_t(x')$ the points q in Lagrangian space L which map to x', *i.e.*, $A_t(x') = \{q \in L | x_t(q) = x'\}$, ρ_i the initial density field and μ_{ti} the eigenvalues of the deformation tensor $\mathcal{M}(q)$, defined by

$$\mathcal{M} v_i = \mu_i v_i \tag{10.8}$$

with eigenvector v_i. The equality in equation (10.7) applies to general deformation tensors[3], since the characteristic polynomial of the deformation tensor can be expressed in terms of the eigenvalues

$$\chi(\lambda) = \det \left[\frac{\partial s_t}{\partial q} - \lambda I \right] = (\mu_{t1} - \lambda)(\mu_{t2} - \lambda)(\mu_{t3} - \lambda) , \tag{10.9}$$

by which

$$\det \left[I + \frac{\partial s_t}{\partial q} \right] = \chi(-1) = (1+\mu_{t1})(1+\mu_{t2})(1+\mu_{t3}) . \tag{10.10}$$

[3]Note that here we use the general convention to represent the deformation eigenvalue field, with $\mu_i(q)$ the i-th eigenvalue of the deformation tensor, $\mathcal{M}(q)$. This differs from the usual convention in cosmology to use the time-independent representation of the deformation field in the context of the Zel'dovich approximation. Within this formalism, the eigenvalues $\lambda_i(q)$ of the deformation field $\psi_{ij} = \partial^2 \Psi(q)/\partial q_i \partial q_j$, are related to the eigenvalues $\mu_i(q)$ via the linear relation $\mu_i(q,t) = -b_+(t)\lambda_i(q)$, in which $b_+(t)$ is the growing mode growth factor. See Appendix A for further details.

By substituting derived quantities like density in the, often more familiar, Eulerian fluid equations, we may obtain a closed set of differential equations for the Lagrangian map x_t or the displacement map s_t. Note that for practical reasons in this chapter we will sometimes suppress the time index of the eigenvalue fields, *i.e.*, $\mu_i = \mu_{ti}$.

Equation (10.7) applies to a fluid with three spatial dimensions. For simplicity, we will restrict explicit expressions to the 3-dimensional case [4]. The arguments presented in this chapter straightforwardly generalize to a Lagrangian fluid with an arbitrary number of spatial dimension and it is straightforward to generalize equation (10.7) to d-dimensional fluids in d-dimensional space.

The appearance of singularities in equation (10.7) is central to our discussion concerning the nature of these singularities. They occur when a mass element reaches an infinite density. More formally stated, as we will see in section 10.3, an infinite density occurs when for at least one of the $i = 1, \ldots, d$,

$$1 + \mu_i = 0\,. \tag{10.11}$$

The regions, in which the mapping x_t becomes degenerate and the density becomes infinite are known as *foldings*, *caustics* or *shocks*. They mark important features in the Lagrangian fluid and are the object of study in this chapter.

While these eigenvalue conditions provide the necessary condition for a mass element to pass through a caustic, and reach infinite density, it does not yield the full information necessary to infer the geometric structure, spatial connectivity and identity of the caustic. As mass elements pass through a multistream region, the spatial properties of the flow will determine the complexity of the folding of the phase-space sheet in which they are embedded. In this study we demonstrate that the corresponding eigenvectors are instrumental in establishing the spatial outline and identify of the corresponding caustics. This key realization emanates from the so-called *caustic conditions*.

Throughout our study, we assume that the displacement map s_t is continuous and sufficiently differentiable. The corresponding eigenvalues are the roots of the characteristic polynomial of the matrix $\mathcal{M} = \partial s_t/\partial q$. Since the characteristic equation is a nonlinear equation, in principle the eigenvalues could develop singularities and become nondifferentiable. However, it can be shown that the eigenvalues can be ordered such that they are continuous. Furthermore the eigenvalues will be assumed to be differentiable whenever

[4]Formally, it would be appropriate to describe the fluids as $(d+1)$-dimensional fluids, a combination of their embedding in a d-dimensional space along with their evolution along time dimension t.

the eigenvalues are distinct. When two eigenvalues coincide, the eigenvalue fields may become non-differentiable.

10.2.1 Hamiltonian fluid dynamics

For fluids moving with no dissipation of energy, the Hamiltonian formalism may be applied. Hamiltonian fluids have a potential velocity field

$$v = \nabla\phi \tag{10.12}$$

with the velocity potential ϕ. The mass density ρ and the velocity potential serve as conjugate variables for the Hamiltonian $\mathcal{H}$, with the equations of motion

$$\begin{aligned} \frac{\partial\rho}{\partial t} &= +\frac{\delta\mathcal{H}}{\delta\phi} = -\nabla\cdot(\rho v)\,, \\ \frac{\partial\phi}{\partial t} &= -\frac{\delta\mathcal{H}}{\delta\rho}\,. \end{aligned} \tag{10.13}$$

A simple example of a Hamiltonian is

$$\mathcal{H} = \int \mathrm{d}x \left(\frac{1}{2}\rho(\nabla\phi)^2 + e(\rho)\right)\,, \tag{10.14}$$

where $e(\rho)$ is the internal energy as a function of density ρ. The first equation of motion in equation (10.13) is equivalent to the continuity equation, while the second equation implies the Euler equation

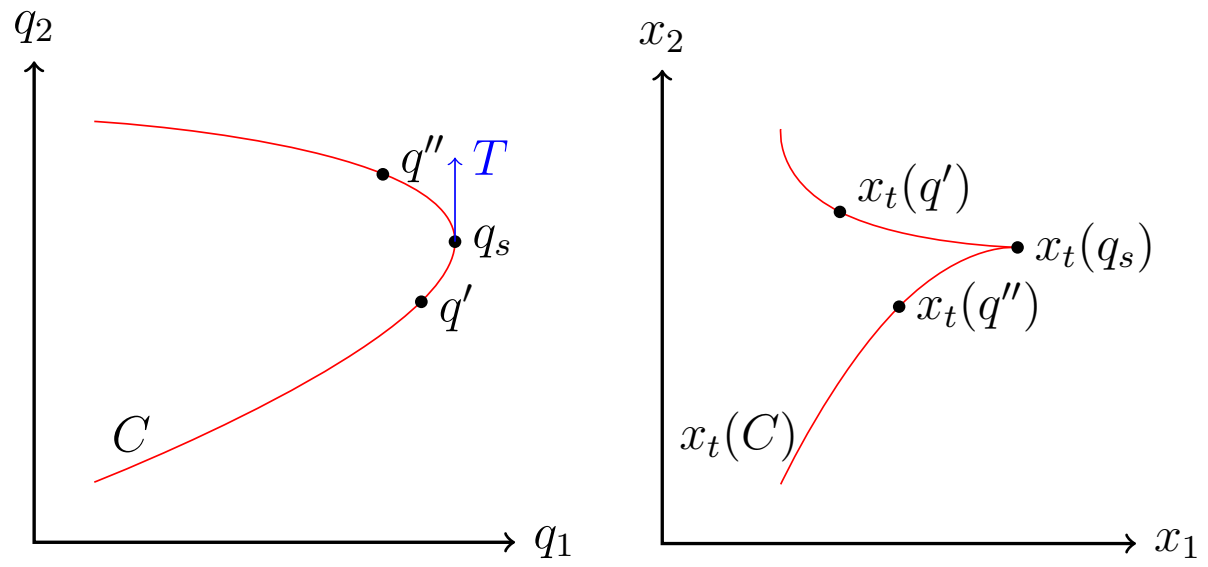
$$\frac{\partial v}{\partial t} + v\cdot\nabla v = -\frac{1}{\rho}\nabla p\,, \tag{10.15}$$

in which p is the pressure of the fluid. For a thorough discussion of fluid mechanics we refer to the seminal volumes of [232], and [233]. For detailed and extensive treatments and analyses of Hamiltonian mechanics and Hamiltonian fluids, we refer to the reviews and textbooks by [16], [19], [167], [255], and [288].

10.3 Shell-crossing conditions

The caustics mentioned above result from the folding of the phase space fluid. At the initial time, $t = 0$, the fluid has not yet evolved. The displacement map s is therefore the zero map (eqn. (10.2)), *i.e.*,

$$s_0(q) = 0 \tag{10.16}$$



(a) Lagrangian space L

(b) Eulerian space E

Figure 10.2: The shell-crossing process of a curve C in a Lagrangian map x_t. The left panel shows Lagrangian space, describing the initial positions of the fluid. The right panel shows Eulerian space, describing the positions of the fluid at time t. The fluid undergoes shell-crossing in point q_s on the curve C (red) at time t. The neighboring points q' and q'' have passed through the opposing segments of C. The Lagrangian mapping of the curve $x_t(C)$ (red) develops a non-differentiable point in $x_t(q_x)$, which is known as a caustic. The arrow T (blue) is the tangent vector of the curve C in point q_s.

for all $q \in L$. The map $x_0(q)$ is one-to-one, *i.e.*, each Eulerian coordinate x corresponds to one Lagrangian position q. Throughout the entire volume, the fluid only contains single-stream regions. As the fluid evolves and nonlinearities start to emerge, we see the development of *multi-stream regions* in the fluid. At the boundary of a multi-stream region, the volume of a mass element vanishes and its density – following eqn. (10.7) – becomes infinite. At such locations in phase space the map $x_t(q)$ attains a n-to-one character, with n an odd positive integer ($n = 3, 5, 7, \ldots$). It means that at any one Eulerian location x, streams from n different Lagrangian positions cross.

The key question we address here is that of inferring the conditions under which a mass element with Lagrangian coordinate q undergoes shell-crossing. Here we derive the neces-

sary and sufficient conditions for the process of shell-crossing to occur. These conditions are called *shell-crossing conditions*. They are the foundation on the basis of which we infer – in section 10.4 – the related conditions on the displacement field for the occurrence of the various classes of caustics. These are called the *caustic conditions*. We infer the caustic conditions for generic as well as Hamiltonian fluid dynamics.

10.3.1 Shell-crossing condition: the derivation

A typical configuration resulting from the *shell-crossing process* – the name by which it is usually indicated – is illustrated in figure 10.2. It focuses on points $q = (q_1, q_2)$ that lie on a smooth curve C in Lagrangian space L (fig. 10.2a). In this context, smooth refers to the assumption that the curve C is C^1 continuous. At time t, the points on the Lagrangian curve C map to the variety $x_t(C)$ in Eulerian space E (fig. 10.2b)[5]. The fluid in point q_s undergoes shell-crossing at time t. The neighboring points q' and q'' have passed through the opposing segments of C. As a result of this, the curve C develops a non-differentiable point in $x_t(q_x)$, which is known as a *caustic*.

In a time sequence of three steps, figure 10.3 illustrates the dynamical process that is underlying the formation of the caustic at $x_t(q_s)$. The singularity at $x_t(q_s) \in x_t(C)$ forms as the result of a folding process in phase space. We may appreciate the emerging structure when assessing the fate of two neighboring points $q', q'' \in C$ on both sides of q_s. While the phase space sheet $x_t(C)$ is folded, the points $x_t(q')$ and $x_t(q'')$ turn around while passing through $x_t(q_s)$. In figure 10.3 we observe how the initially single-stream phase space sheet (lefthand panel) morphs into a configuration marked by shell-crossing as different mass elements q pile up at the same Eulerian position $x_t(q_s)$ (central panel). Subsequently, around $x_t(q_s)$ we notice the formation of a multi-stream region, with the presence of mass elements q' having passed into a region where mass elements from other Lagrangian locations q are to be found.

To infer the shell-crossing conditions, we investigate a curve C in Lagrangian space along which we have points q that will find themselves incorporated in a singularity at Eulerian position $x_s(q_s)$. In the case of shell-crossing, points q near the Lagrangian location q_s will map onto the same Eulerian position $x(q_s)$. The key realization is that this occurs as points q along a direction T tangential to C are all folded on to a single Eulerian position $x_s(q_s)$. This translates the question of the shell-crossing condition into one on the identity

[5]In algebraic geometry, a *variety* is the zero set of a function f, ie. the set of solutions $x \in E$ such that $f(x) = 0$.

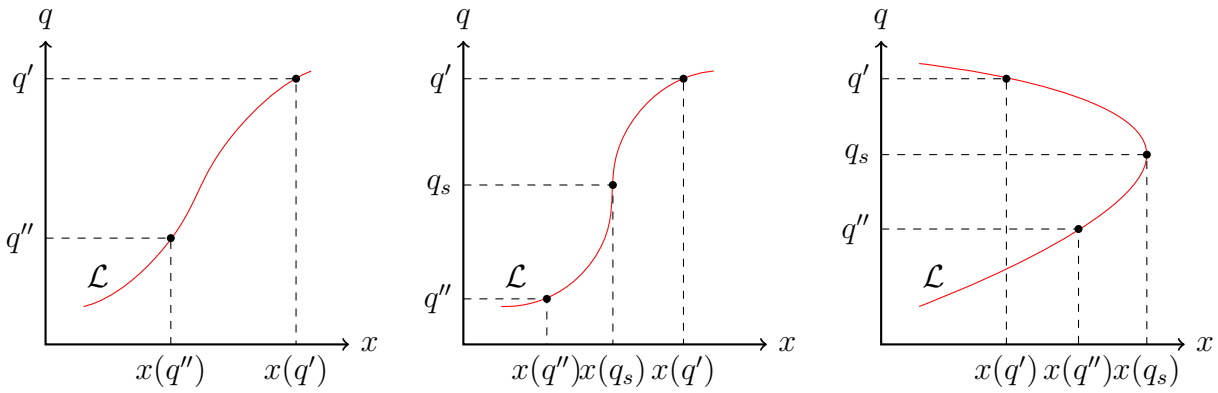


Figure 10.3: Folding of a one-dimensional fluid in phase space $\mathcal{C}$. The three panels show the time evolution of the Lagrangian submanifold $\mathcal{L}$ (red) of the fluid in phase space. We track the evolution of two points $(q', x(q')), (q'', x(q''))$ forming a multi-stream region and mark the point undergoing shell-crossing by $(q_s, x(q_s))$. Left panel: the fluid – early in its evolution – consisting of a single-stream region. Middle panel: a fluid during the process of shell-crossing. Right panel: a fluid consisting of a multi-stream region.

of a tangential direction $T(q)$ along which shell-crossing may or will occur. In other words, whether on a particular curve C – or, more general, a manifold M – there are points q where along one or more tangential directions $T(q)$ to that curve or manifold shell-crossing may or will take place.

Zooming in on two points q' and q'' in the vicinity of the singularity point q_s, we see that as a result of the folding process the ratio of the distances of the two points in the Lagrangian and Eulerian manifold, must go to zero in the limit that we zoom in on points q' and q'' along the Lagrangian curve C at an infinitesimal distance from q_s, *i.e.*,

$$\frac{\Delta x}{|\Delta q|} = \frac{\|x_t(q') - x_t(q'')\|}{\|q' - q''\|} \to 0 \qquad q', q'' \to q_s \,. \tag{10.17}$$

The direct implication of this is, following equation (10.7), that the density in a caustic is infinite: the volume of the mass element associated to q_s vanishes at time t. In essence it informs us that during shell crossing the points q near Lagrangian location q_s, along the tangential direction T to the Lagrangian curve C, map onto the same Eulerian position $x(q_s)$. This means that the norm of the directional derivative of x_t along the tangential direction vanishes. In other words, along the non-zero tangent vector T along C,

$$\left\| \frac{\partial x_t}{\partial q} T \right\| = 0 \,, \tag{10.18}$$

where $\partial x_t/\partial q$ is the Jacobian of x_t evaluated in q_s (see Figure 10.2a). This is equivalent to requiring that

$$\frac{\partial x_t}{\partial q} T \,=\, 0\,. \tag{10.19}$$

In terms of the displacement map s_t, this condition can be expressed as

$$T + \frac{\partial s_t}{\partial q} T = 0\,, \tag{10.20}$$

with the Jacobian $\frac{\partial s_t}{\partial q}$ also evaluated in q_s[6]. Subsequently consider the eigenvalues μ_i and eigenvectors v_i of the deformation tensor $\mathcal{M} = \frac{\partial s_t}{\partial q}$, defined by

$$\mathcal{M} v_i = \mu_i v_i\,. \tag{10.21}$$

Under the assumption that the deformation tensor is diagonalizable[7], we can construct the diagonal matrix $\mathcal{M}_d = \mathrm{diag}(\mu_1, \ldots, \mu_d)$ and the eigenvector matrix $\mathcal{V} = (v_1, \ldots, v_d)$. For an analysis of the case of non-diagonalizable deformation tensors see appendix 10.B. In three dimensions, with the eigenvalues μ_i and eigenvectors $v_i = (v_{i,1}, v_{i,2}, v_{i,3})$, the diagonal matrix $\mathcal{M}_d$ and eigenvector matrix $\mathcal{V}$ are given by

$$\mathcal{M}_d = \begin{pmatrix} \mu_1 & 0 & 0 \\ 0 & \mu_2 & 0 \\ 0 & 0 & \mu_3 \end{pmatrix}, \qquad \mathcal{V} = \begin{pmatrix} v_{1,1} & v_{2,1} & v_{3,1} \\ v_{1,2} & v_{2,2} & v_{3,2} \\ v_{1,3} & v_{2,3} & v_{3,3} \end{pmatrix}. \tag{10.22}$$

In terms of $\mathcal{V}$ and $\mathcal{M}_d$, condition (10.20) reduces to

$$0 = (I + \mathcal{M})\mathcal{V}\mathcal{V}^{-1}T = \mathcal{V}(I + \mathcal{M}_d)\mathcal{V}^{-1}T \tag{10.23}$$

since $\mathcal{V}$ is always invertible[8], using the identity

$$\mathcal{M}\mathcal{V} = \mathcal{M}(v_1, \ldots, v_d) = (\mathcal{M}v_1, \ldots, \mathcal{M}v_d) = (\mu_1 v_1, \ldots \mu_d v_d) = \mathcal{V}\mathcal{M}_d\,. \tag{10.24}$$

We thus obtain the condition

$$(I + \mathcal{M}_d)\mathcal{V}^{-1}T = 0\,, \tag{10.25}$$

[6]Unless mentioned otherwise, we will assume all Jacobians to be evaluated in q_s.

[7]In practice, the assumption of diagonalizability is not really restrictive: non-diagonalizable matrices are unstable, which means that they can be turned diagonalizable by means of a small perturbation in the initial conditions.

[8]That is to say, the eigenvectors can always be chosen to be linearly independent.

which holds for general diagonalizable deformation tensors. Note that the rows of $\mathcal{V}^{-1}$ consist of the dual vectors $\{v_i^*\}$ of the eigenvectors $\{v_i\}$, defined by $v_i \cdot v_j^* = \delta_{ij}$ for all i and j. Explicitly, this means that $\mathcal{V}^{-1}$ in three dimensions is given by

$$\mathcal{V}^{-1} = \begin{pmatrix} v_{1,1}^* & v_{1,2}^* & v_{1,3}^* \\ v_{2,1}^* & v_{2,2}^* & v_{2,3}^* \\ v_{3,1}^* & v_{3,2}^* & v_{3,3}^* \end{pmatrix} , \tag{10.26}$$

with $v_i^* = (v_{i,1}^*, v_{i,2}^*, v_{i,3}^*)$. The product $\mathcal{V}^{-1}T$ is the vector composed out of the inner product of these dual vectors with the tangent vector T, so that in three dimensions equation (10.25) reduces to

$$\begin{pmatrix} (1+\mu_1)v_1^* \cdot T \\ (1+\mu_2)v_2^* \cdot T \\ (1+\mu_3)v_3^* \cdot T \end{pmatrix} = 0\,. \tag{10.27}$$

This represents the proof for the shell-crossing condition for one-dimensional submanifolds. It states the condition for the tangential direction T along which Lagrangian points get folded into an Eulerian singularity point. The obtained condition is a telling expression for the central role of both the deformation eigenvalues and eigenvectors in determining the occurrence of a singularity.

10.3.2 Shell-crossing condition: theorems

Following the proof outlined in the previous subsection 10.3.1, we arrive at the following two theorems stipulating the conditions for the formation of singularities by curves C and arbitrary manifolds M in Lagrangian space L,

Theorem: 1 *A C^1 continuous curve $C \subset L$ forms a singularity under the mapping x_t in the point $x_t(q_s) \in x_t(C) \subset E$, meaning that $x_t(C)$ is not smooth in $x_t(q_s)$, if and only if*

$$(1+\mu_{it}(q_s))v_{it}^*(q_s) \cdot T = 0 \tag{10.28}$$

for all $i = 1, 2, \ldots, dim(L)$, with T a nonzero tangent vector of C in q_s.

Note that the derived caustic condition is independent of the dynamics of the fluid. In general, both the eigenvalue and eigenvector fields are complex-valued. For Hamiltonian fluids, the relation condition simplifies since the eigenvalue and eigenvector fields are forced to be real-valued and the eigenvectors can be chosen to coincide with their dual vectors, *i.e.*, $v_i^* = v_i$.

A similar argument holds for higher dimensional submanifolds of L, *e.g.*, sheets and volumes. These manifolds can be n-dimensional, with $n = 1, \ldots, 3$ for three-dimensional fluids. Given an arbitrary manifold $M \subset L$ we can consider all curves $C \subset M$ passing through the point $q_s \in M$. The variety $x_t(M)$ contains a singularity at $x_t(q_s)$ if and only if at least one such curve $C \subset M$ gets folded under the map x_t. Hence for an arbitrary submanifold M, we should consider the one-dimensional shell-crossing condition for the subset of vectors $T \in M$ [9]. This proves the general shell-crossing condition:

Theorem: 2 *A manifold $M \subset L$ forms a singularity under the mapping x_t in the point $x_t(q_s) \in x_t(M) \subset E$ at time t, meaning that $x_t(M)$ is not smooth in $x_t(q_s)$, if and only if there exists at least one nonzero tangent vector $T \in T_{q_s}M$ satisfying*

$$(1 + \mu_{it}(q_s))v_{it}^*(q_s) \cdot T = 0 \tag{10.29}$$

for all $i = 1, 2, \ldots, dim(L)$.

From this theorem, we immediately observe that the eigenvectors v_i are of key importance in determining the nature of the singularity, in that the shell-crossing condition is not simply that of $1 + \mu_i = 0$ for at least one i. More explicitly, the shell-crossing condition says that

$$1 + \mu_{it}(q_s) \;=\; 0 \qquad OR \qquad v_{it}^*(q_s) \cdot T \;=\; 0 \qquad \text{for all } i\,, \tag{10.30}$$

[9] $T_{q_s}M$, , *i.e.*, all tangent vectors T constrained to be located in the vector space $T_{q_s}M$ [10]. In other words, the one-dimensional shell-crossing condition is considered for all vectors T in the vector space of all tangential vectors to the manifold M in $q_s \in M$.

indicating that, in addition to one or more eigenvalue constraints $1 + \mu_i = 0$, the shell-crossing condition consists of complementary constraints. These single out those points q_s where the eigenvectors $v_j^*(q_s)$ (with $j \neq i$) are orthogonal to a vector T that is restricted to be located in the plane tangent to the manifold M in which the singularity emerges. It is this constraint that is instrumental in defining the area occupied by the corresponding caustic.

Note that the shell-crossing conditions are manifestly independent of coordinate choices. While in general the eigenvalue and eigenvector fields generally do depend on the choice of coordinates, it can be shown that they are invariant if the corresponding coordinate transformation is orthogonal and global. These transformations include rotations and translations. See appendix 10.C for more details.

10.4 Caustic conditions

In section 10.3, we inferred the general condition for shell-crossing. The condition establishes the relation between the eigenvalue and eigenvector fields of the deformation tensor in Lagrangian space, and the Lagrangian regions that get incorporated in features of infinite density in Eulerian space. Moreover, it allows us to establish the identity of the resulting singularity in Eulerian space.

The stable singularities that emerge can be classified by Lagrangian catastrophe theory in the A_k, D_k and E_k series (see [18], [165] and [280]). This is described in some detail in section 10.5 [11]. The A_k series is of co-rank 1, in which co-rank is the number of independent directions in which the Hessian is degenerate. The A_k series corresponds to the caustics for which the density diverges due to only one eigenvalue. The D_k series is of co-rank 2 and corresponds to the points for which the density diverges due to two eigenvalue fields. The E_k series is of co-rank 3 and corresponds to the points for which three eigenvalue fields. However, for three-dimensional fluids, the points for which all eigenvalues simultaneously satisfy this condition are degenerate. For this reason we will not discuss them here.

In this section we apply the shell-crossing condition to three-dimensional Lagrangian fluids to obtain the *caustics conditions* which relate the classification of caustics to the eigenvalue and eigenvector field. These conditions have not been derived in earlier work and are necessary to perform a quantitative study of caustics in large scale structure formation. In section 10.5, we summarize the classification of caustics in its traditional form and compare them to the caustic conditions derived here.

[11] The classification ultimately has its origin in the classification of Coxeter groups

10.4.1 The A family

The A family of caustics form when

$$1 + \mu_i \quad = \quad 0 \tag{10.31}$$

for one i. For diagonalizable deformation tensors, the eigenvector fields $\{v_i\}$ and their dual vector fields $\{v_i^*\}$ are linearly independent.

For three-dimensional fluids, the A family consists of 5 classes running from the trivial A_1 class, corresponding to the points that never form caustics, the sheetlike A_2 fold, the curvelike A_3 cusp, the A_4 swallowtail, to the pointlike A_5 butterfly singularity.

The trivial A_1 class

The A_1 class labels the points which never form caustics. According to the shell-crossing condition, q_s will form a singularity at time t if and only if there exists a nonzero tangent vector $T \in T_{q_s}L$ for which

$$(1 + \mu_i(q_s))v_i^*(q_s) \cdot T = 0 \tag{10.32}$$

for all i. The point q_s will not satisfy this condition if $1 + \mu_i(q_s) \neq 0$ for all i since the three dual vectors $\{v_i^*\}$ of the (generalized) eigenvectors span the tangent space $T_{q_s}L$

From the shell-crossing condition we therefore conclude that the three-dimensional variety A_1,

$$A_1 = \{q \in L | 1 + \mu_{ti}(q) \neq 0 \text{ for all } i \text{ and } t\}\,, \tag{10.33}$$

consists of the points never forming caustics. In this respect we should note that the displacement map at the initial time is the zero map, so that the eigenvalues at the initial time are equal to zero, *i.e.*, $\mu_{0i}(q) = 0$ for all $q \in L$. Since the eigenvalues are a continuous function of time, for the cosmologically interesting case of potential flow the requirement for a point q to belong to A_1 is equivalent to $\mu_{ti}(q) > -1$.

The A_2 caustics

Based on the discussion above, we may conclude that for a given i, $i = 1 \ldots 3$, at time t the points

$$A_2^i(t) = \{q \in L | 1 + \mu_{ti}(q) = 0\} \tag{10.34}$$

form a singularity. For three-dimensional fluids, the set $A_2(t)$ forms a two-dimensional sheet, sweeping through space as the fluid evolves. These singularities can be associated to the A_2 fold singularity class.

From this, we conclude that the set of points which form a A_2 fold singularity at a time $t \in [0, \infty)$ is given by

$$A_2^i = \{q \in L | 1 + \mu_{ti}(q) = 0 \text{ for some } t\} \,. \tag{10.35}$$

The A_3 caustics

Following up on the folding of the fluid to the A_2^i singularity, the A_2^i manifold itself may be folded into a more complex configuration. The result is a so-called A_3 singularity. To guide understanding in the emergence of cusps we may refer to the eigenvalue contour map of figure 10.4.

To infer the identity of the A_3^i caustic, we restrict the criterion for shell-crossing to points on the A_2^i manifold. In other words, we look for points q_s on the surface of the sheetlike variety $A_2^i(t)$ that fulfill the criterion for shell-crossing.

A point $q_s \in A_2^i(t)$ forms a singularity if there exists a nonzero tangent vector T, $T \in T_{q_s} A_2^i(t)$, orthogonal to the $\mathrm{Span}_{\mathbb{C}}\{v_j^* | j \neq i\}$. Writing the tangent vector T as a linear combination of the eigenvectors v_i,

$$T \;=\; \alpha_1 v_1 + \alpha_2 v_2 + \alpha_3 v_3 \,, \tag{10.36}$$

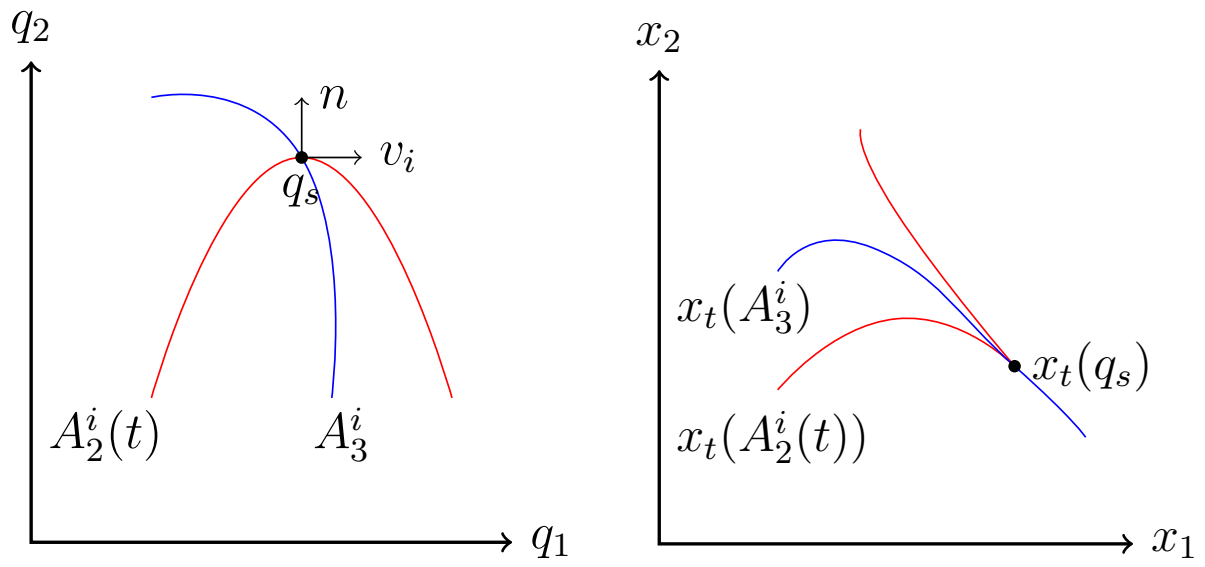
with $\alpha_i \in \mathbb{C}$. The caustic conditions tell us that

$$\alpha_j \;=\; v_j^*(q_s) \cdot T = 0 \quad \text{for } j \neq i \,. \tag{10.37}$$

Given that we know that the ith eigenvalue is real, $\mu_i \in \mathbb{R}$, the eigenvector v_i is also real. This means that this condition is satisfied if and only if the tangent vector T is parallel to v_i. This is equivalent to the condition that v_i is orthogonal to the normal $n = \nabla \mu_{ti}$ of the manifold $A_2^i(t)$ in the point q_s. Explicitly, this means that the inner product of n with v_i is equal to 0,

$$\mu_{ti,i} \equiv v_i \cdot \nabla \mu_{ti} = 0 \,. \tag{10.38}$$

Note that this is the condition that Arnol'd [17] found for the A_3 line for the 2-dimensional Zel'dovich approximation. As we see from the derivation above, the condition is valid in any dimensional space and for general flow configurations.



(a) Lagrangian space

(b) Eulerian space

Figure 10.4: The formation of a cusp (A_3) singularity in a Lagrangian map x_t. The left panel shows Lagrangian space, describing the initial positions of the fluid. The right panel shows Eulerian space, describing the positions of the fluid at time t. The fluid undergoes shell-crossing along the fold $A_2^i(t)$ (red) at time t. The fold gets mapped under the Lagrangian map to $x_t(A_2)$ (red), which is folded into a cusp in the point $x_t(q_s)$ corresponding to q_s. The cusp forms if and only if the normal n of $A_2^i(t)$ is orthogonal to the eigenvector field v_i in q_s. Over time, the cusp traces out the curve A_i (blue) which is mapped to $x_t(A_3^i)$ (blue).

The points q forming a cusp at time t corresponding to eigenvalue field μ_i is given by the one-dimensional variety

$$A_3^i(t) = \{q \in L | q \in A_2^i(t) \ \wedge \ \mu_{ti,i}(q) = 0\}\,. \tag{10.39}$$

Extrapolating this to the set of all points q that at some time $t \in [0, \infty)$ have belonged to or will be incorporated in a cusp singularity defines a two-dimensional variety

$$A_3^i = \{q \in L | q \in A_2^i(t) \ \wedge \ \mu_{ti,i}(q) = 0 \text{ for some } t\}\,, \tag{10.40}$$

which is the assembly of all $A_3^i(t)$ over the time interval $t \in [0, \infty)$.

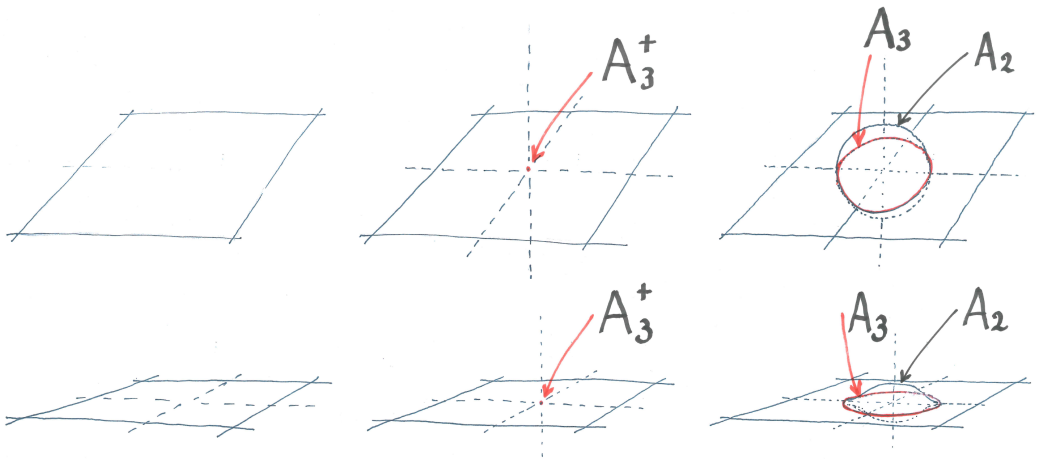


Figure 10.5: The creation/annihilation of a fold (A_2) sheet in a A_3^+ point. The upper three panels show the unfolding of a A_3^+ singularity in Lagrangian space. The lower three panels show the corresponding unfolding in Eulerian space. The two panels on the left show the cusp (A_3) plane on which the cusps form. The middle panels show the appearance of a A_3^+ singularity in which a fold sheet is formed/removed. The right panels show the resulting fold (A_2) sheet. The fold sheet gets folded into a cusp (A_3) curve (red). This configuration is known as the *Zel'dovich pancake* (Zel'dovich 1970).

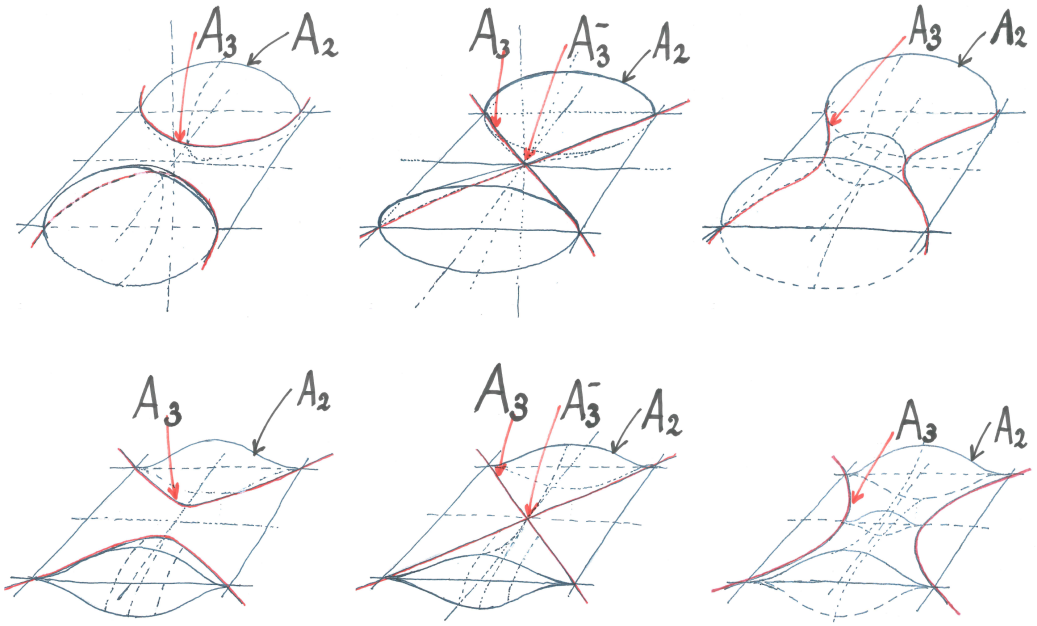


Figure 10.6: The merger/splitting of a fold (A_2) sheet in a A_3^- point. The upper three panels show the unfolding of a A_3^- singularity in Lagrangian space. The lower three panels show the corresponding unfolding in Eulerian space. The two panels on the left show two fold (A_2) sheets, two cusp (A_3) curves (red) and the cusp (A_3) plane on which the cusps form. The middle panels show the merger/splitting of the two fold (A_2) sheets in a A_3^- singularity. The right panels show the resulting merged fold (A_2) sheet. This configuration is known as the *Kissing Lips*.

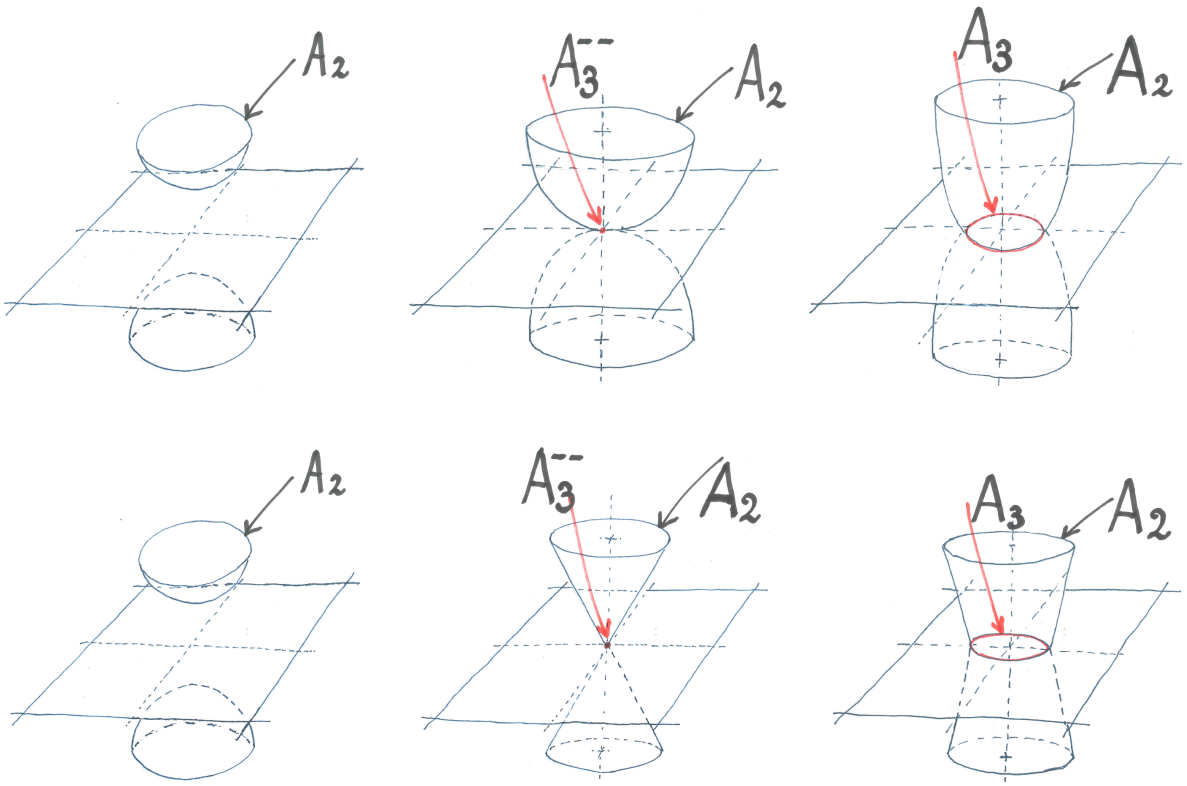
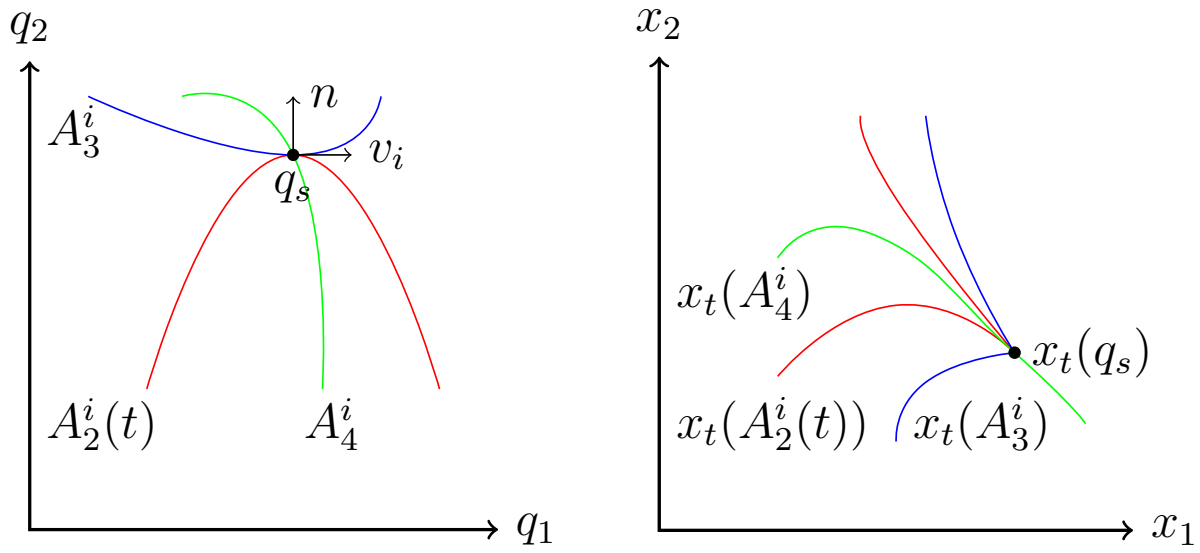


Figure 10.7: The merger/splitting of a fold (A_2) sheet in a A_3^{--} point. The upper three panels show the unfolding of a A_3^{--} singularity in Lagrangian space. The lower three panels show the corresponding unfolding in Eulerian space. The two panels on the left show two fold (A_2) sheets, and the cusp (A_3) plane on which the cusps form. The middle panels show the merger/splitting of the two fold (A_2) sheets in a A_3^{--} singularity. The right panels show the resulting merged fold (A_2) sheet with the corresponding cusp (A_3) curve.



(a) Lagrangian space

(b) Eulerian space

Figure 10.8: The formation of a swallowtail (A_4) singularity in a Lagrangian map x_t. The left panel shows the Lagrangian space describing the initial positions of the fluid. The right panel shows the Eulerian space describing the positions of the fluid at time t. The fluid undergoes shell-crossing along $A_2^i(t)$ (red) at time t. The fold gets mapped in Eulerian space, under the Lagrangian map, to $x_t(A_2)$ (red), which is folded into a cusp in the point $x_t(q_s)$ corresponding to q_s. The cusp forms if and only if the normal n of $A_2^i(t)$ is orthogonal to the eigenvector field v_i in q_s. Over time, in Lagrangian space the cusp traces out the curve A_i (blue) which in Eulerian space is mapped to $x_t(A_3^i)$ (blue). Since the cusp (A_3^i) curve is tangential to the fold (A_2) curve in q_s, the cusp curve $x_t(A_3^i)$ forms a swallowtail (A_4) singularity. Over time, the swallowtail traces out A_4^i (green), which in Eulerian space is mapped into $x_t(A_4^i)$ (green).

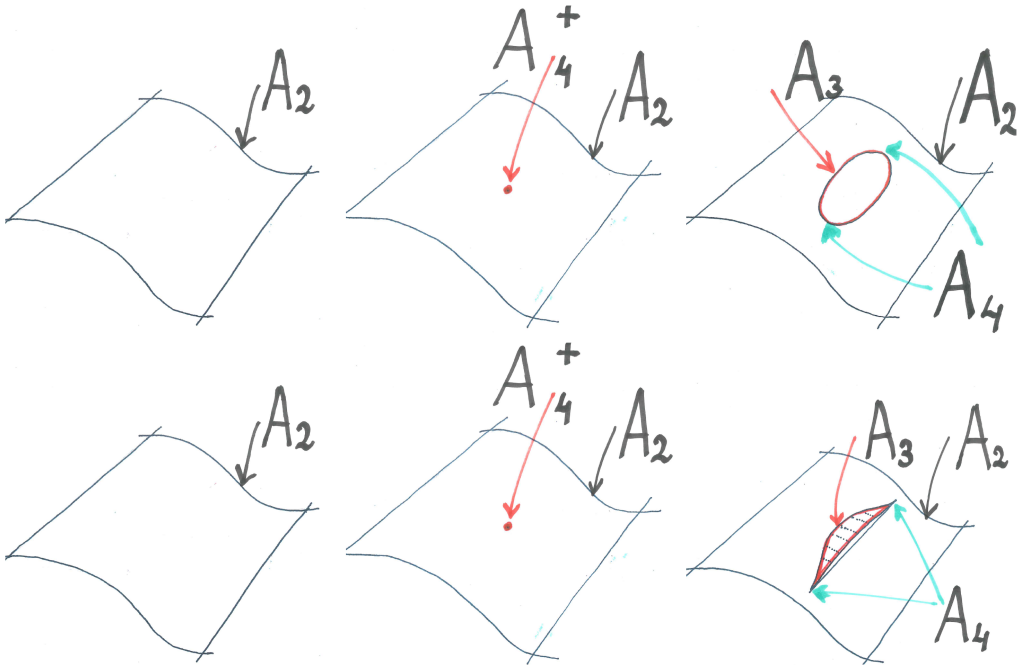


Figure 10.9: The creation/annihilation of a swallowtail (A_4) singularity in a A_4^+ point. The upper three panels show the unfolding of a A_4^+ singularity in Lagrangian space. The lower three panels show the corresponding unfolding in Eulerian space. The two panels on the left show a fold (A_2) sheet. The middle panels show a A_4^+ point on the fold (A_2) sheet. The A_4^+ point leads to the creation/annihilation of two swallowtail (A_4) singularities. The right panels show the resulting cusp (A_3) curves and swallowtail (A_4) singularities.

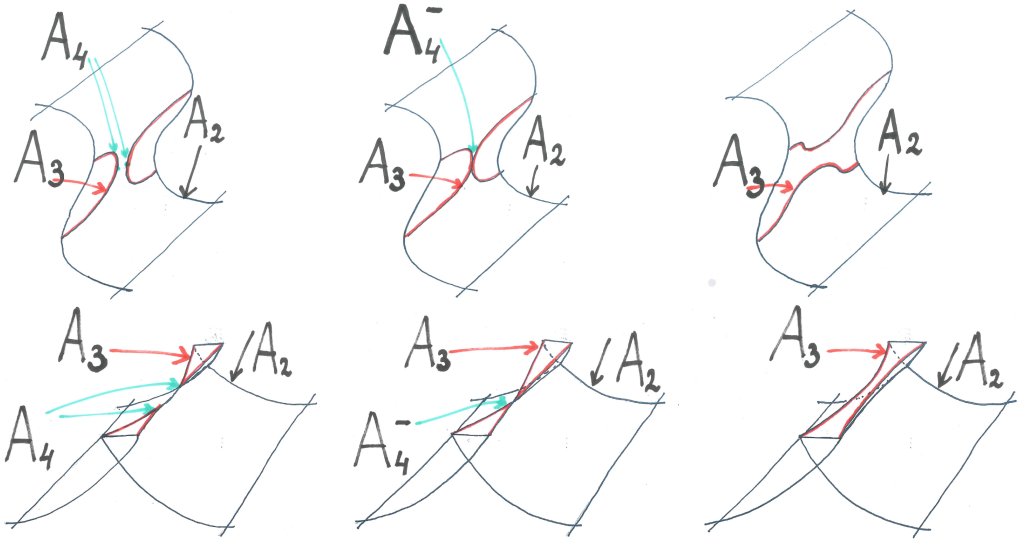


Figure 10.10: The merger/splitting of a cusp (A_3) curve in a A_4^- point. The upper three panels show the unfolding of a A_4^- singularity in Lagrangian space. The lower three panels show the corresponding unfolding in Eulerian space. The two panels on the left show a fold (A_2) sheet, cusp (A_3) curves and swallowtail (A_4) singularities. The middle panels show the merger/splitting of the cusp (A_3) curves in a A_4^- point. The right panels show the resulting fold (A_2) sheet and cusp (A_3) curves singularities.

The $A_3^{\pm}$ points

The topology of the sheetlike $A_2^i(t)$ variety changes as a function of time. These topological changes occur at critical points of the corresponding eigenvalue field μ_{ti}. It is at these points where in Eulerian space we see the emergence of new features, the disappearance of features and/or the merging of features. The critical points are classified as cusp singularities.

At minima of the μ_i field, a feature gets created. At maxima, a feature gets annihilated. Particularly interesting points are the saddle points. In three-dimensional space, there are two classes of saddles in the eigenvalue field μ_{ti}. The index 1 saddles have a Hessian signature $(--+)$, with 1 positive eigenvalue, while the index 2 saddles have a signature $(-++)$.

Based on their impact on caustic structure, Arnol'd used a slightly different classification scheme, in which the distinguished between A_3^{++}, A_3^{+-} and A_3^{--} points [17]. The A_3^{++} point are identified with the minima[12], while the A_3^{+-} points are the saddle points for which the A_3 sheet intersects the two disjoint A_2 sheets. This is illustrated in the upper left panel in Figure 10.6. The additional A_3^{--} points correspond to saddle points for which the A_3 sheet does not intersect the disjoint A_2 sheets. Because this concerns a non-generic situation, we do not treat it here. Also note that higher dimensional fluids will have additional A_3 points.

In the context of this chapter we therefore use a slightly shorter notation for the maxima, minima and saddles, classifying them as the cusp singularities A_3^+ an A_3^-,

$$\begin{aligned} A_3^{i+} &= \{q \in L | q \in A_2^i(t) \wedge \mu_{ti}(q) \text{ max-/minimum of } \mu_{ti} \text{ at some time } t\}\,, \\ A_3^{i-} &= \{q \in L | q \in A_2^i(t) \wedge q \text{ saddle point of } \mu_{ti} \text{ at some time } t\}. \end{aligned} \tag{10.41}$$

Note that in this scheme, the saddle points with index 1 and 2 belong to the same singularity class A_3^{i-}. For an illustration of the A_3^+, A_3^- and A_3^{--} singularities, we refer to figures 10.5, 10.6 and 10.7. From the caustics conditions we may directly infer that the $A_3^{i\pm}$ points are located on the A_3^i variety.

The A_4 caustics

In Eulerian space the $A_3^i(t)$ variety gets folded in points associated with A_4 swallowtail singularities. The identity of the points defining the variety $A_4^i(t)$ can be inferred by the

[12]Note that in Arnol'd's notation, related to the Zel'dovich formalism (see appendix A), these are the maxima of the eigenvalue field

application of the general shell-crossing condition (eqn. (10.29)) to the $A_3^i(t)$ variety (see Figure 10.8). As a consequence, the A_4^i variety is defined as

$$A_4^i(t) = \{q \in L | q \in A_3^i(t) \wedge \mu_{ti,ii}(q) = 0\}, \tag{10.42}$$

with $\mu_{ti,ii}(q)$ the inner product of the normal $n = \nabla\mu_{ti,i}$ with the eigenvector v_i,

$$\mu_{ti,ii} \equiv v_i \cdot \nabla\mu_{ti,i}\,. \tag{10.43}$$

Integrated over time, the points on the varieties $A_4^i(t)$ trace out the 1-dimensional variety A_4^i, , *i.e.*, the 1-dimensional line A_4^i is the set of all points $A_4^i(t)$ over the time interval $t \in [0, \infty)$,

$$A_4^i = \{q \in L | q \in A_3^i(t) \wedge \mu_{ti,ii}(q) = 0 \text{ for some } t\}\,. \tag{10.44}$$

The $A_4^{\pm}$ points

The topology of the variety $A_3^i(t)$ changes as a function of time. To this end, we identify the critical points of the real field $\mu_{i,i}$,

$$\mu_{ti,i} \equiv v_i \cdot \nabla\mu_{ti}\,. \tag{10.45}$$

Constraining the location of these singularities to the one-dimensional curvelike variety $A_3^i(t)$, and thus implicitly also to the two-dimensional membrane of the variety $A_2^i(t)$, these $A_4^{\pm}$ points mark the locations at which topological changes occur. They represent the sites at which we see the birth of new singularities in Eulerian space, or the annihilation of and/or merging of such features. These singularities are classified as swallowtail singularities.

The birth or death of features on $A_3^i(t)$ takes place at maxima and minima of $\mu_{ti,i}$, and is identified with A_4^{i+} singularities. The merging or splitting of features happens at the saddle points of the same field $\mu_{ti,i}$. The latter mark the A_4^{i-} singularities,

$$\begin{aligned} A_4^{i+} &= \{q \in L | q \in A_3^i(t), \mu_{ti,i}(q) \text{ max-/minimum of } \mu_{ti,i}|_{A_2^i(t)} \text{ for some } t\}, \\ A_4^{i-} &= \{q \in L | q \in A_3^i(t) \text{ saddle point of } \mu_{ti,i}|_{A_2^i(t)} \text{ for some } t\}. \end{aligned} \tag{10.46}$$

The $A_4^{\pm}$ critical points are constrained to lie on the curvelike variety $A_2^i(t)$. Their identity is therefore determined by the interplay between the geometric properties of two entities. One of these is the geometry of the field $\mu_{ti,i}$, the other that of the geometry of the curvelike variety $A_3^i(t)$. For illustrations of the A_4^+ and A_4^- singularities we refer to figure 10.9 and 10.10.

From the caustic conditions – as expressed in eqn. (10.42) – we may also immediately observe that the $A_4^{i\pm}$ points belong to the A_4^i variety. In fact, this also represents a condition on the topology of the field $\mu_{ti,i}$ and that of the $A_2^i(t)$ variety.

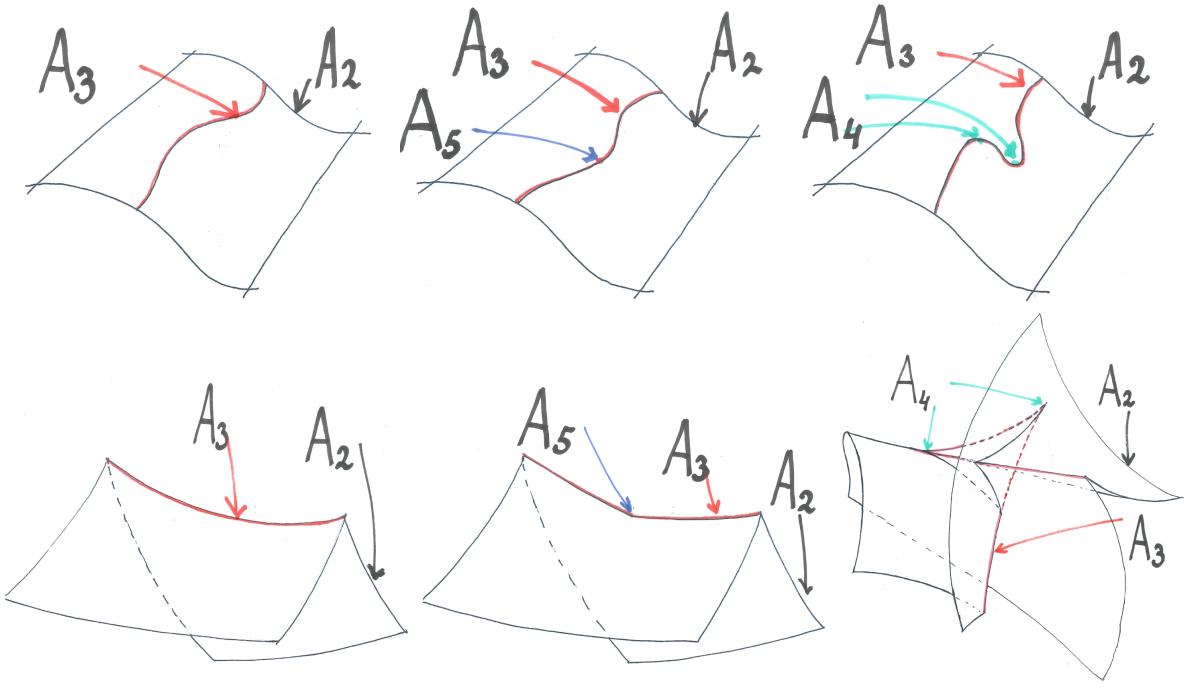


Figure 10.11: The creation/annihilation of swallowtail singularities in a butterfly (A_5) singularity. The upper three panels show the unfolding of a A_5 singularity in Lagrangian space. The lower three panels show the corresponding unfolding in Eulerian space. The two panels on the left show a fold (A_2) sheet, and cusp (A_3) curve. The middle panels show the creation/annihilation of the butterfly (A_5) singularity on the cusp (A_3) curve. The right panels show the resulting fold (A_2) sheet, cusp (A_3) curve and swallowtail (A_4) singularities.

The A_5 caustics

Finally, also the swallowtail curves A_4^i curve get folded in Eulerian space. It leads to the emergence of so-called butterfly singularities, or A_5 singularities. Following the same reasoning as for the A_3^i and A_4^i varieties, we may infer from the general shell-crossing condition that the A_4^i curve gets folded in the points A_5^i. In general this happens when there exists a tangent vector of A_4 parallel to v_i, *i.e.*,

$$A_5^i = \{q \in L | q \in A_4^i \text{ and } v_i \in T_q A_4^i\} \tag{10.47}$$

In the case three-dimensional case, when the displacement field $s_t(q)$ is separable into temporal and spatial parts, time evolution can be seen as a progression through a series of surfaces. The folding points can then be found from the relation,

$$A_5^i = \{q \in L | q \in A_4^i(t) \text{ and } \mu_{ti,iii} = 0 \text{ for some time } t\}. \tag{10.48}$$

Figure 10.11 shows an illustration of a A_5 singularity.

The butterfly singularity is the highest dimensional singularity that may surface in three-dimensional Lagrangian fluids. It is important to realize that the butterfly singularity only exists at one point in spacetime.

10.4.2 The D family

The D family of caustics correspond to manifolds for which the condition

$$1 + \mu_i = 0\,, \tag{10.49}$$

holds for two eigenvalue fields simultaneously. From this, we may immediately infer that these caustics form at the intersection of two $A_2(t)$ fold sheets, the $A_2^i(t)$ and $A_2^j(t)$ varieties. In all, for three-dimensional fluids three classes of D caustics can be identified, the D_4^- elliptic, the D_4^+ hyperbolic and the D_5 parabolic umbilic caustic.

The D_4 caustics

The D_4 caustics are defined by the points q in Lagrangian space, at which two of the eigenvalues have the same value. For instance, the $D_4^{ij}(t)$ caustic, with $i \neq j$, is outlined by the points q for which at the time t the eigenvalues μ_i(t) and $\mu_j(t)$ are equal, $\mu_{ti} = \mu_{tj}$.

While the eigenvalue μ_{ti} defines the fold sheet A_2^i, and the eigenvalue μ_{ti} the fold sheet A_2^j, the umbilic D_4^{ij} caustic consist of the set of points q for which

$$D_4^{ij}(t) = \{q \in L | q \in A_2^i(t) \cap A_2^j(t)\} \,. \tag{10.50}$$

In three-dimensional space, one would expect that the intersection of the two sheets $A_2^i(t)$ and $A_2^j(t)$ to consist of one-dimensional curves. This would certainly be true for two sheets that would be entirely independent of each other. However, the situation at hand concerns a highly constrained situation, in which the two eigenvalues μ_i and μ_j are strongly correlated.

Because of the latter, the intersection between the folds A_2^i and A_2^j is considerably more complex. Instead of a continuous curve, the intersection consists of isolated, singular points. A telling illustration – and discussion – of this, for the two-dimensional situation, can be found in [203].

D_4 singularities and A_3 varieties

To investigate the geometry and structure of the set $D_4^{ij}(t)$ we focus on the particular situation of the set $D_4^{12}(t)$, in which the two first eigenvalues μ_1 and μ_2 have the same value, $\mu_{t1} = \mu_{t2} = -1$. Without loss of generality, we transform the coordinate system such that the third eigenvector v_3 defines the q_3 axis. This transformation makes the q_1q_2-plane the one in which we see the folding and collapse of the phase space sheets to the A_2^1 and A_2^2 caustics.

Assuming that the deformation tensor $\mathcal{M}$ is diagonalizable, in this coordinate system it has the form,

$$\mathcal{M} \;=\; \begin{pmatrix} M_{11} & M_{12} & 0 \\ M_{12} & M_{22} & 0 \\ 0 & 0 & \mu_3 \end{pmatrix} \,, \tag{10.51}$$

in which μ_3 is the third eigenvalue of $\mathcal{M}$. Because the eigenvalues are equal, we get the following 2 conditions for the D_4^{12} caustic.

$$\begin{aligned} M_{11}(q) &= M_{22}(q) \,, \\ M_{12}(q) &= 0 \,. \end{aligned} \tag{10.52}$$

Hence, the deformation tensor is

$$\mathcal{M} = \begin{pmatrix} \mu & 0 & 0 \\ 0 & \mu & 0 \\ 0 & 0 & \mu_3 \end{pmatrix} \,. \tag{10.53}$$

As a consequence of the inferred constraints (10.52) for the D_4 singularities is that D_4^{ij} points will always be located on the two corresponding A_3 varieties, A_3^i and A_3^j. We may infer this from the following observation. In the coordinate system introduced above (cf. eq. (10.51)), the eigenvector for the third eigenvalue μ_3 is given by $v_3 = (0, 0, 1)$. The eigenvectors v_1 and v_2 both lie in the q_1q_2-plane, and since the matrix upper 2×2 matrix is degenerate we have the freedom to take them to be orthogonal to the gradient of the corresponding eigenvalue fields which will also lay in the q_1q_2-plane. This means that

$$\begin{aligned} v_1 \cdot \nabla \mu_1 &= \mu_{1,1} = 0\,, \\ v_2 \cdot \nabla \mu_2 &= \mu_{2,2} = 0. \end{aligned} \tag{10.54}$$

This proves the unfolding $D_4^{ij} \to A_3^i$ and $D_4^{ij} \to A_3^j$. For the relations between the singularity classes see section 10.7.1. For a formal proof see [203]. For the case of a non-diagonalizable deformation tensor we note that a small perturbation in the initial condition generically makes the deformation tensor diagonalizable.

The D_4^+ and D_4^- caustics

Shell-crossing for A caustics is a one-dimensional process. A direct implication of this is that the related critical points are equivalent up to diffeomorphisms. For the D family this is no longer true. Shell-crossing for the D-family is two dimensional. As a consequence, the D_4 class consist of hyperbolic (D_4^+) and elliptic (D_4^-) umbilic points, *i.e.*,

$$D_4^{ij}(t) = D_4^{+ij}(t) \cup D_4^{-ij}(t)\,. \tag{10.55}$$

In order to infer the corresponding caustic conditions we consider the two constraint quantities $Q_1(q)$ and $Q_2(q)$ (see eq. (10.52)),

$$\begin{aligned} Q_1(q) &= \frac{M_{11}(q) - M_{22}(q)}{2}\,, \\ Q_2(q) &= M_{12}(q)\,, \end{aligned} \tag{10.56}$$

which at the D_4 singularity location vanish, *i.e.*, $Q_1(q_s) = 0$ and $Q_2(q_s) = 0$. By a Taylor expansion of $Q_1(q)$ and $Q_2(q)$ in a neighbourhood around the D_4 singularity, we find that for points located in the q_1q_2-plane,

$$\begin{aligned} Q_1(q) &= a\,q_1 + b\,q_2\,, \\ Q_2(q) &= c\,q_1 + d\,q_2\,. \end{aligned} \tag{10.57}$$

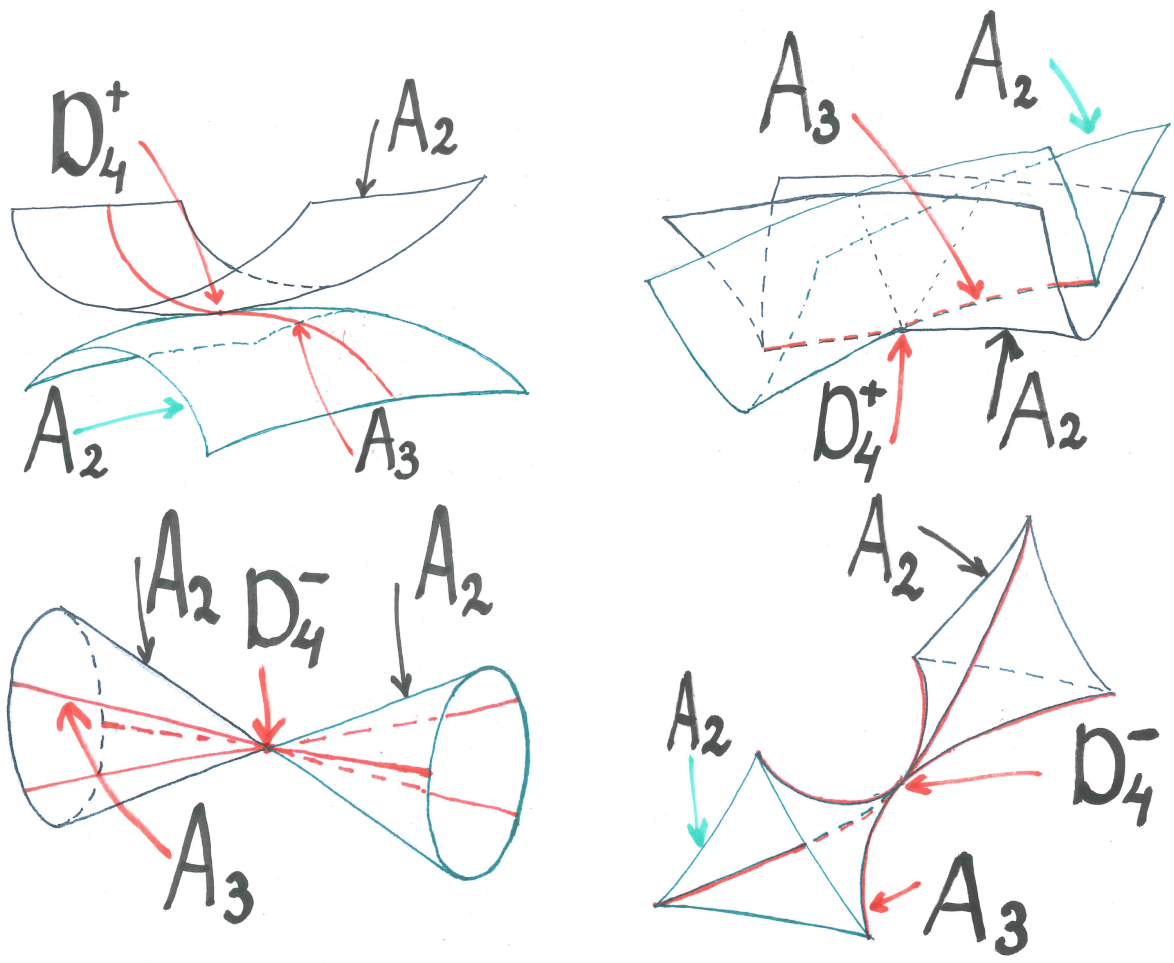


Figure 10.12: The hyperbolic/elliptic umbilic ($D_4^{\pm}$) singularities. The upper two panels show the elliptic umbilic (D_4^{+}) singularity. The lower panels show the hyperbolic umbilic (D_4^{-}) singularity. The two panels on the left are their representations in Lagrangian space and the two panels on the right their representation in Eulerian space. The black sheets are fold (A_2) sheets corresponding to one eigenvalue field. The green sheets are fold (A_2) sheets corresponding to a second eigenvalue field. The red lines are cusp (A_3) curves. The point in the center depict the hyperbolic/elliptic umbilic ($D_4^{\pm}$) singularities. The hyperbolic umbilic (D_4^{+}) and elliptic umbilic (D_4^{-}) singularity are also known as the purse and pyramid singularity.

In this expansion, we have taken the D_4 singularity to define the origin of the coordinate system. The parameters a, b, c and d are the derivatives of $Q_1(q)$ and $Q_2(q)$ at the D_4 location,

$$a = \frac{1}{2}\frac{\partial(M_{11} - M_{22})}{\partial q_1},\ b = \frac{1}{2}\frac{\partial(M_{11} - M_{22})}{\partial q_2},\ c = \frac{\partial M_{12}}{\partial q_1},\ d = \frac{\partial M_{12}}{\partial q_2}\,. \tag{10.58}$$

As proposed by [94], the determinant $S_{\mathcal{M}}$ of the corresponding Q_1Q_2 map,

$$S_{\mathcal{M}} = bc - ad = \frac{1}{2}\left[(M_{112} - M_{222})M_{112} - (M_{111} - M_{122})M_{122}\right]\,, \tag{10.59}$$

is invariant under rotations in the q_1q_2-plane [13]. In the expression above, we have used the notation

$$M_{iik} \;=\; \frac{\partial M_{ii}}{\partial q_k}\,, \quad M_{ikk} \;=\; \frac{\partial M_{ik}}{\partial q_k}\,. \tag{10.60}$$

Using the relations between the matrix elements M_{11}, M_{22} and M_{12} and the eigenvalues μ_1 and μ_2, we may recast the determinant $S_{\mathcal{M}}$ in an explicit expression incorporating these eigenvalues,

$$S_{\mathcal{M}} \;=\; \frac{1}{2}\left[(\mu_1 - \mu_2)_{,2}\mu_{1,2} - (\mu_1 - \mu_2)_{,1}\mu_{2,1}\right]\,. \tag{10.61}$$

As [94] pointed out, the transformation can be shown to consist of two branches. Their identification surfaces via a rescaling of the determinant via the multiplication by a *positive* number. We then find that the two branches correspond to two separate singularity classes of the D_4 family,

$$D_4^{\pm ij}(t) = \{q \in L | q \in A_2^i(t) \cap A_2^j(t) \wedge \mathrm{sign}(S_{\mathcal{M}}) = \pm 1\}\,, \tag{10.62}$$

where the points $q \in A_2^i(t) \cap A_2^j(t)$ are the points for whom at time t the caustic conditions are simultaneously valid for two eigenvalues, *i.e.*, $1 + \mu_i = 1 + \mu_j = 0$. Integrated over time, these $D_4^{\pm ij}(t)$ points trace out the curves $D_4^{\pm ij}$,

$$D_4^{\pm ij} \;=\; \{q \in L | q \in A_2^i(t) \cap A_2^j(t) \wedge \mathrm{sign}(S_{\mathcal{M}}) = \pm 1,\ \text{for some time } t\}. \tag{10.63}$$

For an illustration of the hyperbolic/elliptic umbilic ($D_4^{\pm}$) caustic see Figure 10.12.

[13] In fact, it can be shown that this determinant is a third-order invariant under rotations [94].

The $D_4^{\pm}$ points

The topology of the $D_4^{\pm ij}(t)$ variety changes at $D_4^{\pm}$ and D_5 points. The $D_4^{\pm}$ points are analogous to the $A_4^{\pm}$ points of the A-family. The $D_4^{\pm}$ points occur when ith and jth eigenvalue field, μ_i and μ_j, restricted to the points q in the $D_4^{\pm ij}$ variety reaches a minimum or maximum, *i.e.*,

$$\begin{aligned} D_4^{ij+} &= \{q \in L | q \in D_4^{+ij}(t) \ \wedge \ \mu_{tk}(q) \text{ max-/min. of } \mu_{tk}|_{D_4^{+ij}} (k = i \text{ or } k = j) \text{ for some } t\} \\ D_4^{ij-} &= \{q \in L | q \in D_4^{-ij}(t) \ \wedge \ \mu_{tk}(q) \text{ max-/min. of } \mu_{tk}|_{D_4^{-ij}} (k = i \text{ or } k = j) \text{ for some } t\} \end{aligned} \tag{10.64}$$

Particularly interesting is the fact that the $D_4^{\pm}$ points are always created as a pair. Two D_4^{+} points are created simultaneously, as are D_4^{-} points. By implication, also the $D_4^{\pm}$ curves (eq. (10.63)) are always created in pairs. This is in contrast to the D_5 points, which go along with the creation of a pair consisting of a D_4^{+} and a D_4^{-} point.

The D_5 caustics

The shell-crossing condition applied to the D_4^{ij} variety yields the caustic conditions for the D_5 parabolic umbilic singularity. The manifold D_4^{ij} forms a singularity in the point $q_s \in D_4^{ij}(t)$ if and only if the tangent vector $T \in T_{q_s} D_4^{ij}$ is normal to v_k^*, with $k \neq i, j$. Hence, the tangent vector $T \in \text{span}_{\mathbb{C}}\{v_i, v_k\}$, *i.e.*,

$$D_5^{ij} = \{q \in L | q \in D_4^{ij} \text{ and } \text{span}\{v_i, v_j\} \cap T_q D_4^{ij} \neq \emptyset\} \tag{10.65}$$

For three dimensional fluids in which the deformation tensor is separable in a time factor and a spatial factor, the normal $n = \nabla(\mu_{ti} - \mu_{tj})$, is orthogonal to both v_i and v_j,

$$\begin{aligned} (\mu_i - \mu_j)_{,i} \equiv v_i \cdot \nabla(\mu_{ti} - \mu_{tj}) &= 0\,, \\ (\mu_i - \mu_j)_{,j} \equiv v_j \cdot \nabla(\mu_{ti} - \mu_{tj}) &= 0\,. \end{aligned} \tag{10.66}$$

The collection of all such points form the variety

$$D_5^{ij} = \{q \in L | q \in D_4^{ij}(t) \quad \wedge \quad (\mu_i - \mu_j)_{,i} = (\mu_i - \mu_j)_{,j} = 0 \quad \text{for some time } t\}\,. \tag{10.67}$$

The D_5^{ij} lays on the A_4^i and A_4^j variety. The elliptic and hyperbolic umbilic ($D_4^{\pm}$) points merge in parabolic umbilic (D_5) points, since $D_5^{ij}(t) \subset D_4^{ij}(t)$ and

$$S_{\mathcal{M}} = \frac{1}{2}\left\{(\mu_i - \mu_j)_{,j}\mu_{i,j} - (\mu_i - \mu_j)_{,i}\mu_{j,i}\right\} = 0. \tag{10.68}$$

The D_5 points are stable singularities in the classification of Lagrangian singularities. For general dynamics they are unstable and not included in the classification scheme.

10.4.3 Caustic conditions: physical significance

For a visual appreciation of the process leading to the formation of the various classes of caustics identified in the subsections above, it is helpful to consider the phase-space manifold on which all mass elements are located in 6-D phase-space $L \times E$. This is called the *phase-space sheet* (see *e.g.* [2, 296]). The dynamical evolution of a system leads to the folding of this phase-space sheet. In a sense, we can recognize a hierarchical process in which the phase-space sheet is wrapped into an increasingly complex pattern. In this process we see the emergence of a hierarchy of complex spatial folds.

The phase-space sheet folding process generates higher order singularities within the A_2 caustic itself. These can only be identified with the help of the complementary eigenvector conditions. Restricting the manifold M to the points q_s located in the A_2 caustic, one may identify the subset of points for whom a nonzero vector T exists that (a) is tangent to the A_2 manifold and (b) is orthogonal to the span of dual eigenvectors $\mathrm{Span}\{v_j^*|j \neq i\}$. This subset fulfils the shell-crossing conditions and maps into a higher order singularity. Proceeding along the sequence of caustic conditions leads to the identification of the entire hierarchy of caustics.

The classification of A family caustic involves one eigenvalue for which $1 + \mu_i = 0$. It is straightforward to see that a similar procedure follows for configurations involving more than one eigenvalue for which $\mu_k = -1$. For example, if both $1 + \mu_1 = 0$ and $1 + \mu_2 = 0$, then T will be a vector orthogonal to the dual eigenvector v_3^*. The eigenvalue conditions therefore trace a line through three-dimensional Lagrangian space. The points q along this line are singularity points. Along this line we subsequently seek to identify higher-order singularities, by identifying points q_s along the line for which a tangent vector T exists fulfilling the shell-crossing conditions.

Conversely, note that if $1 + \mu_i \neq 0$ for all i, then there does not exist any T satisfying the general shell-crossing condition.

10.4.4 Spatial connectivity: singularities and eigenvalue fields

With the purpose to provide a guide that evokes a visual intuition for the connection between the structure and geometry of the eigenvalue fields and the formation of the various

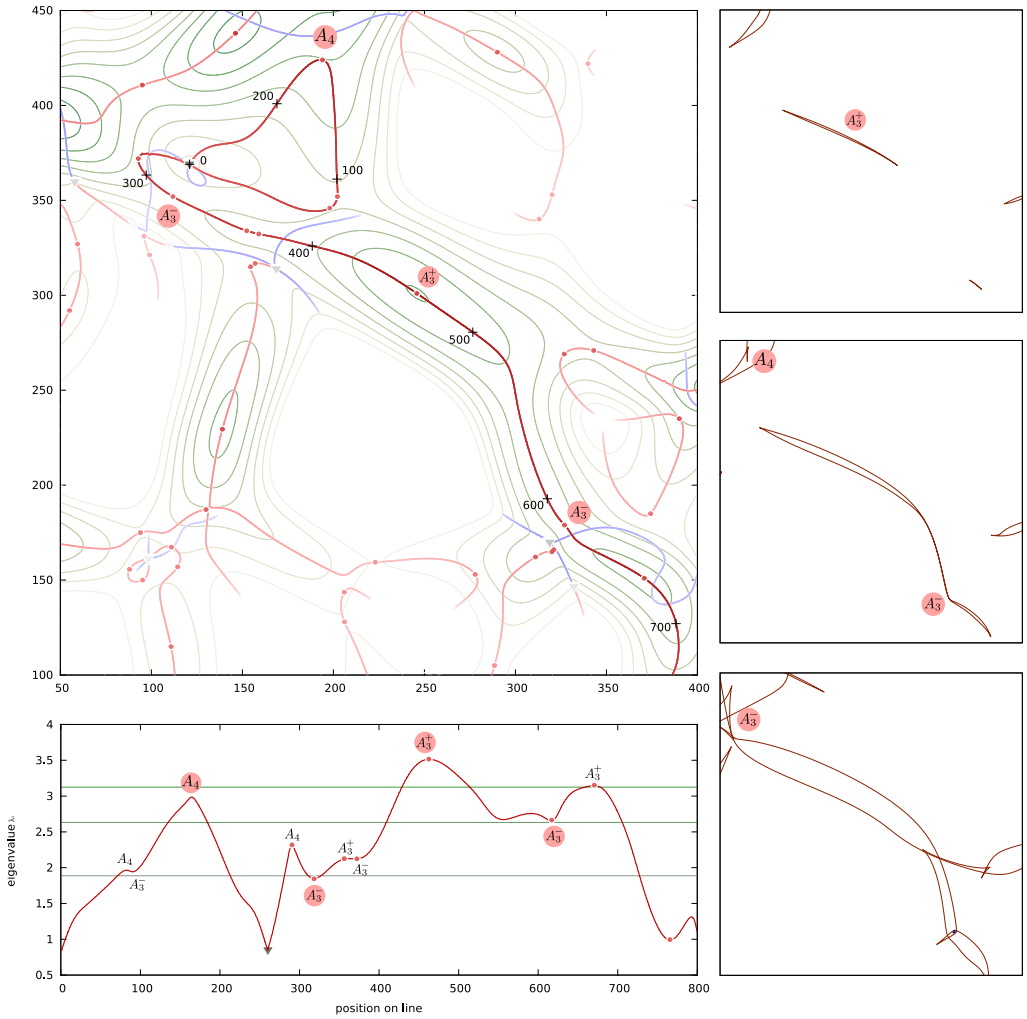


Figure 10.13: Eigenvalue field and singularity points. In the case of the two-dimensional Zel'dovich approximation (see appendix 10.A). The Zel'dovich approximation concerns the specific situation of potential flow, for which the eigenvalues and eigenvectors are real, and shows the field of the lowest eigenvalue. Top lefthand frame: the contour map illustrates the typical structure of the eigenvalue field λ_1 corresponding to a 2-D Gaussian random density field. Indicated are the positions of different A-family singularity points and varieties. The run of the A_3 line is particular noteworthy. One may appreciate how the identity of the various singularities is determined by the specific geometric character of the eigenvalue field $\lambda_1(q)$, as expressed in its derivatives. Bottom lefthand frame: the panel depicts the run of the eigenvalue field along the A_3 curve (in the contour map of top lefthand frame). Note the location of the $A_3^{\pm}$ points and A_4 points on the extrema of the curve. The green curve represents the level $b_+(t)^{-1}$ indicating which parts of the A_3 line has formed at three instances depicted in the righthand panels. Righthand panels: the three panels show the evolution, in Eulerian space, of the A_3 line. Note the appearance of the corresponding caustics and the relation between the geometry of the A_3 line in Eulerian space and the $A_3^{\pm}$ and A_4 points corresponding to the three green lines in the lower lefthand panel. This is a more extensive version of figure 8 in Hidding et al. 2014 [203].

singularities, in particular those of the A-family, we include figure 10.13. It shows a contour map representing the typical structure of the eigenvalue field μ_i. This field corresponds to a two-dimensional Gaussian random density field. For reasons of convenience, we have assumed higher eigenvalues to correspond to earlier collapse, and negative ones to no collapse (in other words, we have mirrored μ_i). The geometry and topology of the eigenvalue landscape is decisive for the occurrence of singularities. This may already be inferred from the positions of different A-family singularity points and varieties, whose positions are indicated on the contour map.

The landscape defined by the eigenvalue contours is varied, characterized by several peaks, connected by ridges with lower μ_i values. These, in turn, are connected to valleys in which μ_i attains negative values that will prevent collapse – along the direction of the eigenvector v_i – of the corresponding mass elements at any time. From the density relation (eqn. (10.7)), we know that the region of space that has undergone collapse before the current epoch (, *i.e.*, attained an infinite density) is the superlevel set of the eigenvalue field defined by the current value μ_{ti}. For each time t, the positive value contours correspond to the $A_2(t)$ fold sheets. Collapse occurs first at the maxima in the field. These mark the birth of new features, and are designated by the label of A_3^+ points. Evidently, the steepness of the hill around these maxima, *i.e.*, the gradient $\nabla\mu_i(q)$, will determine how and which mass elements around the hill will follow in outlining the emerging feature around the A_3^+ points.

The run of the A_3 line is particularly noteworthy. The key significance of the A_3 curve is evident from the observation that all A-family singularities are aligned along the ridge. In two-dimensional space, the A_3 curves delineate the points where the eigenvalues μ_i are maximal along the direction of the corresponding local eigenvector. At these points, along the eigenvector direction, the gradient of the eigenvalues is zero, *i.e.*, they are the points where the eigenvector v_i is perpendicular to the local gradient of $\nabla\mu_i$ of the eigenvalue field. Below, in section 10.4.1, we will see that this follows directly from the shell-crossing conditions that were derived in the previous section. Because of this there is a line-up and accumulation of neighbouring mass elements that simultaneously pass through the singularity. When mapped to Eulerian space, this evokes the formation of an A_3 cusp.

To illustrate the connection between A_3 curves and the various singularities even more strongly, the bottom lefthand panel depicts the run of the eigenvalue field along the A_3 curve. In particular noteworthy is the location of the $A_3^\pm$ points and A_4 points on the extrema of the curve. A prominent aspect of this is the presence of the A_3^- points at saddle junctions in the eigenvalue field. These are topologically the most interesting locations, as they evoke the merging of separate fold sheets into a single structure. In other words, they are the points where the topological structure of the field undergoes a transition and

where the connectivity of the emerging structural features is established. To establish this even more strongly, the three righthand panels of figure 10.13 represent a time sequence of the evolving structure along the A_3 line as it is mapped to its appearance in Eulerian space. The evolution follows the linear Lagrangian Zel'dovich approximation (see [344] and appendix 10.A). We may note the appearance and merging of the corresponding caustics.

10.5 Classification of singularities

The form and morphology in which the various singularities that were inventorized in the previous section will appear in the reality of a physical system depends on several aspects. The principal influence concern the dynamics of the system, as well as its dimensionality. The dynamics determines the way the fluid evolves, to a large extent via its dominant influence on the accompanying flow of the fluid. This affects the morphology of the fluid, and in particular the occurrence of singularities. Evidently, also the dimensionality of the fluid process will bear strongly on the occurrence and appearance of singularities. Higher spatial dimensions may enlarge the number of ways in which a singularity may form. It also influences the ways in which singularities can dynamically transform into one another.

In this section, we provide an impression of the variety in appearance of singularities. To this end, we will first discuss the generic singularity classification scheme that we follow. It is not the intention of this study to provide an extensive listing of all possible classes of fluids. Instead, to make clear in how different physical situations may affect the appearance of singularities, we restrict our presentation of classification schemes to two different classes of fluids. We also restrict our inventory to fluids in a three- dimensional context. It is the most representative situation, and at the same time offers a good illustration of other configurations.

10.5.1 Classes of Lagrangian fluids

To appreciate the role of the dynamics in constraining the evolution and appearance of a fluid, and that of the formation and fate of the singularities in the fluid, it is important to understand and describe its evolution in terms of six-dimensional phase space.

One way of defining phase space $\mathcal{C}$ is in terms of the Cartesian product of Lagrangian and Eulerian manifolds L and E, , *i.e.*, $\mathcal{C} = L \times E$. In this context, the phase space coordinates of a mass element are (q, x). Every point in phase space $(q, x) \in \mathcal{C}$ represents

the initial and final position q and x of a mass element at some time t. Evidently, one may also opt for the more conventional definition consisting of space coordinates x and canonical momenta p, in which case the phase space coordinate of a mass element are given by (x, p). However, for the description of Lagrangian fluid dynamics it is more convenient to follow the first convention. We should note that for this description of phase space Liouville's theorem does not apply, specifically not for the Euclidean notion of volumes.

At the initial time $t = 0$, the Lagrangian map is the identity map, *i.e.*, for all $q \in L$ $x_0(q) = q$. In phase space $\mathcal{C}$, the fluid then occupies the submanifold $\mathcal{L}_0 = \{(q, x_0(q)) \in \mathcal{C} | q \in L\}$. If we equip $\mathcal{C}$ with a symplectic structure ω, we can prove this to be a so-called Lagrangian submanifold (for a precise definition of Lagrangian submanifolds see appendix 10.D).

Differences in the dynamics of a fluid reveal themselves in particular through major differences in the phase space structure and topology of the manifolds delineated by the mass elements. To provide an impression of the differences in morphology and classification of singularities emerging in fluids of a different nature, specifically that of fluids with a different dynamical behaviour, we concentrate the discussion on two different classes of Lagrangian fluids:

1. *Generic Lagrangian fluids.*
   Lagrangian fluids for which the map $x_t : L \to E$ is a generic continuous and differentiable mapping from L to E for every time t. The dynamics does not restrict the map x to any extent. We describe the classification up to local diffeomorphisms, *i.e.*, two singularities are considered equivalent if and only if there exist local coordinate transformations, which map them into each other.

2. *Lagrangian fluids with Hamiltonian dynamics.*
   The evolution of the fluid is governed by a Hamiltonian. This assumption restricts the possible evolution of the fluid. Formally, the map x corresponds uniquely to a so-called Lagrangian map. The singularities of Lagrangian maps, known as Lagrangian singularities, are classified up to Lagrange equivalence.

Lagrangian fluids with Hamiltonian dynamics form an important class of fluids: fundamental theories of particle physics generally allow for a Hamiltonian description. Nonetheless, in a range of practical circumstances we may encounter fluids that are either more or less constrained. An example are fluids with effective dynamics. They contain friction terms

which are not described by Hamiltonian systems. Such fluid systems are less restrictive than those that are specifically Hamiltonian. On the other hand, there are also Hamiltonian fluids that are characterized by additional constraints.

10.5.2 Singularity classification: generic fluids

For the classification of singularities of generic one-family maps $x : L \to E$, with L and E three-dimensional, we follow the classification by Bruce [60]. Bruce showed that the singularities that emerge in generic mappings are equivalent to those emerging in the simple linear maps

$$x_t(q) = q + t\, u(q)\,, \tag{10.69}$$

in which u is a vector field on L. In general, the vector field $u(q)$ consists of both a longitudinal and a transversal part,

$$u(q) \;=\; u_l(q) + u_t(q)\,. \tag{10.70}$$

The longitudinal component corresponds to potential motion and has curl zero, $\nabla \times u_l = 0$, while the transversal component has divergence zero, $\nabla \cdot u_t = 0$.

The classification of singularities in general Lagrangian fluid dynamics is expressed by theorem 3. We restrict ourselves to listing the classification scheme, in terms of the generic expressions for the maps $x_t(q)$ of each of the classified singularities. In appendix 10.E we show that these normal forms indeed satisfy the corresponding caustic conditions. Note that the classification was derived using the classification of jet-spaces. It successfully cauterized the properties of caustics appearing in Lagrangian maps but did not provide a practical way to detect them in realizations.

Theorem: 3 *A stable singularity occurring in a Lagrangian fluid with generic dynamics is, up to local diffeomorphisms, equivalent to one of the following classes:*

Singularity class	*Map* $x_t(q)$	*Singularity name*
A_1	$x_t(q) = q$	*trivial case*
A_2	$x_t(q) = q + t\,(0, 0, q_3^2 - q_3)$	*fold*
A_3	$x_t(q) = q + t\,(0, 0, q_1 q_3 + q_3^3 - q_3)$	*cusp*
A_4	$x_t(q) = q + t\,(0, 0, q_1 q_3 + q_3^4 - q_3)$	*swallowtail*
A_5	$x_t(q) = q + t\,(0, 0, q_1 q_3 + q_2 q_3^2 + q_3^5 - q_3)$	*butterfly*
$D_4^{\pm}$	$x_t(q) = q + t\,(0, q_2 q_3 - q_2, q_2^2 \pm q_3^2 + q_1 q_2 - q_3)$	*hyperbolic/elliptic*
$A_3^{\pm}$	$x_t(q) = q + t\,(0, 0, (q1^2 \pm q_2^2) q_3 + q_3^3 - q_3)$	
$A_4^{\pm}$	$x_t(q) = q + t\,(0, 0, q_1 q_3 \pm q_2^2 q_3^2 + q_3^4 - q_3)$	

Note: The normal forms $x_t(q)$ *form the singularity at the origin* $q = 0$*, at* $t = 1$*. The first five singularity classes are the A-family. The subsequent class is the D-family. The last two are the normal forms of the* A_3 *and* A_4 *points. The* A_k *class has co-rank* 1 *and co-dimension* $k - 2$*. The* $D_4^{\pm}$ *singularities have co-rank* 2 *and are one-dimensional.*

10.5.3 Singularity classification: Hamiltonian fluids

The evolution of Lagrangian fluids with Hamiltonian dynamics is more constrained than that of generic Lagrangian fluids. As the fluid develops complex multistream regions, the phase space submanifold $\mathcal{L}_t = \{(q, x_t(q)) | q \in L\}$ for fluids with Hamiltonian dynamics remains a Lagrangian submanifold.

A key step in evaluating the emerging singularities is that of connecting the displacement map $s_t(q)$ to the Lagrangian map. In appendix 10.D.2, we describe in some detail how a given Lagrangian map can be constructed from a Lagrangian submanifold $\mathcal{L}$. A Lagrangian map can develop regions in which multiple points in the Lagrangian manifold are mapped to the same point in the base space.

Lagrangian singularities are those points at which the number of pre-images of the Lagrangian map undergoes a change. Lagrangian catastrophe theory [14, 21] classifies the stable singularities. This refers to the stability of singularities with respect to small deformations of the Lagrangian manifold of $\mathcal{L}$. This is true up to Lagrangian equivalence, a concept that is a generalization of equivalence up to coordinate transformation. For a more formal and precise definition of Lagrangian equivalence see appendix 10.D.

It can be demonstrated (see [21]) that every Lagrangian map $l : \mathcal{L} \to \mathcal{C} \to E$ is locally

Lagrangian equivalent to a so-called gradient map, *i.e.*, the map x_t is locally equivalent to

$$x_t(q) = \nabla_q S_t \,, \tag{10.71}$$

for some $S_t : L \to \mathbb{R}$. By recasting S_t in terms of a function $\Psi_t : L \to \mathbb{R}$,

$$S_t \;=\; \frac{1}{2}q^2 + \Psi_t(q) \,, \tag{10.72}$$

we find that locally the map x can be written in the form

$$x_t(q) = q + \nabla_q \Psi_t(q) \,. \tag{10.73}$$

Evidently, this implies that the displacement map is longitudinal, and that the corresponding Jacobian $\partial s_t/\partial q$ is symmetric.

The classification of singularities of a Lagrangian fluid with Hamiltonian dynamics is expressed by theorem 4. In appendix 10.E it is shown that these normal forms indeed satisfy the corresponding caustic conditions. For proofs we refer to Arnol'd [14]. Note that the classification was derived using the classification of critical points of scalar functions and the theory of generating functions. It successfully characterized the properties of caustics appearing in Lagrangian maps but did not provide a practical way to detect them in realizations.

Theorem: 4 *A stable Lagrangian singularity of a Lagrangian fluid with Hamiltonian dynamics, is locally Lagrange equivalent to one of the following classes:*

Singularity class	*Map* $x_t(q)$	*Singularity name*
A_1	$x_t(q) = q$	*trivial case*
A_2	$x_t(q) = q + t\,(0, 0, q_3^2 - q_3)$	*fold*
A_3	$x_t(q) = q + t\,\big(\frac{1}{2}q_3^2, 0, q_3(q_1 - 1)\big)$	*cusp*
A_4	$x_t(q) = q + t\,\big(\frac{1}{2}q_3^2, 0, q_1 q_3 + q_3^4 - q_3\big)$	*swallowtail*
A_5	$x_t(q) = q + t\,\big(\frac{1}{2}q_3^2, \frac{1}{3}q_3^3, q_1 q_3 + q_2 q_3^2 + q_3^5 - q_3\big)$	*butterfly*
$D_4^{\pm}$	$x_t(q) = q + t\big(\pm q_1 q_2 - q_1,$ $\pm\big(\frac{1}{2}q_1^2 + \frac{3}{2}q_2^2\big) + 2q_2 q_3 + 2q_2^3 - q_2, q_2^2\big)$	*hyperbolic/elliptic*
D_5	$x_t(q) = q + t\,(0, q_2^3 - q_2, q_3^3 - q_3)$	*parabolic*
$A_3^{\pm}$	$x_t(q) = q + t\,(q_1 q_3^2, \pm q_2 q_3^2, (q_1^2 \pm q_2^2)q_3 + q_3^3 - q_3)$	
$A_4^{\pm}$	$x_t(q) = q + t\,\big(\frac{1}{2}q_3^2, \pm\frac{2}{3}q_2 q_3^3, q_1 q_3 \pm q_2^2 q_3^2 + q_3^4 - q_3\big)$	

Note: The normal forms $x_t(q)$ form the singularity at the origin $q = 0$, at $t = 1$. The first five singularity classes are the A-family. The subsequent two are the D-family. The last two are the normal forms of the A_3 and A_4 points. The A_k class has co-rank 1 and co-dimension $k - 2$. The D_k singularities have co-rank 2 and co-dimension $k - 2$ [14].

Comparing the classification schemes for generic Lagrangian singularities and those for Lagrangian fluids with Hamiltonian dynamics, we may note the similarities. Both classifications have an A and a D family. It can be demonstrated that the A singularity classes of the scheme for Lagrangian fluids with Hamiltonian dynamics are contained in those corresponding to the generic Lagrangian fluid. Concretely, this means that a displacement field corresponding to the Hamiltonian A_k class is also an element of the generic A_k class.

The D families are some what different. The Hamiltonian D_4 class is contained in the generic D_4 class. However, the Hamiltonian D_5 class has no analogue in the generic classification scheme. This is a result of the D_5 singularity not being stable under coordinate transformations.

A final remark concerns the singularity classification schemes for higher dimensional fluids. For these a more elaborate classification scheme applies. This classification scheme is described in appendix 10.D.

10.5.4 Unfoldings

Singularities generally change their class upon small, but finite, deformations of the displacement map s_t. The corresponding evolution of a singularity follows the universal unfolding process of singularities. The general behavior is described in the following unfolding diagram, in which the arrows indicate the singularity into which specific singularities can transform.

$$\begin{array}{ccccccccc} A_1 & \leftarrow & A_2 & \leftarrow & A_3 & \leftarrow & A_4 & \leftarrow & A_5 \\ & & & & & \nwarrow & & \nwarrow & \\ & & & & & & D_4 & \leftarrow & D_5 \end{array}$$

For $i \geq 2$, the A_i singularities decay into A_{i-1} singularities. For $i \geq 5$, the D_i singularities decay into either A_{i-1} or D_{i-1} singularities. In section 10.7 we will describe how the decay of singularities is connected to the evolution of the large-scale structure in the Universe and in outlining the spine of the cosmic-web.

10.6 The caustic skeleton & the cosmic-web

The process of formation and evolution of structure in the Universe is driven by the gravitational growth of tiny primordial density and velocity perturbations. When it reaches a stage at which the matter distribution starts to develop nonlinearities, we see the the emergence of complex structural patterns. In the current universe we see this happening at Megaparsec scales. On these scales, cosmic structure displays a marked intricate weblike pattern. Prominent elongated filamentary features define a pervasive network. Forming the dense boundaries around large tenuous sheetlike membranes, the filaments connect up at massive, compact clusters located at the nodes of the network and surround vast, underdense and near-empty voids.

The gravitational structure formation process is marked by vast migration streams, known as cosmic flows. Inhomogeneities in the gravitational force field lead to the displacement of mass out of the lower density areas towards higher density regions. Complex structures arise at the locations where different mass streams meet up. Gravitational collapse sets in as this happens. In terms of six-dimensional phase space, it corresponds to the local folding of the phase space sheet along which matter – in particular the gravitationally dominant dark matter component – has distributed itself.

10.6.1 The caustic skeleton

The positions where streams of the dark matter fluid cross are the sites where gravitational collapse occurs. The various types of caustics described and classified in our study mark the different configurations in which this process may take place. Their locations trace out a Lagrangian skeleton of the emerging cosmic-web, marking key structural elements and establishing their connectivity (also see the discussion in [203]). In other words, the A_3, A_4, A_5, D_4, D_5 varieties, in combination with the corresponding $A_3^{\pm}, A_4^{\pm}$, and $D_4^{\pm}$ points, are the dynamical elements whose connectivity defines the weaving of the the cosmic-web [344, 49, 319, 12, 71]. On the basis of this observation, we may obtain the skeleton of the cosmic-web by mapping the caustic varieties defined above to Eulerian space with the Lagrangian map x_t. Following the identification of the various caustic varieties and caustic points in Lagrangian space, the application of the map x_t will produce the corresponding weblike structure in Eulerian space.

Of central significance in our analysis and description of the cosmic-web is the essential role of the deformation tensor *eigenvector* fields in outlining the caustic skeleton and in

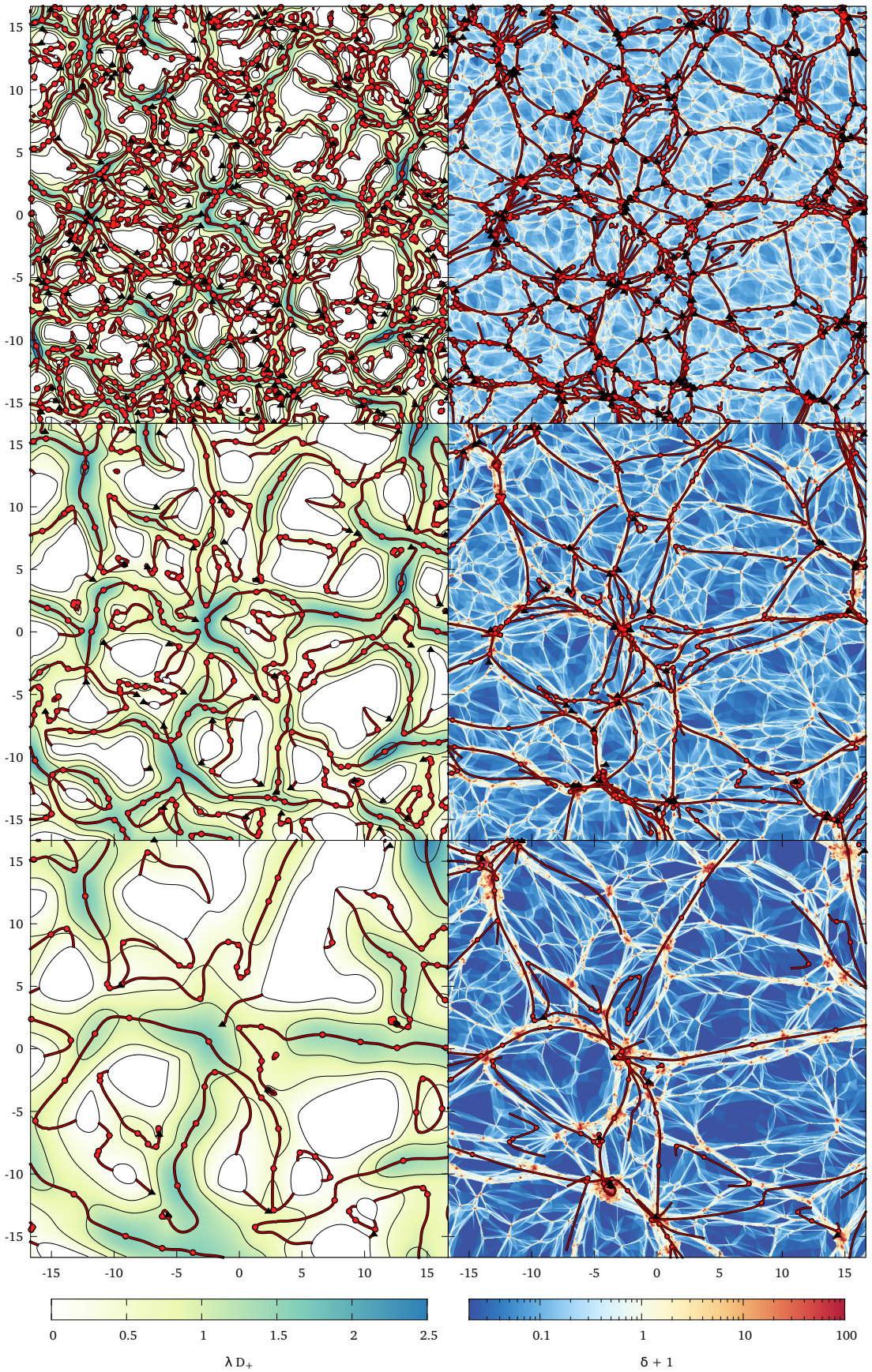

15
10
5
0
-5
-10
-15
-15 -10 -5 0 5 10 15
0 0.5 1 1.5 2 2.5
λ D+
0.1 1 10 100
δ + 1

Figure 10.14: Spatial distribution of singularities in the Lagrangian and Eulerian cosmic-web. The figure compares the spine of the cosmic-web with the mass distribution in a 2-D N-body simulation. Left panel: initial field of density fluctuations and the skeleton of identified singularities/catastrophes. Right panel: density field of an evolved 2D cosmological N-body simulation, in which the Lagrangian skeleton of singularities is mapped by means of the Zel'dovich approximation. From Feldbrugge et al. [138].

establishing the spatial connections between the various structural features. So far, Lagrangian studies of the cosmic-web have usually been based on the role of the *eigenvalues* of the deformation tensor (for recent work see [76, 338, 234]). Nearly without exception, they ignore the information content of the eigenvectors of the deformation tensor. In this work we actually emphasize that the eigenvectors are of key importance in tracing the spatial locations of the different types of emerging caustic features and, in particular, in establishing their mutual spatial connectivity. This important fact finds its expression in terms of the *caustic conditions* that we have derived in this study.

The study by Hidding et al. [203] illustrated the important role of the deformation field eigenvectors in outlining the skeleton of the cosmic-web, for the specific situation of A_3 cusp lines in the 2-D matter distribution evolving out of a Gaussian initial density field. The present study describes the full generalization for the evolving matter distribution (a) for each class of emerging caustics in (b) in spaces of arbitrary dimension D.

10.6.2 The 2-D caustic skeleton and cosmic-web

A telling and informative illustration of the intimate relationship between the caustic skeleton defined by the derived caustic conditions and the evolving matter distribution is that offered by the typical patterns emerging in the two-dimensional situation. Figure 10.14 provides a direct and quantitative comparison between the caustic skeleton of the cosmic-web and the fully nonlinear mass distribution in an N-body simulation. The three panels in the lefthand column show the Lagrangian skeleton for a two-dimensional fluid. The fluid is taken to evolve according to the Zel'dovich approximation [344] (see appendix 10.A), which represents a surprisingly accurate first-order Lagrangian approximation of a gravitationally evolving matter distribution (see *e.g.* [298]). The initial density field of the displayed models is that of a Gaussian random density field [4, 26], which according to the latest observations and to current theoretical understanding is an accurate description of the observed primordial matter distribution [274, 225, 84].

To enable our understanding of the hierarchical process of structure formation and the resulting multiscale structure of the cosmic-web, we assess the caustic structure of the Lagrangian matter field at three different resolutions. In figure 10.14 the field resolution decreases from the top panels to the bottom panels, as the initial density field was smoothed by an increasingly large Gaussian filter. The contour maps that form the background in these panels represent the resulting initial density fields. The red lines trace the A_3 variety, *i.e.*, the A_3 lines, for the largest eigenvalue μ_1 field (also see fig. 10.13 to appreciate how they are related). Also the $A_3^{\pm}$ points and $D_4^{\pm}$ points are shown, the first as red dots, the latter as black triangles.

The resulting weblike structure in Eulerian space is depicted in the corresponding right-hand panels. The A_3 lines, $A_3^{\pm}$ points and $D_4^{\pm}$ points are mapped to their Eulerian location by means of the Zel'dovich approximation. The red lines, red dots and black triangles represent the Eulerian skeleton corresponding to the Zel'dovich approximation. These are superimposed on the density field of the corresponding N-body simulations. The comparison between the latter and the Eulerian skeleton reveal that the caustic skeleton – the assembly of A_3 lines, $A_3^{\pm}$ points and $D_4^{\pm}$ points – trace the principal elements and connections of the cosmic-web seen in the N-body simulations remarkably well (see Table 10.1 for the identification of the lines and points to the cosmic-web). Moreover, by assessing the caustic structure at different resolutions of the density field, one obtains considerable insight into the multiscale structure and topology of the cosmic-web.

10.6.3 The 3-D caustic skeleton and cosmic-web

One of the unique features facilitated by the caustic conditions that we have derived in the previous sections is the ability to go beyond the two-dimensional case and construct and explore the full caustic skeleton of the three-dimensional mass distribution. In the case of the skeleton of the cosmic-web defined by the three-dimensional mass distribution, the cusp (A_3) sheets correspond to the walls or membranes of the large scale structure [49, 319, 71, 240]. The swallowtail (A_4) and elliptic/hyperbolic umbilic ($D_4^{\pm}$) lines correspond to the filaments of the cosmic-web and the butterfly (A_5) and parabolic umbilic (D_5) points correspond to the cluster nodes of the network [49, 319, 11, 71, 240]. The identification of the caustics in the three dimensional cosmic-web is summarized in table 10.1.

To appreciate the impressive level at which the caustic skeleton is outlining the three-dimensional weblike mass distribution, figure 10.15 provides an instructive illustration. The figure depicts elements of the caustic skeleton of the Zel'dovich approximation in a $200h^{-1}$ Mpc box. The resulting skeleton is superposed on the log density field of a dark

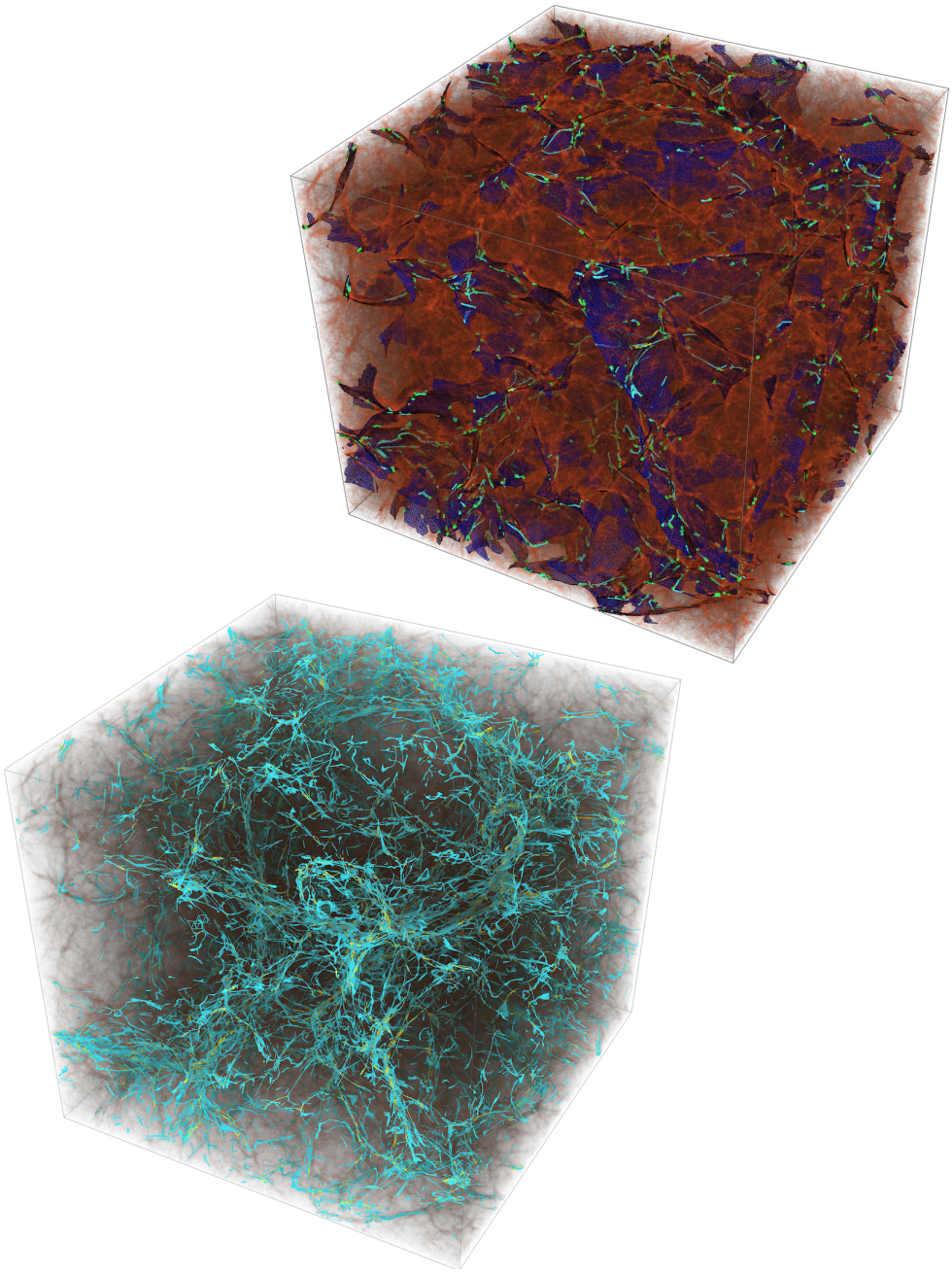


Figure 10.15: The log density field of a dark matter N-body simulation with ΛCDM cosmology in a box of $200h^{-1}$ Mpc with 512^3 particles and elements of the caustic skeleton of the Zel'dovich approximation [240]. Top right panel: the cusp (A_3) sheets (dark blue), the swallowtail (A_4) lines (light blue) and the elliptic/hyperbolic umbilic lines (yellow) corresponding to the lowest eigenvalue field of the caustic skeleton. Note that the Zel'dovich approximation concerns a potential flow, which means that the eigenvalue fields can be ordered. The initial density field was smoothed on the scale $6.3h^{-1}$ Mpc. Bottom left panel: the swallowtail (A_4) lines (light blue) and the elliptic/hyperbolic umbilic lines (yellow) corresponding to the lowest eigenvalue field of the caustic skeleton. The initial density field is smoothed at $3.1h^{-1}$ Mpc.

Singularity class	Singularity name	Feature in the 2D cosmic-web	Feature in the 3D cosmic-web
A_2	fold	shell-crossed region	shell-crossed region
A_3	cusp	filament	wall or membrane
A_4	swallowtail	cluster or knot	filament
A_5	butterfly	not stable	cluster or knot
D_4	hyperbolic/elliptic	cluster or knot	filament
D_5	parabolic	not stable	cluster or knot

Table 10.1: The identification of the different caustics in the 2- and 3-dimensional cosmic-web

matter N-body simulation in a ΛCDM cosmology with 512^3 particles [240]. We should emphasize that the Zel'dovich approximation is linear and that the corresponding skeleton is completely local in the initial conditions. While a full and detailed analysis of these three-dimensional weblike patterns is the subject of an upcoming accompanying paper [145], the illustrations of figure 10.15 already give a nice impression of the ability of the caustic conditions to outline the spine of the cosmic-web.

The top righthand panel contains the cusp (A_3) sheet (dark blue colour) and the swallowtail (A_4) and elliptic/hyperbolic umbilic ($D_4^{\pm}$) lines (light blue colour) corresponding to the lowest eigenvalue field, superimposed on the density field of the N-body simulation (red shaded log density field values). The pattern concerns the caustics obtained for a displacement field that is filtered at a length scale of $6.3h^{-1}$ Mpc. Close inspection reveals the close correspondence between the cusp sheets of the caustic skeleton and the flattened - two-dimensional - features in the mass distribution of the cosmic-web. Notwithstandig this, one may also observe that the two-dimensional skeleton does not capture all the structures present in the N-body simulation. This is predominantly an issue of scale, as the corresponding displacement field cannot resolve and trace features whose size is more refined than the $6.3h^{-1}$ Mpc filter scale.

An impression of the more refined structure can be obtained from the bottom left panel of figure 10.15, which follows the line-like elements of the caustic skeleton at a length scale of $3.1h^{-1}$ Mpc. More specifically, it shows the swallowtail (A_4) and elliptic/hyperbolic umbilic ($D_4^{\pm}$) lines of the caustic skeleton. The correspondence of these with the prominent and intricate filamentary pattern in the cosmic mass distribution is even more outstanding than that of the A_3 sheets with the membranes in the density field. It is important to realize, and

emphasize, that blue curves were generated using only the eigenvalue field corresponding to the first collapse. This already creates a filament in the network of caustics, without the need to involve the second eigenvalue. In other words, collapse along the second eigenvector is not necessary to create a filament-like structure (also see [203]). This leads to a radical new insight on structure formation, in that it suggests the different possible late-time morphologies for filaments [204]. We may even relate this to the prominence of the corresponding filamentary features: as they concern features that have experienced collapse along two directions, the umbilic $D_4^{\pm}$ filaments will have a higher density and contrast than the filigree of more tenuous $A_4^{\pm}$ filaments. An additional observation of considerable interest is that the line-like A_4 and $D_4^{\pm}$ features trace the connectivity of the cosmic-web in meticulous detail.

10.6.4 Caustic density profiles

Also of decisive interest in their embedding in the cosmic-web, is the expected mass distribution in and around the various classes of caustics.

Vesilev [321] inferred the density profiles of the various classes of singularities, in case they emerge as a result of potential motion in a collision-less self-gravitating medium. For each of the mass concentrations in and around these singularities, he found scale free power-law profiles. The radially average profiles display the following decrease of density $\rho(r)$ as a function of radius r.

Singularity class	Singularity name	Profile $\rho(r)$
A_2	fold	$\rho(r) \propto r^{-1/2}$
A_3	cusp	$\rho(r) \propto r^{-2/3}$
A_4	swallowtail	$\rho(r) \propto r^{-3/4}$
A_5	butterfly	$\rho(r) \propto r^{-4/5}$
D_4	hyperbolic/elliptic	$\rho(r) \propto r^{-1}$
D_5	parabolic	$\rho(r) \propto r^{-1} \log(1/r)$

With respect to these radially averaged profiles, we should realize that the mass distribution in and around the singularities is highly anisotropic. This is true for any dimension in which we consider the structure around the singularities.

Notwithstanding this, we do observe that the steepest density profiles are those around the point singularities A_5 and D_5. However, they are mere transient features that will only exist for a single moment in time. The point singularities A_4 and D_4 display a less pronounced behaviour. However, they move over time. Also, we see that the cusp singularity A_3 possesses a steeper mass distribution that that in and around the sheet singularity A_2.

10.6.5 Higher order Lagrangian perturbations

Evidently, the details of the dynamical evolution will bear a considerable influence on the developing caustic structure. This not only concerns the dynamics of the system itself, but also its description. The examples that we presented in the previous sections showed the caustic features developing as the dynamics is predicated on the first-order Lagrangian approximation of the Zel'dovich formalism [344]. The visual comparison with the outcome of the corresponding N-body simulations demonstrated the substantial level of agreement. Nonetheless, given the nature of singularities, the process of caustic formation might be very sensitive to minor deviations of the mass element deformations and hence the modelling of the dynamics. This may even strongly affect the predicted population of caustics and their spatial organization in the skeleton of the cosmic-web. Some indications on the level to which the spatial mass distribution is influenced may be obtained from an early series of papers by Buchert and collaborators [61, 62, 64, 63, 65], who were the first to explore the formation of structure in higher-order Lagrangian perturbation schemes and investigate in how far they would effect the occurrenc and location of multistream regions. An important finding from their work is that 2nd order effects are substantial, while 3rd order ones are minimal. Elaborated and augmented by additional work [52, 294], 2nd order Lagrangian perturbations – usually designated by the name 2LPT – have been established as key ingredients of any accurate analytical modeling of cosmic structure growth. In a follow-up to the present study, we investigate in detail the repercussions of different analytical prescriptions for the dynamical evolution of the cosmic mass distribution for the full caustic skeleton of the cosmic-web.

In addition to 2LPT, we will systematically investigate the caustic skeleton in the context of the *adhesion approximation* [175, 298, 322, 176, 205, 202]. Representing a fully nonlinear extension of the Zel'dovich formalism, it includes an analytically tractable gravitational source term for the later nonlinear stages. It accomplishes this via an artificial viscosity term that emulates the effects of gravity, resulting in the analytically solvable Burger's equation. With the effective addition of a gravitational interaction term for the emerging structures, unlike the Zel'dovich approximation the adhesion model is capable of

following the hierarchical buildup of structure and the cosmic-web [205, 204, 202]. At early epochs, the resulting matter streams coincide with the ballistic motion of the Zel'dovich approximation. At the later stages, as the mass flows approach multistream regions a solid structure is created at the shell-crossing location. Matter inside these structures is confined to stay inside, while outside collapsed structures the results from the Zel'dovich approximation and adhesion are identical. The caustics from the Zel'dovich approximation are compressed to infinitesimally thin structures, hence unifying the Zel'dovich' idea of collapsed structures in terms of shell crossing with a hierarchical formation model. While offering a complete model for the formation and hierarchical evolution of the cosmic-web, it does accomplish this by seriously altering the flow pattern involved in the buildup of cosmic structure. This, in turn, is expected to affect at least to some extent the properties and evolution of the caustic population and its connectivity.

10.6.6 Gaussian statistics of the caustic skeleton

In addition to characterizing the geometric and topological outline of the cosmic-web in terms of the caustic skeleton, our study points to another important and related application of the formalism described. The fact that the linear Zel'dovich approximation provides such an accurate outline of the skeleton of the cosmic-web establishes an important relation between the primordial density and flow field and the resulting cosmic-web. Via the Zel'dovich approximation, we may relate the caustic skeleton directly to the statistical nature and characteristics of the primordial density field. In other words, we may directly relate the structure of the cosmic-web to the nature of the Gaussian initial density field. This, in turn, establishes a direct link between the geometric and topological properties of the cosmic-web and the underlying cosmology. Hence a probabilistic analysis of the caustic skeleton may define a path towards a solidly defined foundation and procedure for using the structure of the observed cosmic-web towards constraining global cosmological parameters and the cosmic structure formation process.

The fact that we may invoke Gaussian statistics facilitates the calculation of a wide range of geometric and topological characteristics of the cosmic-web, as they are directly related to the primordial Gaussian deformation field, its eigenvalues and eigenvectors. For an example of such a statistical treatment of 2-dimensional fluids, we refer to [136]. It describes how one may not only analytically compute the distribution of maxima, or minima, but also the population of singularities and the length of caustic lines. In an accompaying study, we present an extensive numerical analysis of the statistics of 2- and 3-dimensional caustic skeleton will follow in [145]. This will establish the reference point for

the subsequent solid analytical study of interesting geometric properties of the cosmic-web (for the initial steps towards this program see [138]).

This will represent a major extension of statistical descriptions that were solely based on the eigenvalue fields. The latter would make it possible to study the number density of clusters and void basins, make predictions on the statistical properties of angular momentum, and even several aspects of the cosmic skeleton (*e.g.* [116, 276]). As we have argued extensively in previous sections, it is only by invoking the information contained in the corresponding eigenvector fields that we may expect to obtain a more complete census of intricate spatial properties of the cosmic-web.

10.7 Dynamics and evolution of caustics

The caustic conditions presented in this study reveal the profound relationship between the various classes of singularities that may surface in Lagrangian fluids. Besides the aspect of the identification and classification of singularities, we need to have insight in the transformation and evolution of caustics and caustic networks that accompanies the dynamical evolution of a fluid. The evolution of the fluid, dictated by the dynamics of the system, generally involves the development of ever more distinctive structures and the proliferation of complex structural patterns.

Tracing the evolution of a fluid starts at an initial time $t = 0$. At that time, the displacement map s_t is the zero map. Amongst others, this implies the fluid does not (yet) contain singularities. Starting from these near uniform initial conditions, the structure in the evolving fluid becomes increasingly pronounced. The phase space sheet that it occupies in six-dimensional space gets increasingly folded. Its projection on Euclidian space follows this process, and it is as a result of the folding process that we see the fluid developing singularities. While the dynamical evolution proceeds to more advanced stages, we not only see the appearance of more singularities, but also the transformation of one class of singularities into another one. A complementary process that may underlie the changes of local geometry that of the merging of singularities into a new singularity, itself a manifestation of the hierarchical buildup of structural complexity.

The eigenvalue landscape in figure 10.13 offers an instructive tool for facilitating and guiding our understanding and visual intuition for the iterative folding of singularities in phase space and the accompanying caustic transformations.

10.7.1 Caustic mutations and transformations: evolutionary sequence

The dynamical evolution of a fluid goes along with a rich palet of local processes. These involve fundamental mutations in the local singularity structure that lead to significant topological changes of the spatial pattern forming in the fluid. In some systems and situations this will be a key element in the hierarchical buildup of structure.

The fundamental notion in these structural mutations in the evolving fluid is that of the ruling dynamics of the system evoking changes in the deformation field. Small deformations will lead to the decay of singularities into different ones belonging to other singularity classes. Conversely, they may get folded according to a rigid order.

The sequence of singularity mutations is not random and arbitrary. Due to the strict geometric conditions and constraints corresponding to the various singularities, expressed in the caustic conditions discussed extensively in this study, a given singularity is only allowed to transform into a restricted set of other singularities. Conversely, a given singularity may only have emanated from a restricted set of other singularities.

In most situations a particular singularity can have decayed from only one distinctive class of singularities. Some may have descended from two other singularity classes. Likewise, most singularities can decay only into one distinctive other class of singularity. This is true for all A-family singularities. D-family singularities have a richer diversity of options, with the D_5 points being able to decay into 3 different ones, while the D_4 points may decay into 2 distinct A_3 points. The entire singularity transformation and unfolding sequence may be transparently summarized in the unfolding diagram below.

$$\begin{array}{ccccccccc} & & A_2^i & \leftarrow & A_3^i & \leftarrow & A_4^i & \leftarrow & A_5^i \\ & \swarrow & & & & \nwarrow & & \nwarrow & \\ A_1 & & & & & & D_4^{ij} & \leftarrow & D_5^{ij} \\ & \nwarrow & & & & \swarrow & & \swarrow & \\ & & A_2^j & \leftarrow & A_3^j & \leftarrow & A_4^j & \leftarrow & A_5^j \end{array}$$

The unfolding diagram follows directly from Lagrangian catastrophe theory, although it can also be derived from the caustic conditions. The unfoldings of an A_k^i singularities into an A_{k-1}^i singularities, with $k \geq 2$, follow trivially from the caustic conditions. The same holds for the unfolding of the D_5^{ij} singularities into the D_4^{ij} singularities. The decay from

the D_4 to the A_3 singularities are proven in section 10.4.2. The mutations $D_5^{ij} \rightarrow A_4^i$ and $D_5^{ij} \rightarrow A_4^j$ follow directly since the shell-crossing of the D_5^{ij} caustic is analogous to the shell-crossing condition on the A_4^i and A_4^j caustics.

10.7.2 Singularity transformations

The principal family of singularities – principal in terms of rate of occurrence and spatial dominance – is the A-family. They are induced by singularities in the geometric structure of one of the eigenvalue fields. In physical terms, they involve one-dimensional collapse on to the emerging singularity. Of a more challenging nature within the evolutionary unfolding of the patterns emerging in fluid flow is the formation of the D-family of singularities. They occur when two fold sheets corresponding to different eigenvalue fields intersect. Amongst others, this means that the D singularities connect A singularities corresponding to two eigenvalue fields.

Evolving A-family caustics

The most prominent and abundant singularities are those of the two-dimensional fold sheets $A_2^i(t)$. In Eulerian space, they mark the regions where mass elements are turned inside out as the density attains infinity. This happens while they represent the locations where separate matter streams are crossing each other. As time proceeds, the fold sheets $A_2^i(t)$ sweep over an increasingly larger Lagrangian region. Ultimately, integrating over time, they mark an entire Lagrangian volume, which is labelled as A_2^i. The A_2^i set forms a three-dimensional variety.

When we wish to identify where a particular individual fold sheet is born, we turn to the cusp points A_3^{i+}. They are the points on the fold sheets where the corresponding eigenvalue field attains an extremum. Because of this, they mark the sites of birth of the fold singularities. As the $A_2^i(t)$ sheets unfold, at the edges their surface gets wrapped in a higher order singularity, the cusp curves $A_3^i(t)$. In time, these curves move through space and trace out cusp sheets A_3^i. In the context of the Megaparsec scale matter distribution in the Universe, the cusp sheets are to be associated with the walls or membranes in the cosmic-web [49, 319, 11, 71, 240].

A dynamically interesting process occurs at the cusp points A_3^{i-}, which are the saddle points of the corresponding eigenvalue field μ_{ti} that at a given time are encapsulated by the fold sheet A_2^i. At the A_3^{i-} points, we see the merging or annihilation of fold sheets A_2^i into

a larger structure (cf. figure 10.13). Mathematically, they mark the key locations where the topology of the eigenvalue field changes abruptly. Physically, they are associated with the merging of separate structural components, a manifestation of the hierarchical buildup of structural complexity [319, 71].

Also the cusp curves $A_3^i(t)$ can get folded. In Eulerian space, the folding of the cusp curves manifests itself as $A_4^i(t)$ swallowtail points. As time proceeds, these points move through space and define the swallowtail curve A_4^i. It is of interest to note that the swallowtail curve is embedded in the cusp sheet, *i.e.*, $A_4^i \subset A_3^i$. In the context of the cosmic structure formation process, the swallowtail curves outline and trace perhaps the most outstanding feature of the cosmic-web, the pronounced elongated filaments that form the of spine the weblike network [319, 11, 71].

Also these features build up in a hierarchical process of small filaments merging into ever larger and more prominent arteries. In the context of the evolving singularity structure that we study, this process is represented by the A_4^{i+} points and A_4^{i-} points. They define the decisive junctions where significant changes in topology occur. For the $A_4^{i\pm}$ points this concerns their identity in the gradient of the eigenvalue field, in which the A_4^{i+} are maxima and minima and A_4^{i-} points are the saddle points. The implication of this is that cusp curves get created or annihilated at A_4^{i+} points, while they merge or separate at A_4^{i-} points.

The final morphological constituent in this structural hierarchy of singularities is that of the butterfly points A_5^i. They conclude the A-family of singularities, *i.e.*, the family of singularities that correspond to the spatial characteristics of the field of one eigenvalue μ_i. The swallowtail curves A_4^i get folded at A_5^i butterfly points. In the three-dimensional structural pattern that formed in the fluid, these will represent nodes. In the cosmic-web, they define the nodal junctions, connecting to the various filamentary extensions that outline its spine [49, 77, 319, 11, 71]. In principle, for a given initial field and dynamical evolution, one might use these identifications to *e.g.* evaluate how many filaments are connected to the network nodes [12, 276].

Evolving D-family caustics

The $A_2^i(t)$ and $A_2^j(t)$ sheets, with $i \neq j$, intersect in the elliptic and hyperbolic umbilic points $D_4^{\pm ij}(t)$. In contrast to the A family of singularities, the collapse into D singularities is two-dimensional. It leads to the birth of the socalled *umbilic* points. Over time, they

trace out the umbilic curve $D_4^{\pm ij}$. The collapse process may occur in two distinctive ways, indicated by the labels + and −.

The topology of the variety $D_4^{\pm ij}(t)$ changes at $D_4^{ij\pm}$ and D_5 points. An interesting characteristic of umbilic curves is that they are always created or annihilated in pairs. The $D_4^{ij\pm}$ points correspond to the creation or annihilation of two $D_4^{\pm ij}$ curves of the same signature. By contrast, the D_5^{ij} points correspond to the creation or annihilation of a pair with one D_4^{+ij} and one D_4^{-ij} point.

10.8 Discussion & conclusions

In this study we have developed a general formalism for identifying the caustic structure of a dynamically evolving mass distribution, in an arbitrary dimensional space. Through a new and direct derivation of the caustic conditions for the classification and characterization of singularities that will form in an evolving matter field, our study enables the practical implementation of a toolset for identifying the spatial location and outline of each relevant class of emerging singularities. By enabling the development of such instruments, and the application of these to any cosmological primordial density and velocity field, our study opens the path towards further insight into the dynamics of the formation and evolution of the morphological features populating the cosmic-web. In particular significant is that it will enable us to obtain a fundamental understanding of the spatial organization of the cosmic-web, *i.e.*, of the way in which these structural components are arranged and connected.

10.8.1 Phase-space structure of the cosmic-web

Caustics are prominent features emerging in advanced stages of dynamically evolving fluids. They mark the positions where fluid elements cross and multi-stream regions form. They are associated with regions of infinite density, and often go along with the formation of shocks. In the context of the gravitationally evolving mass distribution in the universe, caustics emerge in regions in which nonlinear gravitational collapse starts to take place. As such, they are a typical manifestation of the structure formation process at the stage where it transits from the initial linear evolution to that of more advanced nonlinear configurations involving gravitational contraction and collapse. The overall spatial organization of matter at the corresponding scale is that of the cosmic-web, which assembles flattened walls, elongated filaments and tendrils and dense, compact cluster nodes in an intricate multiscale weblike network that pervades the Universe.

Over the past decades our understanding of the formation and evolution of the cosmic-web has advanced considerably. The availability of large computer simulations have been instrumental in this, as they enabled us to follow the cosmic structure formation process in detail (see *e.g.* [302, 331, 290]). In combination with new theoretical insights [49, 319], this has led to the development of a general picture of the emergence of the weblike matter and galaxy distribution. The full phase-space dynamics of the process and its manifestation in the emerging matter distribution is an instrumental aspect of this that only recently received more prominent attention. While the study by Zel'dovich [344] already underlined the importance of a full phase-space description for understanding cosmic structure formation (see also [298, 297]), with the exception of a few prominent studies [20] the wealthy information content of full 6-D phase-space escaped attention.

A series of recent publications initiated a resurgence of interest in the phase-space aspects of the cosmic structure

formation process. They realized that the morphology of components in the evolving matter distribution is closely related to its multistream character [2, 134, 262, 296, 281] (for an early study on this observation see [64]). This realization is based on the recognition that the emergence of nonlinear structures occurs at locations where different streams of the corresponding flow field cross each other. Looking at the appearance of the evolving spatial mass distribution as a 3D *phase space sheet* folding itself in 6D phase space, this establishes a connection between the structure formation process and the morphological classification of the emerging structure. Moreover, to further our understanding of the dynamical evolution and buildup of the cosmic matter distribution, we also need to answer the question in how far the various emerging structural features connect up in the overall weblike network of the cosmic-web.

10.8.2 Singularities and caustics

To be able to answer the questions, we study the emergence of singularities and caustics in a dynamically evolving mass distribution. Our analysis is built on the seminal work by Arnol'd, specifically his classification of singularities in Lagrangian catastrophe theory. In a three-dimensional setting we can recognize two series of singularities, the A_k and D_k series. The 4 classes of A_k singularities – A_2, A_3, A_4 and A_5 – are the singularities for which the caustic condition holds for one eigenvalue. The D-family of umbilic singularities – including the D_4^+, D_4^- and D_5 – are caustics for which the caustic conditions are satisfied by two eigenvalue simultaneously. In three-dimensional fluids, the case in which all three eigenvalues simultaneously satisify the caustic conditions, the E-family caustics, is non-degenerate.

In order to detect these caustics in practice, we derived the *caustic conditions*, which classify them in terms of both *eigenvalue* and the *eigenvector* fields of the deformation tensor. The derivation differs from the classical derivation of catastrophe theory, in terms of generating functions and the classification of its degenerate critical points, in that we work with the geometry of the system. Moreover, the caustic conditions are not restricted to Hamiltonian dynamics and apply to all systems which allow for a description with a sufficiently differentiable Lagrangian map.

10.8.3 Caustic skeleton and cosmic-web

On the basis of the derived formalism, we show how the caustics of a Lagrangian fluid form an intricate skeleton of the nonlinear evolution of the fluid. The family of newly derived caustic conditions allow a significant extension and elaboration of the work described in Arnold et al. (1982) [20]. Arnol'd et al. classified the caustics that develop in one- and two-dimensional systems that evolve according to the Zel'dovich approximation. While [17] did offer a qualitative description of caustics in the three-dimensional situation, this did not materialize in a practical application to the full three-dimensional cosmological setting. The expressions derived in our study, and the specific identification of the important role of the deformation tensor eigenvectors, have enabled us to breach this hiatus. To identify the full spatial distribution and arrangement of caustics in the evolving three-dimensional cosmic matter distribution, we follow the philosophy exposed in the two-dimensional study by Hidding et al. 2014 [203, 136]. By relating the singularity distribution to the spatial properties of the initial Gaussian deformation field, [203] managed to identify and show the spatial connectivity of singularities and establish how in a hierarchical evolutionary sequence they evolve and may ultimately merge with surrounding structures.

When applied to the Zel'dovich approximation for cosmic structure formation, the caustic conditions form a skeleton of the caustic web. In the context of the cosmic-web, we may identify these singularities with different components. This observation by itself leads to some radically new insights into the origin of the structural features in the cosmic-web. The A_3 cusp singularities are related to the *walls* of the skeleton of the comsic web. The A_4 swallowtail singularities trace the filamentary ridges and tendrils in the cosmic-web. Also the $D_4^{\pm}$ hyperbolic and elliptic umbilic singularities are related to the filamentary spine of the spine, as they define the dense filamentary extensions of the cluster nodes. The butterfly (A_5) and parabolic umbilic (D_5) singularities are both connected with the nodes of the weblike pattern. One immediate observation of considerable interest is that the line-like A_4 and $D_4^{\pm}$ features trace the connectivity of the cosmic-web in meticulous detail.

Perhaps equally or even more interesting, and of key importance for our understanding of the dynamical evolution of the cosmic-web, is the observation that both filaments and tendrils, as well as nodes, may have formed due to the folding by the phase-space sheet induced by only one deformation eigenvalue: the filamentary A_4 caustics and nodal A_5 caustic belong to the one eigenvalue A family of caustics. In other words, collapse along the second eigenvector is not necessary to create a filament-like structure, and not even collapse along both second and third eigenvector is needed for the appearance of nodes (see [203, 204]). This is a new insight as it suggests the existence of different possible late-time morphologies for filaments and nodes [204].

A realization of key importance emanating from our work is that it is not sufficient to limit a structural analysis to the eigenvalues of the deformation tensor field. Usually neglected, we argue – and show by a few examples – that it is necessary to include the information contained in the (local) deformation tensor eigenvectors, our study has demonstrated and emphasized that for the identification of the full spatial outline of the cosmic-web's skeleton. In an accompanying numerical study of the caustic skeleton in cosmological N-body simulations, we illustrate how essential it is to invoke the deformation eigenvectors in the analysis [145]. This study will present a numerical and statistical comparison between the matter distribution in the simulation and the caustic skeleton of the three-dimensional cosmic-web.

10.8.4 Extensions and applications

Amongst the potentially most important applications of the current project is the fact that the caustic skeleton inferred from the Zel'dovich approximation adheres closely to the spine of the full nonlinear matter distribution. The direct implication is that we may directly link the outline of the cosmic-web to the initial Gaussian density and velocity field. On the basis of the corresponding deformation field, one may then attempt to calculate a range of properties analytically. The fact that we may invoke Gaussian statistics facilitates the calculation of a wide range of geometric and topological characteristics of the cosmic-web, as they are directly related to the primordial Gaussian deformation field, its eigenvalues and eigenvectors. The first step towards this program were taken by [138]. A few examples of results of such a statistical treatment for 2-dimensional fluids are described in [136]. It describes how one may not only analytically compute the distribution of maxima, or minima, but also the population of singularities and the length of caustic lines. This will represent a major extension of statistical descriptions that were solely based on the eigenvalue fields (see *e.g.* [116, 276]). Moreover, the ability to infer solid analytical results

for a range of parameters quantifying the cosmic-web will be a key towards identifying properties of the cosmic-web that are sensitive to the underlying cosmology. This, in turn, would enable the use of these properties to infer cosmological parameters, investigate the nature of dark matter and dark energy, trace the effects of deviations from standard gravity, and other issues of general cosmological interest.

Notwithstanding the observation that the caustic skeleton inferred from the Zel'dovich approximation appears to closely adhere to the full nonlinear structure seen in N-body simulations, an aspect that still needs to be addressed in detail is the influence of the dynamical evolution on the the developing caustic structure. This concerns in particular the description of the dynamics of the system. Given the nature of singularities, the process of caustic formation might be very sensitive to minor deviations of the mass element deformations and hence the modelling of the dynamics. This may even strongly affect the predicted population of caustics and their spatial organization in the skeleton of the cosmic-web. The Zel'dovich formalism [344] is a first-order Lagrangian approximation. A range of studies have shown that second order Lagrangian descriptions, often named 2LPT, provide a considerably more accurate approximation of in particular the mildly nonlinear phases that are critical for understanding the cosmic-web [61, 64, 63, 52, 294]. In addition to a follow-up study in which we explore the caustic structure according to 2LPT and possible systematic differences with that predicated by the Zel'dovich approximation, we will also systematically investigate the caustic skeleton in the context of the adhesion formalism [175, 176, 205, 202]. Representing a fully nonlinear extension of the Zel'dovich formalism through the inclusion of an effective gravitational interaction term for the emerging structures, it is capable of following the hierarchical buildup of structure. While it provides a highly insightful model for the hierarchically evolving cosmic-web, it also affects the flow patterns and hence the multistream structure in the cosmic mass distribution. In how far this will affect the caustic skeleton remains a major question for our work.

Finally, of immediate practical interest to our project will be identification of the various classes of singularities that are populating the Local Universe. On the basis of advanced Bayesian reconstruction techniques, various groups have been able to infer constrained realizations of the implied Gaussian primordial density and velocity field in a given cosmic volume [215, 221, 235, 234]. From these constrained initial density and deformation fields, we may subsequently determine the caustic structure in the Local Universe (see *e.g.* [204]). The resulting caustic skeleton of the local cosmic-web may then be confronted with the structures – clusters, groups and galaxies – that surveys have observed. Ultimately, this will enable us to reconstruct the cosmic history of objects and structures in the local Universe.

10.8.5 Summary

In summary, the ability to relate the formation and hierarchical evolution of structure in the Universe to the tale of the emergence and fate of singularities in the cosmic density field provides the basis for a dynamical theory for the development of the cosmic-web, including that of its substructure. This will be the principal question and subject of the sequel to the work that we have presented here.

10.A Zel'dovich approximation

The Zel'dovich approximation is the first order approximation of a Lagrangian pressureless fluid evolving under self gravity, [344]. The Zel'dovich approximation is the simplest example of a Lagrangian fluid with Hamiltonian dynamics and serves as a good illustration of the caustic conditions. The displacement map of the Zel'dovich approximation factors into a term depending on time and a term depending on the initial conditions

$$s_t(q) = -b_+(t)\nabla_q \Psi(q), \tag{10.74}$$

with the linearized velocity potential $\Psi(q)$ and growing mode $b_+(t)$. The growing mode can be obtained from linear Eulerian perturbation theory. Up to linear order, the linearized velocity potential is proportional to the linearly extrapolated gravitational potential at the current epoch $\phi_0(q)$, *i.e.*,

$$\Psi(q) = \frac{2}{3\Omega_0 H_0^2}\phi_0(q), \tag{10.75}$$

with current Hubble constant H_0 and current energy density Ω_0. The linearized velocity potential $\Psi(q)$ encodes the initial conditions while the growing mode $b_+(t)$ encodes the cosmological evolution of the fluid. For the Zel'dovich approximation it is common to define the deformation tensor as

$$\psi_{ij} = \frac{\partial^2 \Psi(q)}{\partial q_i \partial q_j} \tag{10.76}$$

with eigenvalues $\lambda_i(q)$ satisfying $\mu_i(q,t) = -b_+(t)\lambda_i(q)$. The density in the Zel'dovich approximation can be expressed as

$$\rho(x',t) \;=\; \sum_{q\in A(x',t)} \frac{\rho_i(q)}{|1-b_+(t)\lambda_1(q)||1-b_+(t)\lambda_2(q)||1-b_+(t)\lambda_d(q)|}, \tag{10.77}$$

with ρ_i the initial density field. Caustics occur at q at time t if and only if

$$\lambda_i(q) = \frac{1}{b_+(t)} \tag{10.78}$$

for at least one i. The eigenvalues λ_i are functions determined by the initial gravitational field. Equation (10.78) can be pictured as a hyperplane at height $1/b_+(t)$. Since the Zel'dovich approximation concerns potential flow, which means that the eigenvalues are real and can be ordered such that $\lambda_1 \geq \lambda_2 \geq \lambda_3$. The intersection of this plane with the graph of the eigenvalues undergoes shell-crossing at that time. For the Zel'dovich approximation the caustic conditions in terms of the eigenvalues λ_i are given by

$$\begin{aligned}
A_1 &= \{q \in L | \lambda_i(q) \neq 1/b_+(t) \text{ for all } t \text{ and } i\}, && (10.79)\\
A_2^i(t) &= \{q \in L | \lambda_i(q) = 1/b_+(t)\}, && (10.80)\\
A_3^i(t) &= \{q \in L | q \in A_2^i(t) \text{ and } \lambda_{i,i}(q) = 0\}, && (10.81)\\
A_4^i(t) &= \{q \in L | q \in A_3^i(t) \text{ and } \lambda_{i,ii}(q) = 0\}, && (10.82)\\
A_5^i(t) &= \{q \in L | q \in A_4^i(t) \text{ and } \lambda_{i,iii}(q) = 0\}, && (10.83)\\
D_4^{\pm ij}(t) &= \{q \in L | \lambda_i(q) = \lambda_j(q) = 1/b_+(t) \text{ and } \mathrm{sign}(S_{\mathcal{M}}) = \pm 1\}, && (10.84)\\
D_5^{ij}(t) &= \{q \in L | q \in D_4^{ij}(t) \text{ and } (\lambda_i - \lambda_j)_{,i}(q) = (\lambda_i - \lambda_j)_{,j}(q) = 0\}, && (10.85)
\end{aligned}$$

and the points at which the topology of above sets changes

$$\begin{aligned}
A_3^{i+} &= \{q \in L | q \in A_2^i \wedge \lambda_i(q) \text{ max-/minimum of } \lambda_i\}, && (10.86)\\
A_3^{i-} &= \{q \in L | q \in A_2^i \text{ saddle point of } \lambda_i\}, && (10.87)\\
A_4^{i+} &= \{q \in L | q \in A_3^i \wedge \lambda_{i,ii}(q) \text{ max-/minimum of } \lambda_{i,ii}|_{A_2}\}, && (10.88)\\
A_4^{i-} &= \{q \in L | q \in A_3^i \text{ saddle point of } \lambda_{i,ii}|_{A_2}\}, && (10.89)\\
D_4^{ij\pm} &= \{q \in L | q \in D_4^{\pm ij} \wedge \lambda_i(q) = \lambda_j(q) \text{ max-/minimum of } \lambda_i|_{D_4^{\pm ij}} = \lambda_j|_{D_4^{\pm ij}}\}. && (10.90)
\end{aligned}$$

with the direction derivatives $\lambda_{i,i} = \nabla\lambda_i \cdot v_i$, $\lambda_{i,ii} = \nabla\lambda_{i,i} \cdot v_i$ and $\lambda_{i,iii} = \nabla\lambda_{i,ii} \cdot v_i$. Note that the eigenvectors are defined modulo multiplication by a real number and really represent lines.

10.B Non-diagonalizable deformation tensors

In sections 10.3 and 10.4 we derived the shell-crossing conditions and caustic conditions under the assumption that the deformation tensor $\frac{\partial s_t}{\partial q} = \mathcal{M}$ is diagonalizable. We here extend these conditions to non-diagonalizable deformation tensors.

The eigenvalues μ_i are the roots of the characteristic function $\chi(\lambda) = \det(\mathcal{M} - \lambda I)$ corresponding to the deformation tensor. The deformation tensor is diagonalizable if and only if the algebraic multiplicity – the order of the root – is equal to the geometric multiplicity – the number of eigenvectors corresponding to the root – for all eigenvalues. Hence, in order for the deformation tensor to be non-diagonalizable, two or more eigenvalues need to coincide while there are fewer corresponding eigenvectors. In this case we can extend the set of eigenvectors by adding the necessary generalized eigenvectors to put the deformation tensor in Jordan normal form

$$\mathcal{M} = \mathcal{V}\mathcal{M}_J\mathcal{V}^{-1} \tag{10.91}$$

where $\mathcal{V}$ is the generalized modal matrix consisting of the eigenvectors and generalized eigenvectors and $\mathcal{M}_J$ is the upper triangular matrix of Jordan normal form containing the eigenvalues.

In the three-dimensional case, these non-diagonalisable deformation tensors correspond to the hyperbolic/elliptic umbilic ($D_4^{\pm}$) caustics. For simplicity lets restrict to the three-dimensional case where the shell-crossing occurs due to the first and the second eigenvalue fields $1+\mu_1 = 1+\mu_2 = 0$. In this case, we extend the set of eigenvectors $\{v_1, v_3\}$ by adding the generalized eigenvector $\bar{v}_2$. The Jordan matrix now takes the form

$$\mathcal{M}_J = \begin{pmatrix} \mu_1 & 1 & 0 \\ 0 & \mu_2 & 0 \\ 0 & 0 & \mu_3 \end{pmatrix} . \tag{10.92}$$

Substituting equation (10.91) in equation (10.20) we obtain the condition that there needs to exist a non-zero tangent vector of $D_4^{\pm 12}$ for which

$$(1+\mu_1)v_1^* \cdot T + \bar{v}_2^* \cdot T = 0 \tag{10.93}$$
$$(1+\mu_2)\bar{v}_2^* \cdot T = 0 \tag{10.94}$$
$$(1+\mu_3)v_3^* \cdot T = 0\,, \tag{10.95}$$

replacing equations (10.27). We thus see that the D_4^{12} variety forms a D_5^{12} caustic if and only if $\bar{v}_2^* \cdot T = 0$ and $v_3^* \cdot T = 0$ (for a diagonalizable deformation tensor we obtain the second condition). This is equivalent to the condition that T is parallel to the eigenvector v_1.

Finally note that the deformation tensor can only be non-diagonalizable for non-Hamiltonian dynamics for which the parabolic umbilic caustic (D_5) is not stable (see section 10.5.2). This condition is thus not relevant for Hamiltonian and generic non-Hamiltonian Lagrangian fluids in three dimensions.

This analysis straightforwardly generalizes to the case in which the geometric and algebraic multiplicity of the eigenvalues differs by more than one for higher dimensional fluids.

10.C Shell-crossing conditions: coordinate transformation

The shell-crossing conditions are manifestly independent of coordinate choices. However, the eigenvalue and eigenvector fields generally do depend on the choice of coordinates. By themselves, they do therefore not provide valid descriptions of the dynamics of the fluid. Suppose the displacement field can be written as $s = \nabla\psi$ for some potential ψ. The Hessian H_x of ψ,

$$H_{ij} \;=\; \frac{\partial^2\psi}{\partial x_i \partial x_j}\,, \tag{10.96}$$

transforms non-trivially under the local coordinate transformation $x \to X(x)$, *i.e.*,

$$H \to \tilde{H} \;=\; J^T H J + J^T \nabla(J) \nabla\psi\,, \tag{10.97}$$

with J the Jacobian between the coordinate systems X and x,

$$J_{ij} \;=\; \frac{\partial X_i}{\partial x_j}\,. \tag{10.98}$$

From this we immediately infer that the eigenvalue field and eigenvector fields are invariant if the transformation is orthogonal and global, *i.e.*, if $J^T = J^{-1}$ and $\nabla(J) = 0$. As may be expected, these transformations include rotations and translations.

10.D Lagrangian maps and Lagrangian equivalence

We here shortly describe the mathematical background of symplectic manifolds, Lagrangian manifolds and Lagrangian maps. For a detailed description and derivations we refer to [13, 21].

10.D.1 Symplectic manifolds and Lagrangian maps

A $2n$-dimensional symplectic manifold (M, ω) is a smooth $2n$-dimensional manifold M, equipped with a closed nondegenerate bilinear 2-form ω called the symplectic form. Symplectic manifolds are always even dimensional for ω to be nondegenerate. In Hamiltonian dynamics the symplectic form ω can be associated to the Poisson brackets which encodes the dynamics of the theory. A Lagrangian manifold L of a $2n$-dimensional symplectic manifold (M, ω) is a n-dimensional submanifold of M on which the symplectic form ω vanishes. Let (B, π) be a Lagrangian fibration of (M, ω), which is a n-dimensional manifold with a projection map $\pi : M \to B$ for which the fibers $\pi^{-1}(b)$ are Lagrangian manifolds for all $b \in B$.

An example of a symplectic manifold is phase space consisting of position and canonical momenta $(q_1, \ldots, q_n, p_1, \ldots p_n)$ with the symplectic form $\omega = \sum_i^n \mathrm{d}q_i \wedge \mathrm{d}p_i$. An example of a Lagrangian fibration is $\{(q_1, \ldots, q_n), \pi\}$ with the projection map $\pi(q_1, \ldots, q_n, p_1, \ldots p_n) = (q_1, \ldots, q_n)$.

Give a symplectic manifold (M, ω) with a Lagrangian fibration (B, π) we can for every Lagrangian manifold L define a Lagrangian map $(\pi \circ i) : L \to M \to B$, with i being the inclusion map sending L into M. Two Lagrangian maps $(\pi_1 \circ i_1) : L_1 \to M_1 \to B_1$ and $(\pi_2 \circ i_2) : L_2 \to M_2 \to B_2$ are defined to be Lagrangian equivalent if there exist diffeomorphisms σ, τ and ν such that $\tau \circ i_1 = i_2 \circ \sigma, \nu \circ \pi_1 = \pi_2 \circ \tau$ and $\tau^* \omega_2 = \omega_1$, or equivalently the diagram below commutes

$$\begin{array}{ccccc} L_1 & \xrightarrow{i_1} & (M_1, \omega_1) & \xrightarrow{\pi_1} & B_1 \\ \sigma \downarrow & & \tau \downarrow & & \nu \downarrow \\ L_2 & \xrightarrow{i_2} & (M_2, \omega_2) & \xrightarrow{\pi_2} & B_2 \end{array}$$

10.D.2 Displacement as Lagrangian map

Given a Lagrangian submanifold $\mathcal{L}$ we can construct a corresponding Lagrangian map. First map the Lagrangian submanifold $\mathcal{L}$ with the inclusion map $i : \mathcal{L} \to \mathcal{C}$ to the corresponding points in phase space $\mathcal{C}$, *i.e.*, $i : (q, x) \mapsto (q, x)$ for all $(q, x) \in \mathcal{L}$. Subsequently map these points to a base manifold B with the projection map $\pi : \mathcal{C} \to B$. In Lagrangian fluid dynamics it is convenient to pick the Eulerian manifold E as the base manifold B and define the projection map as $\pi : (q, x) \mapsto x$ for all $(q, x) \in \mathcal{C}$. As there will always be an exact correspondence between the Lagrangian manifold L and the Lagrangian submanifold $\mathcal{L}_t \subset \mathcal{C}$ (there exists a unique point $x \in E$ such that $(q, x) \in \mathcal{L}_t$ for every $q \in L$), we can

associate the Lagrangian map corresponding to $\mathcal{L}_t$ with the map x_t. In summary, the map x_t corresponds uniquely to a Lagrangian map for fluids with Hamiltonian dynamics.

A Lagrangian map can develop regions in which multiple points in the Lagrangian manifold are mapped to the same point in the base space. The points at which the number of pre-images of the Lagrangian map changes are known as Lagrangian singularities. Lagrangian catastrophe theory classifies the stable singularities, stable with respect to small deformations of $\mathcal{L}$, up to Lagrangian equivalence. Lagrangian equivalence is a generalization of equivalence up to coordinate transformations. For a precise definition of Lagrangian equivalence we refer to appendix 10.D.

10.D.3 Lagrangian map germs

In catastrophe theory it is important to consider the Lagrangian map at a point. This is achieved by means of Lagrangian germs. Starting with a point $p \in M$ we can consider Lagrangian functions $F_i : U_i \to B$ for $i = 1, 2$ for small environments U_i of p which coincide on the intersection $U_1 \cap U_2$. The equivalence classes of such Lagrangian functions are Lagrangian germs. The Lagrange equivalence of Lagrangian maps straightforwardly extends to Lagrange equivalence of Lagrangian germs. These are the equivalence classes used in the classification of stable Lagrangian maps, where a Lagrangian germ is stable if and only if every sufficiently small fluctuation on the germ is Lagrange equivalent to the germ.

10.D.4 Gradient maps

Every Lagrangian germ is Lagrange equivalent to the germ of a gradient map. That is to say, for every Lagrangian map $l = \pi \circ i : \mathcal{L} \to \mathcal{C} \to E$ we can for a point $(q, x) \in \mathcal{L}$ locally write the map as

$$l(q_1, \ldots, q_n, x_1, \ldots, x_n) = \left(\frac{\partial S}{\partial q_1}, \frac{\partial S}{\partial q_2}, \ldots, \frac{\partial S}{\partial q_n} \right) \tag{10.99}$$

for some function $S : \mathbb{R}^n \to \mathbb{R}$. The corresponding map x is given by

$$x(q_1, \ldots, q_n, t) = \left(\frac{\partial S}{\partial q_1}, \frac{\partial S}{\partial q_2}, \ldots, \frac{\partial S}{\partial q_n} \right) \tag{10.100}$$

for some time t. By writing $S = \frac{1}{2}q^2 + \Psi$ for $\Psi : \mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}$ we obtain

$$x(q,t) = q + \frac{\partial \Psi}{\partial q}, \tag{10.101}$$

with the gradient field

$$s = \frac{\partial \Psi}{\partial q}. \tag{10.102}$$

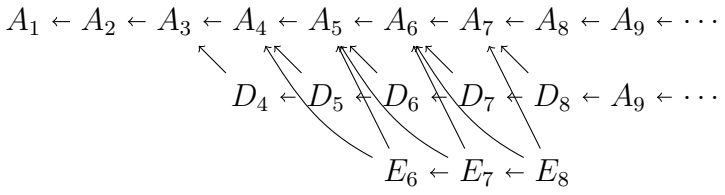
The Jacobian of the displacement map

$$\left[\frac{\partial s}{\partial q}\right]_{ij} = \frac{\partial^2 \Psi}{\partial q_i \partial q_j} \tag{10.103}$$

is symmetric. The set of eigenvectors $\{v_i\}$ can be taken to be orthonormal by which the dual vectors coincide with the eigenvectors, *i.e.*, $v_i^* = v_i$ for all i. A Lagrangian map is locally equivalent to the Zel'dovich approximation.

10.D.5 Arnol'd's classification of Lagrangian catastrophes

In section 10.4, we described the classification of Lagrangian singularities in up to three dimensions. However the classification extends to higher dimensional singularities. A $(n+1)$-dimensional fluid can contain stable singularities in the A_i, D_i and E_i classes with $i \leq n+2$, where the D-class range starts at $i = 4$ and the E-class is only defined for $i = 6, 7, 8$. These singularities decompose into lower-dimensional singularities as illustrated in the unfolding diagram below.



$A_1: \ x(q,1) = (q_1, q_2, q_3)$	$1+\mu_1 = 1$	$1+\mu_2 = 1$	$1+\mu_3 = 1$
$A_2: \ x(q,1) = (q_1, q_2, q_3^2)$	$1+\mu_1 = 1$	$1+\mu_2 = 1$	$1+\mu_3 = 2q_3$ $\mu_{3,3} = 2$
$A_3: \ x(q,1) = (q_1, q_2, q_1q_3 + q_3^3)$	$1+\mu_1 = 1$	$1+\mu_2 = 1$	$1+\mu_3 = 3q_3^2 + q_1$ $\mu_{3,3} = 6q_3$ $\mu_{3,333} = 6$
$A_4: \ x(q,1) = (q_1, q_2, q_1q_3 + q_3^4)$	$1+\mu_1 = 1$	$1+\mu_2 = 1$	$1+\mu_3 = q_1 + 4q_3^3$ $\mu_{3,3} = 12q_3^2$ $\mu_{3,33} = 24q_3$ $\mu_{3,333} = 24$
$A_5: \ x(q,1) = (q_1, q_2, q_1q_3 + q_2q_3^2 + q_3^5)$	$1+\mu_1 = 1$	$1+\mu_2 = 1$	$1+\mu_3 = q_1 + 2q_2q_3 + 5q_3^4$ $\mu_{3,3} = 2q_2 + 20q_3^3$ $\mu_{3,33} = 60q_3^2$ $\mu_{3,333} = 120q_3$ $\mu_{3,3333} = 120$
$A_3^{\pm}: \ x(q,1) = (q_1, q_2, (q_1^2 \pm q_2^2)q_3 + q_3^3)$	$1+\mu_1 = 1$	$1+\mu_2 = 1$	$1+\mu_3 = q_1^2 \pm q_2^2 + 3q_3^2$ $\mu_{3,3} = 6q_3$ $\mu_{3,33} = 6$
$A_4^{\pm}: \ x(q,1) = (q_1, q_2, q_1q_3 \pm q_2^2q_3^2 + q_3^4)$	$1+\mu_1 = 1$	$1+\mu_2 = 1$	$1+\mu_3 = q_1 \pm 2q_2^2q_3 + 4q_3^3$ $\mu_{3,3} = \pm 2q_2^2 + 12q_3^2$ $\mu_{3,33} = 24q_3$ $\mu_{3,333} = 24$

Table 10.2: The caustic conditions of the normal forms of the A singularity classes

10.E Caustic conditions of the normal forms

We here verify the caustic conditions for the normal forms in the generic classification of singularities given in section 10.5.2. The normal forms of the the Lagrangian singularities given in section 10.5.3 follow analogously.

The eigenvalue fields and corresponding derivatives in the direction of the eigenvector fields are given in Table 10.2. The eigenvalues of the normal form for the trivial (A_1) case equal 1 and thus satisfy the condition $1 + \mu_i \neq 0$ for all i. The third eigenvalue of the normal form of the fold (A_2) singularity equals -1 in the origin. The derivative of the eigenvalue field in the direction of the corresponding eigenvector field does not vanish in the origin. The normal form thus satisfies the caustic conditions of the fold singularity. The normal forms of the remaining singularities follow analogously.

References

[1] B. P. Abbott, R. Abbott, T. D. Abbott, M. R. Abernathy, F. Acernese, K. Ackley, C. Adams, T. Adams, P. Addesso, R. X. Adhikari, and et al. Observation of Gravitational Waves from a Binary Black Hole Merger. *Physical Review Letters*, 116(6):061102, February 2016.

[2] T. Abel, O. Hahn, and R. Kaehler. Tracing the dark matter sheet in phase space. *Mon. Not. R. Astron. Soc.*, 427:61–76, November 2012.

[3] M. Adams and V. Guillemin. *Measure Theory and Probability*. Birkhauser, Boston, 1996.

[4] R. J. Adler. *The Geometry of Random Fields*. Chichester: Wiley, 1981.

[5] Y. Aharonov and D. Rohrlich. *Quantum Paradoxes: Quantum Theory for the Perplexed*. Wiley-VCH, September 2003.

[6] Yakir Aharonov, David Z. Albert, and Lev Vaidman. How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100. *Physical Review Letters*, 60(14):1351–1354, Apr 1988.

[7] Yakir Aharonov, Peter G. Bergmann, and Joel L. Lebowitz. Time Symmetry in the Quantum Process of Measurement. *Physical Review*, 134(6B):1410–1416, Jun 1964.

[8] S. Albeverio, R. Høegh-Krohn, and S. Mazzucchi. *Mathematical Theory of Feynman Path Integrals: An Introduction*. Lecture Notes in Mathematics. Springer Berlin Heidelberg, 2008.

[9] Mandana Amiri et al. Observations of fast radio bursts at frequencies down to 400 megahertz. *Nature*, 566(7743):230–234, 2019.

[10] Carl D. Anderson. The Positive Electron. *Physical Review*, 43(6):491–494, Mar 1933.

[11] M. A. Aragón-Calvo, E. Platen, R. van de Weygaert, and A. S. Szalay. The Spine of the Cosmic Web. *Astrophysical Journal*, 723:364–382, November 2010.

[12] M. A. Aragón-Calvo, R. van de Weygaert, and B. J. T. Jones. Multiscale phenomenology of the cosmic web. *Mon. Not. R. Astron. Soc.*, 408:2163–2187, November 2010.

[13] E. Arnold, S.M. Gusein-Zade, and A.N. Varchenko. *Singularities of Differentiable Maps, Volume 2: Monodromy and Asymptotics of Integrals*. Modern Birkhäuser Classics. Birkhäuser Boston, 2012.

[14] V. I. Arnol'd. Normal forms for functions near degenerate critical points, the Weyl groups of A_k, D_k, E_k and Lagrangian singularities. *Functional Anal. Appl*, 6:1972, 1972.

[15] V. I. Arnol'd. Wave front evolution and equivalent Morse lemma. *Communications in Pure Applied Mathematics*, 29:557–582, November 1976.

[16] V. I. Arnol'd. *Mathematical methods of classical mechanics*. New York: Springer, 1978.

[17] V. I. Arnol'd. Evolution of singularities of potential flows in collisionless media and transformations of caustics in three-dimensional space. *Trudy Seminar imeni G Petrovskogo*, 8:21–57, 1982.

[18] V. I. Arnold. *Catastrophe theory*. Berlin: Springer, 1984.

[19] V. I. Arnol'd and B. A. Khesin. Topological methods in hydrodynamics. *Annual Review of Fluid Mechanics*, 24:145–166, 1992.

[20] V. I. Arnol'd, S. F. Shandarin, and I. B. Zel'dovich. The large scale structure of the universe. I - General properties One- and two-dimensional models. *Geophysical and Astrophysical Fluid Dynamics*, 20:111–130, 1982.

[21] V.I. Arnol'd, A. Varchenko, and S.M. Gusein-Zade. *Singularities of Differentiable Maps: Volume I: The Classification of Critical Points Caustics and Wave Fronts*. Monographs in Mathematics. Birkhauser Boston, 2012.

[22] Vladimir I. Arnol'd. Remarks on the Stationary Phase Method and Coxeter Numbers. *Russian Mathematical Surveys*, 28(5):19–48, Oct 1973.

[23] Vladimir I. Arnol'd. Critical Points of Smooth Functions and Their Normal Forms. *Russian Mathematical Surveys*, 30(5):1–75, Oct 1975.

[24] R. Arnowitt, S. Deser, and C. W. Misner. Dynamical Structure and Definition of Energy in General Relativity. *Physical Review*, 116:1322–1330, December 1959.

[25] R. F. Baierlein, D. H. Sharp, and J. A. Wheeler. Three-Dimensional Geometry as Carrier of Information about Time. *Physical Review*, 126:1864–1865, June 1962.

[26] J. M. Bardeen, J. R. Bond, N. Kaiser, and A. S. Szalay. The statistics of peaks of Gaussian random fields. *The Astrophysical Journal*, 304:15–61, May 1986.

[27] Itzhak Bars. Wavefunction for the Universe close to its beginning with dynamically and uniquely determined initial conditions. *Physical Review D*, 98(10):103510, Nov 2018.

[28] Itzhak Bars, Shih-Hung Chen, Paul J. Steinhardt, and Neil Turok. Complete Set of Homogeneous Isotropic Analytic Solutions in Scalar-Tensor Cosmology with Radiation and Curvature. *Phys. Rev.*, D86:083542, 2012.

[29] Gökçe Basar, Gerald V. Dunne, and Mithat Ünsal. Resurgence theory, ghost-instantons, and analytic continuation of path integrals. *Journal of High Energy Physics*, 2013:41, Oct 2013.

[30] I. A. Batalin and G. A. Vilkovisky. Relativistic S-matrix of dynamical systems with boson and fermion constraints. *Physics Letters B*, 69:309–312, August 1977.

[31] Lorenzo Battarra and Jean-Luc Lehners. On the No-Boundary Proposal for Ekpyrotic and Cyclic Cosmologies. *JCAP*, 1412(12):023, 2014.

[32] Lorenzo Battarra and Jean-Luc Lehners. On the Creation of the Universe via Ekpyrotic Instantons. *Phys. Lett.*, B742:167–171, 2015.

[33] H. Bauer and R.B. Burckel. *Measure and Integration Theory*. De Gruyter studies in mathematics. W. de Gruyter, 2001.

[34] C. Becchi, A. Rouet, and R. Stora. The abelian Higgs Kibble model, unitarity of the S-operator. *Physics Letters B*, 52:344–346, October 1974.

[35] Alireza Behtash, Gerald V. Dunne, Thomas Schaefer, Tin Sulejmanpasic, and Mithat Unsal. Toward Picard-Lefschetz Theory of Path Integrals, Complex Saddles and Resurgence. *arXiv e-prints*, page arXiv:1510.03435, Oct 2015.

[36] Alireza Behtash, Gerald V. Dunne, Thomas Schafer, Tin Sulejmanpasic, and Mithat Unsal. Complexified path integrals, exact saddles and supersymmetry. *Phys. Rev. Lett.*, 116(1):011601, 2016.

[37] Alireza Behtash, Gerald V. Dunne, Thomas Schäfer, Tin Sulejmanpasic, and Mithat Ünsal. Complexified Path Integrals, Exact Saddles, and Supersymmetry. *Physical Review Letters*, 116(1):011601, Jan 2016.

[38] Alireza Behtash, Gerald V. Dunne, Thomas Schäfer, Tin Sulejmanpasic, and Mithat Ünsal. Critical points at infinity, non-Gaussian saddles, and bions. *Journal of High Energy Physics*, 2018(6):68, Jun 2018.

[39] M. V. Berry. Focusing and twinkling: critical exponents from catastrophes in non-Gaussian random short waves. *Journal of Physics A Mathematical General*, 10(12):2061–2081, Dec 1977.

[40] M. V. Berry. Focused tsunami waves. *Proceedings of the Royal Society of London Series A*, 463(2087):3055–3071, Nov 2007.

[41] M. V. Berry and J. F. Nye. Fine structure in caustic junctions. *nature*, 267(5606):34–36, May 1977.

[42] M. V. Berry, J. F. Nye, and F. J. Wright. The Elliptic Umbilic Diffraction Catastrophe. *Philosophical Transactions of the Royal Society of London Series A*, 291(1382):453–484, Apr 1979.

[43] M. V. Berry and C. Upstill. IV Catastrophe Optics: Morphologies of Caustics and Their Diffraction Patterns. *Progess in Optics*, 18:257–346, Jan 1980.

[44] L. Bianchi. *Sugli spazii a tre dimensioni che ammettono un gruppo continuo di movimenti*, volume 11, page 267. Soc. Ital. Sci. Mem. di Mat., 1898.

[45] A. V. Bilous, S. M. Ransom, and P. Demorest. Unusually Bright Single Pulses from the Binary Pulsar B1744-24A: A Case of Strong Lensing? *Astrophysical Journal*, 877(2):125, Jun 2019.

[46] N. D. Birrell and P. C. W. Davies. *Quantum fields in curved space*. Cambridge: University Press, 1982.

[47] M. R. Blanton, M. A. Bershady, B. Abolfathi, F. D. Albareti, C. Allende Prieto, A. Almeida, J. Alonso-García, F. Anders, S. F. Anderson, B. Andrews, and et al.

Sloan Digital Sky Survey IV: Mapping the Milky Way, Nearby Galaxies, and the Distant Universe. *Astrophysical Journal*, 154:28, July 2017.

[48] Martin Bojowald and Suddhasattwa Brahma. Loops Rescue the No-Boundary Proposal. *Physical Review Letters*, 121(20):201301, Nov 2018.

[49] J. R. Bond, L. Kofman, and D. Pogosyan. How filaments of galaxies are woven into the cosmic web. *Nature*, 380:603–606, April 1996.

[50] H. Bondi and T. Gold. The Steady-State Theory of the Expanding Universe. *Notices of the Royal Astronomical Society*, 108:252, 1948.

[51] Max Born and Emil Wolf. *Principles of Optics*. Cambridge university press, 1999.

[52] F. R. Bouchet, S. Colombi, E. Hivon, and R. Juszkiewicz. Perturbative Lagrangian approach to gravitational instability. *Astron. Astrophys.*, 296:575, April 1995.

[53] M. Boylan-Kolchin, V. Springel, S. D. M. White, A. Jenkins, and G. Lemson. Resolving cosmic structure formation with the Millennium-II Simulation. *Mon. Not. R. Astron. Soc.*, 398:1150–1164, September 2009.

[54] Suddhasattwa Brahma and Dong-han Yeom. No-boundary wave function for loop quantum cosmology. *Physical Review D*, 98(8):083537, Oct 2018.

[55] Sebastian F. Bramberger, Shane Farnsworth, and Jean-Luc Lehners. Wavefunction of anisotropic inflationary universes with no-boundary conditions. *Phys. Rev.*, D95(8):083513, 2017.

[56] Sebastian F. Bramberger, Thomas Hertog, Jean-Luc Lehners, and Yannick Vreys. Quantum transitions through cosmological singularities. *Journal of Cosmology and Astroparticle Physics*, 2017(7):007, Jul 2017.

[57] G. Brooker. *Modern Classical Optics*. Oxford Master Series in Physics. OUP Oxford, 2003.

[58] R. Brout, F. Englert, and E. Gunzig. The Creation of the Universe as a Quantum Phenomenon. *Annals Phys.*, 115:78, 1978.

[59] R. Brout, S. Massar, R. Parentani, S. Popescu, and Ph. Spindel. Quantum source of the back reaction on a classical field. *Physical Review D*, 52(2):1119–1133, Jul 1995.

[60] J.W. Bruce. A classification of 1-parameter families of map germs $r^3, 0 \to r^3, 0$ with applications to condensation problems. *J. Lond. Math. Soc.*, II. Ser. 30(2):375–384, 1985.

[61] T. Buchert. Lagrangian theory of gravitational instability of Friedman-Lemaitre cosmologies and the 'Zel'dovich approximation'. *Mon. Not. R. Astron. Soc.*, 254:729–737, February 1992.

[62] T. Buchert. Lagrangian perturbation theory - A key-model for large-scale structure. *Astron. Astrophys.*, 267:L51–L54, January 1993.

[63] T. Buchert. Lagrangian Theory of Gravitational Instability of Friedman-Lemaitre Cosmologies - a Generic Third-Order Model for Nonlinear Clustering. *Mon. Not. R. Astron. Soc.*, 267:811, April 1994.

[64] T. Buchert and J. Ehlers. Lagrangian theory of gravitational instability of Friedman-Lemaitre cosmologies – second-order approach: an improved model for non-linear clustering. *Mon. Not. R. Astron. Soc.*, 264, September 1993.

[65] T. Buchert, A. L. Melott, and A. G. Weiss. Testing higher-order Lagrangian perturbation theory against numerical simulations I. Pancake models. *Astron. Astrophys.*, 288:349–364, August 1994.

[66] T. S. Bunch and P. C. W. Davies. Quantum field theory in de Sitter space - Renormalization by point-splitting. *Proceedings of the Royal Society of London Series A*, 360:117–134, March 1978.

[67] J. Butterfield. On Time in Quantum Physics. *arXiv e-prints*, June 2014.

[68] J. Butterfield and C. J. Isham. Spacetime and the Philosophical Challenge of Quantum Gravity. *arXiv General Relativity and Quantum Cosmology e-prints*, 3 1999.

[69] R. H. Cameron. A Family of Integrals Serving to Connect the Wiener and Feynman Integrals. *Journal of Math. and Phys.*, 39:126–140, 1960.

[70] M. Capinski and P.E. Kopp. *Measure, Integral and Probability*. Springer Undergraduate Mathematics Series. Springer London, 2013.

[71] M. Cautun, R. van de Weygaert, B. J. T. Jones, and C. S. Frenk. Evolution of the cosmic web. *Mon. Not. R. Astron. Soc.*, 441:2923–2973, July 2014.

[72] S. Chatterjee, C. J. Law, R. S. Wharton, S. Burke-Spolaor, J. W. T. Hessels, G. C. Bower, J. M. Cordes, S. P. Tendulkar, C. G. Bassa, P. Demorest, B. J. Butler, A. Seymour, P. Scholz, M. W. Abruzzo, S. Bogdanov, V. M. Kaspi, A. Keimpema, T. J. W. Lazio, B. Marcote, M. A. McLaughlin, Z. Paragi, S. M. Ransom, M. Rupen, L. G. Spitler, and H. J. van Langevelde. A direct localization of a fast radio burst and its host. *Nature*, 541(7635):58–61, Jan 2017.

[73] Shami Chatterjee, Robert Wharton, Casey J. Law, Jason Hessels, Sarah Burke-Spolaor, Geoffrey C. Bower, Matthew W. Abruzzo, Cees Bassa, Bryan J. Butler, James M. Cordes, Demorest Paul, Victoria M. Kaspi, Maura McLaughlin, Scott M. Ransom, Paul Scholz, Andrew Seymour, Laura Spitler, Shriharsh P. Tendulkar, PALFA Survey Team, VLA+AO FRB121102 Simultaneous Campaign Team, and EVN FRB121102 Campaign Team. Localizing the Fast Radio Burst 121102. In *American Astronomical Society Meeting Abstracts #229*, volume 229 of *American Astronomical Society Meeting Abstracts*, page 330.01, Jan 2017.

[74] Aleksey Cherman, Daniele Dorigoni, and Mithat Unsal. Decoding perturbation theory using resurgence: Stokes phenomena, new saddle points and Lefschetz thimbles. *JHEP*, 10:056, 2015.

[75] CHIME/FRB Collaboration. A second source of repeating fast radio bursts. *Nature*, 566(7743):235–238, Jan 2019.

[76] M. S. Chong, A. E. Perry, and B. J. Cantwell. A general classification of three-dimensional flow fields. *Physics of Fluids*, 2:765–777, May 1990.

[77] J. M. Colberg, K. S. Krughoff, and A. J. Connolly. Intercluster filaments in a ΛCDM Universe. *Mon. Not. R. Astron. Soc.*, 359:272–282, May 2005.

[78] S. Cole, W. J. Percival, J. A. Peacock, P. Norberg, C. M. Baugh, C. S. Frenk, I. Baldry, J. Bland-Hawthorn, T. Bridges, R. Cannon, M. Colless, C. Collins, W. Couch, N. J. G. Cross, G. Dalton, V. R. Eke, R. De Propris, S. P. Driver, G. Efstathiou, R. S. Ellis, K. Glazebrook, C. Jackson, A. Jenkins, O. Lahav, I. Lewis, S. Lumsden, S. Maddox, D. Madgwick, B. A. Peterson, W. Sutherland, and K. Taylor. The 2dF Galaxy Redshift Survey: power-spectrum analysis of the final data set and cosmological implications. *Mon. Not. R. Astron. Soc.*, 362:505–534, September 2005.

[79] W. A. Coles. Interplanetary Scintillation. *Space Science Reviews*, 21(4):411–425, Feb 1978.

[80] W. A. Coles, B. J. Rickett, J. J. Gao, G. Hobbs, and J. P. W. Verbiest. Scattering of Pulsar Radio Emission by the Interstellar Plasma. *The Astrophysical Journal*, 717(2):1206–1221, Jul 2010.

[81] Wm. A. Coles, J. P. Filice, R. G. Frehlich, and M. Yadlowsky. Simulation of wave propagation in three-dimensional random media. *Applied Optics*, 34(12):2089, Apr 1995.

[82] N. Copernicus. *Nicolai Copernici Torinensis de revolvtionibvs orbium coelestium, libri VI. Habes in HOC opere iam recens nato, aedito, studiose lector, motus stellarum, tam fixarum, quam erraticarum, cum EX ueteribus tum etiam EX recentibus obseruationibus restitutos: nouis insuper AC admirabilibus hypothesibus ornatos. Habes etiam tabulas expeditissimas, EX quibus eosdem AD quoduis tempus quam facillime caculare poteris. Igitur eme, lege, fruere. [Line in Greek].* Norimbergae, apud Ioh. Petreium, 1543.

[83] J. M. Cordes, I. Wasserman, J. W. T. Hessels, T. J. W. Lazio, S. Chatterjee, and R. S. Wharton. Lensing of Fast Radio Bursts by Plasma Structures in Host Galaxies. *Astrophysical Journal*, 842(1):35, Jun 2017.

[84] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga. Limits on non-Gaussianities from WMAP data. *Journal of Cosmology and Astroparticle Physics*, 5:004, May 2006.

[85] M. Cristoforetti, F. Di Renzo, A. Mukherjee, and L. Scorzato. Quantum field theories on the Lefschetz thimble. *PoS*, LATTICE2013:197, 2014.

[86] Liang Dai and Wenbin Lu. Probing Motion of Fast Radio Burst Sources by Timing Strongly Lensed Repeaters. *Astrophysical Journal*, 847(1):19, Sep 2017.

[87] Liang Dai and Tejaswi Venumadhav. On the waveforms of gravitationally lensed gravitational waves. *arXiv e-prints*, page arXiv:1702.04724, Feb 2017.

[88] Ingrid Daubechies and John R. Klauder. Quantum-mechanical path integrals with Wiener measure for all polynomial Hamiltonians. II. *Journal of Mathematical Physics*, 26(9):2239–2256, Sep 1985.

[89] Hans De Raedt. Product formula algorithms for solving the time dependent Schrödinger equation. *Computer Physics Reports*, 7(1):1–72, Dec 1987.

[90] Hans de Raedt. *Computer Simulation of Quantum Phenomena in Nanoscale Devices*, pages 107–146. World Scientific Publishing Co. Pte. Ltd., 1996.

[91] W. de Sitter. On the relativity of inertia. Remarks concerning Einstein's latest hypothesis. *Koninklijke Nederlandse Akademie van Wetenschappen Proceedings Series B Physical Sciences*, 19:1217–1225, March 1917.

[92] W. de Sitter. Contributions to a British Association Discussion on the Evolution of the Universe. *Nature*, 128:706–709, October 1931.

[93] B. de Spinoza and W.H. White. *Ethics*. Classics of World Literature Series. Wordsworth Editions, 2001.

[94] T. Delmarcelle. *The Visualization of Second-Order Tensor Fields*. PhD thesis, Stanford university, 1995.

[95] R. Descartes. *Le Monde: Ou Traité de la Lumière*. Janus Series. Abaris Books, 1979.

[96] B. S. Dewitt. Quantum Theory of Gravity. I. The Canonical Theory. *Physical Review*, 160:1113–1148, August 1967.

[97] B. S. Dewitt. Quantum Theory of Gravity. II. The Manifestly Covariant Theory. *Physical Review*, 162:1195–1239, October 1967.

[98] B. S. Dewitt. Quantum Theory of Gravity. III. Applications of the Covariant Theory. *Physical Review*, 162:1239–1256, October 1967.

[99] B. S. Dewitt. The Quantum and Gravity: The Wheeler-DeWitt Equation. *Proceedings of the Eighth Marcel Grossmann Conference, The Hebrew University, Jerusalem, Israel.*, 1997.

[100] C. Dewitt-Morette. *The Pursuit of Quantum Gravity*. Springer-Verlag Berlin Heidelberg, 2011.

[101] Alice Di Tucci, Job Feldbrugge, Jean-Luc Lehners, and Neil Turok. Quantum Incompleteness of Inflation. *arXiv e-prints*, page arXiv:1906.09007, June 2019.

[102] Alice Di Tucci and Jean-Luc Lehners. Unstable no-boundary fluctuations from sums over regular metrics. *Physical Review D*, 98(10):103506, Nov 2018.

[103] Alice Di Tucci and Jean-Luc Lehners. The No-Boundary Proposal as a Path Integral with Robin Boundary Conditions. *arXiv e-prints*, page arXiv:1903.06757, Mar 2019.

[104] Juan Diaz Dorronsoro, Jonathan J. Halliwell, James B. Hartle, Thomas Hertog, and Oliver Janssen. Real no-boundary wave function in Lorentzian quantum cosmology. *Phys. Rev.*, D96(4):043505, 2017.

[105] R. H. Dicke, P. J. E. Peebles, P. G. Roll, and D. T. Wilkinson. Cosmic Black-Body Radiation. *Astrophysical Journal*, 142:414–419, July 1965.

[106] P. Dirac. The Relation between Mathematics and Physics. *Proceedings of the Royal Society (Edinburgh)*, 59:122–129, 1939.

[107] P. A. M. Dirac. The Quantum Theory of the Emission and Absorption of Radiation. *Proceedings of the Royal Society of London Series A*, 114(767):243–265, Mar 1927.

[108] P. A. M. Dirac. The Quantum Theory of the Electron. *Proceedings of the Royal Society of London Series A*, 117(778):610–624, Feb 1928.

[109] P. A. M. Dirac. The Quantum Theory of the Electron. Part II. *Proceedings of the Royal Society of London Series A*, 118(779):351–361, Mar 1928.

[110] P. A. M. Dirac. A Theory of Electrons and Protons. *Proceedings of the Royal Society of London Series A*, 126(801):360–365, Jan 1930.

[111] P. A. M. Dirac. On the Annihilation of Electrons and Protons. *Proceedings of the Cambridge Philosophical Society*, 26(3):361, Jan 1930.

[112] P. A. M. Dirac. Quantised Singularities in the Electromagnetic Field. *Proceedings of the Royal Society of London Series A*, 133(821):60–72, Sep 1931.

[113] P. A. M. Dirac. The Theory of Gravitation in Hamiltonian Form. *Proceedings of the Royal Society of London Series A*, 246:333–343, August 1958.

[114] P.A.M. Dirac. *Lectures on Quantum Mechanics*. Belfer Graduate School of Science, monograph series. Dover Publications, 2001.

[115] J.L. Doob. *Measure Theory*. Graduate Texts in Mathematics. Springer New York, 2012.

[116] A. G. Doroshkevich. Spatial structure of perturbations and origin of galactic rotation in fluctuation theory. *Astrophysics*, 6:320–330, October 1970.

[117] J. Diaz Dorronsoro, J. J. Halliwell, J. B. Hartle, T. Hertog, O. Janssen, and Y. Vreys. Damped Perturbations in the No-Boundary State. *Physical Review Letters*, 121(8):081302, Aug 2018.

[118] Gerald V. Dunne. Functional determinants in quantum field theory. *J. Phys.*, A41:304006, 2008.

[119] Gerald V. Dunne. Functional determinants in quantum field theory. *Journal of Physics A Mathematical General*, 41(30):304006, Aug 2008.

[120] Gerald V. Dunne and Mithat Unsal. What is QFT? Resurgent trans-series, Lefschetz thimbles, and new exact saddles. *arXiv e-prints*, page arXiv:1511.05977, Nov 2015.

[121] F. Dyson. A meeting with Enrico Fermi. *Nature*, 427:297, January 2004.

[122] A. S. Eddington. On the instability of Einstein's spherical world. *Monthly Notices of the Royal Astronomical Society*, 90:668–678, 1930.

[123] A. Einstein. Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen. *Annalen der Physik*, 322(8):549–560, Jan 1905.

[124] A. Einstein. Erklarung der Perihelionbewegung der Merkur aus der allgemeinen Relativitatstheorie. *Sitzungsber. preuss.Akad. Wiss., vol. 47, No.2, pp. 831-839, 1915*, 47:831–839, 1915.

[125] A. Einstein. Zur allgemeinen Relativitätstheorie. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 778-786.*, 1915.

[126] A. Einstein. Kosmologische Betrachtungen zur allgemeinen Relativitätstheorie. *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 142-152.*, 1917.

[127] A. Einstein. Bemerkung zu der Arbeit von A. Friedmann #8222Über die Krümmung des Raumes". *Zeitschrift fur Physik*, 11:326–326, December 1922.

[128] A. Einstein and W. de Sitter. On the Relation between the Expansion and the Mean Density of the Universe. *Proceedings of the National Academy of Science*, 18:213–214, March 1932.

[129] G. F. R. Ellis and M. A. H. MacCallum. A class of homogeneous cosmological models. *Communications in Mathematical Physics*, 12:108–141, June 1969.

[130] J. Elstrodt. *Maß- und Integrationstheorie*. Springer-Lehrbuch. Springer Berlin Heidelberg, 2009.

[131] Xinzhong Er and Adam Rogers. Two families of astrophysical diverging lens models. *Mon. Not. R. Astron. Soc.*, 475(1):867–878, Mar 2018.

[132] L. D. Faddeev. The Feynman integral for singular Lagrangians. *Theoretical and Mathematical Physics*, 1:1–13, October 1969.

[133] L. D. Faddeev and V. N. Popov. Feynman diagrams for the Yang-Mills field. *Physics Letters B*, 25:29–30, July 1967.

[134] B. L. Falck, M. C. Neyrinck, and A. S. Szalay. ORIGAMI: Delineating Halos Using Phase-space Folds. *Astrophysical Journal*, 754:126, August 2012.

[135] J. Feldbrugge, A. Fertig, L. Sberna, , and N. Turok. In preparation: The Spacetime Amplitude Approach to Relativistic Quantum Mechanics: the Schwinger Effect. 2019.

[136] J. Feldbrugge, J. Hidding, and R. van de Weygaert. Statistics of Caustics in Large-Scale Structure Formation. *Proceedings of IAU Symposium 308 "The Zeld'ovich Universe: Genesis and Growth of the Cosmic Web"*, December 2014.

[137] J. Feldbrugge and N. Turok. In preparation. 2019.

[138] Job Feldbrugge. Statistics of caustics in large-scale structure formation. Master's thesis, Rijksuniversiteit Groningen, the Netherlands, 2014.

[139] Job Feldbrugge, Jean-Luc Lehners, and Neil Turok. Lorentzian Quantum Cosmology. *Phys. Rev.*, D95(10):103508, 2017.

[140] Job Feldbrugge, Jean-Luc Lehners, and Neil Turok. No Smooth Beginning for Spacetime. *Physical Review Letters*, 119(17):171301, Oct 2017.

[141] Job Feldbrugge, Jean-Luc Lehners, and Neil Turok. Inconsistencies of the New No-Boundary Proposal. *Universe*, 4(10):100, 2018.

[142] Job Feldbrugge, Jean-Luc Lehners, and Neil Turok. No rescue for the no boundary proposal: Pointers to the future of quantum cosmology. *Phys. Rev.*, D97(2):023509, 2018.

[143] Job Feldbrugge, Ue-Li Pen, and Neil Turok. Oscillatory path integrals for radio astronomy. *arXiv e-prints*, page arXiv:1909.04632, Sep 2019.

[144] Job Feldbrugge, Rien van de Weygaert, Johan Hidding, and Joost Feldbrugge. Caustic Skeleton & Cosmic Web. *Journal of Cosmology and Astro-Particle Physics*, 2018(5):027, May 2018.

[145] Job Feldbrugge, Rien van de Weygaert, Johan Hidding, and Joost Feldbrugge. In preparation. 2019.

[146] R. P. Feynman. Space-Time Approach to Non-Relativistic Quantum Mechanics. *Reviews of Modern Physics*, 20(2):367–387, Apr 1948.

[147] R. P. Feynman. Mathematical formulation of the quantum theory of electromagnetic interaction. *Phys. Rev.*, 80:440–457, 1950.

[148] Richard P. Feynman, Fernando B. Morinigo, and William G. Wagner. *Feynman lectures on gravitation*. CRC Press, 1995.

[149] Richard P. Feynman, Steven Weinberg, and Per H. Andersen. Elementary Particles and the Laws of Physics: The 1986 Dirac Memorial Lectures. *Physics Today*, 41(4):96, Jan 1988.

[150] R.P. Feynman and A.R. Hibbs. *Quantum mechanics and path integrals*. International series in pure and applied physics. McGraw-Hill, 1965.

[151] R. L. Fiedler, B. Dennison, K. J. Johnston, and A. Hewish. Extreme scattering events caused by compact structures in the interstellar medium. *Nature*, 326(6114):675–678, Apr 1987.

[152] A. Friedmann. Über die Krümmung des Raumes. *Zeitschrift fur Physik*, 10:377–386, 1922.

[153] A. Friedmann. Über Wirbelbewegung in einer kompressiblen Flüssigkeit. *Zeitschrift Angewandte Mathematik und Mechanik*, 4:102–107, 1924.

[154] S. A. Fulling. *Aspects of Quantum Field Theory in Curved Spacetime*. Cambridge University Press, September 1989.

[155] G. Galilei. *Sidereus nuncius magna, longeque admirabilia spectacula pandens lunae facie, fixis innumeris, lacteo circulo, stellis nebulosis, ... Galileo Galileo : nuper a se reperti beneficio sunt observata in apprime vero in quatuor planetis circa Iovis stellam disparibus intervallis, atque periodis, celeritate mirabili circumvolutis ... atque Medicea sidera nuncupandos decrevit*. Medicea sidera nuncupandos decrevit, 1610.

[156] Luis J. Garay, Jonathan J. Halliwell, and Guillermo A. Mena Marugan. Path integral quantum cosmology: A Class of exactly soluble scalar field minisuperspace models with exponential potentials. *Phys. Rev.*, D43:2572–2589, 1991.

[157] I. M. Gel'fand and A. M. Yaglom. Integration in Functional Spaces and its Applications in Quantum Physics. *Journal of Mathematical Physics*, 1:48–69, January 1960.

[158] U. H. Gerlach and U. K. Sengupta. Homogeneous Collapsing Star: Tensor and Vector Harmonics for Matter and Field Asymmetries. *Phys. Rev.*, D18:1773–1784, 1978.

[159] G. W. Gibbons. The Einstein Action of Riemannian Metrics and Its Relation to Quantum Gravity and Thermodynamics. *Phys. Lett.*, A61:3–5, 1977.

[160] G. W. Gibbons, S. W. Hawking, and M. J. Perry. Path integrals and the indefiniteness of the gravitational action. *Nuclear Physics B*, 138:141–150, June 1978.

[161] Steven B. Giddings. The Conformal Factor and the Cosmological Constant. *Int. J. Mod. Phys.*, A5:3811–3830, 1990.

[162] Steven B. Giddings. Wormholes, the conformal factor, and the cosmological constant. In *International Colloquium on Modern Quantum Field Theory Bombay, India, January 8-14, 1990*, pages 0518–544, 1990.

[163] Steffen Gielen and Neil Turok. Perfect Quantum Cosmological Bounce. *Phys. Rev. Lett.*, 117(2):021301, 2016.

[164] Steffen Gielen and Neil Turok. Quantum propagation across cosmological singularities. *Physical Review D*, 95(10):103510, May 2017.

[165] R. Gilmore. *Catastrophe theory for scientists and engineers*. Wiley, 1981.

[166] J. Glimm and A. Jaffe. *Quantum Physics: A Functional Integral Point of View*. Springer-Verlag, 1987.

[167] H. Goldstein. *Classical Mechanics*. Addison-Wesley, 1980.

[168] D. L. Goodstein and J. R. Goodstein. *Feynman's lost lecture : the motion of planets around the sun*. New York : Norton, 1996.

[169] W. Gordon. Der Comptoneffekt nach der Schrödingerschen Theorie. *Zeitschrift fur Physik*, 40(1-2):117–133, Jan 1926.

[170] Steven Gratton, Antony Lewis, and Neil Turok. Closed universes from cosmological instantons. *Phys. Rev.*, D65:043513, 2002.

[171] M. B. Green, J. H. Schwarz, and E. Witten. *Superstring theory. Volume 1 - Introduction*. Cambridge University Press, 479 p., 1987.

[172] Gianfranco Grillo and James Cordes. Wave asymptotics and their application to astrophysical plasma lensing. *arXiv e-prints*, page arXiv:1810.09058, Oct 2018.

[173] L. P. Grishchuk, A. G. Doroshkevich, and V. M. Yudin. Long Gravitational Waves in a Closed Universe. *Zh. Eksp. Teor. Fiz.*, 69:1857–1871, 1975.

[174] Steiner Grosche and C. Frank. *Handbook of Feynman Path Integrals*. Berlin: Springer, 1998.

[175] S. N. Gurbatov, A. I. Saichev, and S. F. Shandarin. The large-scale structure of the universe in the frame of the model equation of non-linear diffusion. *Mon. Not. R. Astron. Soc.*, 236:385–402, January 1989.

[176] S. N. Gurbatov, A. I. Saichev, and S. F. Shandarin. Large-scale structure of the Universe. The Zeldovich approximation and the adhesion model. *Physics Uspekhi*, 55:223–249, March 2012.

[177] B.C. Hall. *Quantum Theory for Mathematicians*. Graduate Texts in Mathematics. Springer New York, 2013.

[178] J. Halliwell and S. Hawking. Quantum fluctuations as the cause of inhomogeneity in the Universe. In M. A. Markov, V. A. Berezin, and V. P. Frolov, editors, *Quantum Gravity*, page 509, 1985.

[179] J. Halliwell and S. Hawking. Quantum cosmology - beyond minisuperspace. In R. Ruffini, editor, *Fourth Marcel Grossmann Meeting on General Relativity*, pages 65–83, 1986.

[180] J. J. Halliwell and J. B. Hartle. Wave functions constructed from an invariant sum-over-histories satisfy constraints: three derivations of the Wheeler-DeWitt equation. In M. A. Markov, V. A. Berezin, and V. P. Frolov, editors, *Quantum Gravity V*, page 92, 1991.

[181] J. J. Halliwell, J. B. Hartle, and T. Hertog. What is the no-boundary wave function of the Universe? *Physical Review D*, 99(4):043526, February 2019.

[182] J. J. Halliwell and S. W. Hawking. The Origin of Structure in the Universe. *Phys. Rev.*, D31:1777, 1985.

[183] J. J. Halliwell and M. E. Ortiz. Sum-over-histories origin of the composition laws of relativistic quantum mechanics and quantum cosmology. *Physical Review D*, 48:748–768, July 1993.

[184] Jonathan J. Halliwell. Derivation of the Wheeler-De Witt Equation from a Path Integral for Minisuperspace Models. *Phys. Rev.*, D38:2468, 1988.

[185] Jonathan J. Halliwell and James B. Hartle. Integration Contours for the No Boundary Wave Function of the Universe. *Phys. Rev.*, D41:1815, 1990.

[186] Jonathan J. Halliwell and Jorma Louko. Steepest Descent Contours in the Path Integral Approach to Quantum Cosmology. 1. The De Sitter Minisuperspace Model. *Phys. Rev.*, D39:2206, 1989.

[187] Daniel Harlow. Metastability in Anti de Sitter Space. *arXiv e-prints*, page arXiv:1003.5909, Mar 2010.

[188] J. B. Hartle and S. W. Hawking. Wave Function of the Universe. *Phys. Rev.*, D28:2960–2975, 1983.

[189] James B. Hartle, S. W. Hawking, and Thomas Hertog. The Classical Universes of the No-Boundary Quantum State. *Phys. Rev.*, D77:123537, 2008.

[190] Allen Hatcher. *Algebraic topology*. Cambridge Univ. Press, Cambridge, 2000.

[191] S. W. Hawking. Singularities in the Universe. *Physical Review Letters*, 17:444–445, August 1966.

[192] S. W. Hawking. The Occurrence of Singularities in Cosmology. *Proceedings of the Royal Society of London Series A*, 294:511–521, October 1966.

[193] S. W. Hawking. The Occurrence of Singularities in Cosmology. II. *Proceedings of the Royal Society of London Series A*, 295:490–493, December 1966.

[194] S. W. Hawking. The Boundary Conditions of the Universe. *Pontif. Acad. Sci. Scr. Varia*, 48:563–574, 1982.

[195] S. W. Hawking. The Quantum State of the Universe. *Nucl. Phys.*, B239:257, 1984.

[196] S. W. Hawking and W. Israel. *General relativity. An Einstein centenary survey.* Cambridge University Press, 1979.

[197] W. Heisenberg. Über quantentheoretische Umdeutung kinematischer und mechanischer Beziehungen. *Zeitschrift fur Physik*, 33(1):879–893, Dec 1925.

[198] M. Henneaux. Hamiltonian form of the path integral for theories with a gauge freedom. *Review Section of Physics Letters*, 126:1–66, September 1985.

[199] M. Henneaux and C. Teitelboim. *Quantization of gauge systems.* Princeton, USA: Univ. Pr. 520 p, 1992.

[200] Thomas Hertog and Gary T. Horowitz. Towards a big crunch dual. *JHEP*, 07:073, 2004.

[201] Thomas Hertog and Gary T. Horowitz. Holographic description of AdS cosmologies. *JHEP*, 04:005, 2005.

[202] J.. Hidding. *The Phase-Space Geometry of the Cosmic Web*, Ph.D. thesis, University of Groningen, 2019.

[203] J. Hidding, S. F. Shandarin, and R. van de Weygaert. The Zel'dovich approximation: key to understanding cosmic web complexity. *Mon. Not. R. Astron. Soc.*, 437:3442–3472, February 2014.

[204] J. Hidding, R. van de Weygaert, and S. Shandarin. The Zeldovich & Adhesion approximations and applications to the local universe. In R. van de Weygaert, S. Shandarin, E. Saar, and J. Einasto, editors, *The Zeldovich Universe: Genesis and Growth of the Cosmic Web*, volume 308 of *IAU Symposium*, pages 69–76, October 2016.

[205] J. Hidding, R. van de Weygaert, G. Vegter, B. J. T. Jones, and M. Teillaud. The Sticky Geometry of the Cosmic Web. *ArXiv e-prints*, May 2012.

[206] S. A. Hojman, K. Kuchař, and C. Teitelboim. Geometrodynamics regained. *Annals of Physics*, 96:88–135, January 1976.

[207] F. Hoyle. A New Model for the Expanding Universe. *Notices of the Royal Astronomical Society*, 108:372, 1948.

[208] E. Hubble. A Relation between Distance and Radial Velocity among Extra-Galactic Nebulae. *Proceedings of the National Academy of Science*, 15:168–173, March 1929.

[209] Colin Hunter. "A New Picture of the Big Bang" –The Quantum Universe. *Inside the perimeter*, 2017.

[210] V. Icke. Formation of Galaxies inside Clusters. *Astron. Astrophys.*, 27:1, Aug 1973.

[211] C. J. Isham and J. Butterfield. On the Emergence of Time in Quantum Gravity. *arXiv General Relativity and Quantum Cosmology e-prints*, January 1999.

[212] Akira Ishimaru. Optical scattering and diffusion in turbulence and scatterers (A). *Journal of the Optical Society of America (1917-1983)*, 68:1368, Oct 1978.

[213] K. Itô. Wiener Integral and Feynman Integral. *Proc. Fourth Berkeley Symp. on Math., Stat. and Prob.*, 2:227–238, 1960.

[214] J.D. Jackson. *Classical Electrodynamics*. John Wiley, New York, 1975.

[215] J. Jasche, F. S. Kitaura, C. Li, and T. A. Enßlin. Bayesian non-linear large-scale structure inference of the Sloan Digital Sky Survey Data Release 7. *Mon. Not. R. Astron. Soc.*, 409:355–370, November 2010.

[216] B. J. T. Jones. *Precision Cosmology: The First Half Million Years*. Cambridge University Press, June 2017.

[217] A. Josephy et al. CHIME/FRB Detection of the Original Repeating Fast Radio Burst Source FRB 121102. 2019.

[218] M. Kac. On Some Connections between Probability Theory and Differential and Integral Equations. In *Second Berkeley Symposium on Mathematical Statistics and Probability*, pages 189–215, Jan 1951.

[219] I. Kant and K. Kehrbach. *Kritik der reinen Vernunft: Text der Ausgabe 1781 mit Beifügung sämmtlicher Abweichungen der Ausgabe 1787*. Reclams Universal-Bibliothek. Reclam, 1878.

[220] J. Kepler. *Astronomia nova ..., seu physica coelestis, tradita commentariis de motibus stellae martis*. 1609.

[221] F.-S. Kitaura. The initial conditions of the Universe from constrained simulations. *Mon. Not. R. Astron. Soc.*, 429:L84–L88, February 2013.

[222] John R. Klauder. *The Feynman Path Integral: An Historical Slice*, pages 55–76. World Scientific Publishing Co. Pte. Ltd., 2003.

[223] J.R. Klauder. *A Modern Approach to Functional Integration*. Applied and Numerical Harmonic Analysis. Birkhäuser Boston, 2010.

[224] Oskar Klein. Quantentheorie und fünfdimensionale Relativitätstheorie. *Zeitschrift fur Physik*, 37(12):895–906, Dec 1926.

[225] E. Komatsu, A. Kogut, M. R. Nolta, C. L. Bennett, M. Halpern, G. Hinshaw, N. Jarosik, M. Limon, S. S. Meyer, L. Page, D. N. Spergel, G. S. Tucker, L. Verde, E. Wollack, and E. L. Wright. First-Year Wilkinson Microwave Anisotropy Probe (WMAP) Observations: Tests of Gaussianity. *The Astrophysical Journal Supplement*, 148:119–134, September 2003.

[226] Helge S. Kragh. *Conceptions of Cosmos: From Myths to the Accelerating Universe: A History of Cosmology*. University of Oxford Press, 2006.

[227] Lawrence Krauss. Cosmological antigravity. *Scientific American*, 2002.

[228] I. A. Kravtsov and I. I. Orlov. Caustics, catastrophes, and wave fields. *Uspekhi Fizicheskikh Nauk*, 141:591–627, December 1983.

[229] K. Kuchař. Geometrodynamics regained: A Lagrangian approach. *Journal of Mathematical Physics*, 15:708–715, June 1974.

[230] K. Kuchař. Measure for measure: Covariant skeletonizations of phase space path integrals for systems moving on Riemannian manifolds. *Journal of Mathematical Physics*, 24:2122–2141, August 1983.

[231] K. V. Kuchař. Time and Interpretations of Quantum Gravity. *International Journal of Modern Physics D*, 20:3–86, 2011.

[232] L.D. Landau and E.M. Lifshitz. *Volume 1: Mechanics*. Pergamom, 1959.

[233] L.D. Landau and E.M. Lifshitz. *Volume 6: Fluid Mechanics*. Butterworth-Heinemann, 1976.

[234] F. Leclercq, J. Jasche, G. Lavaux, B. Wandelt, and W. Percival. The phase-space structure of nearby dark matter as constrained by the SDSS. *Journal of Cosmology and Astroparticle Physics*, 6:049, June 2017.

[235] F. Leclercq, J. Jasche, and B. Wandelt. Bayesian analysis of the dynamic cosmic web in the SDSS galaxy survey. *Journal of Cosmology and Astroparticle Physics*, 6:015, June 2015.

[236] Jean-Luc Lehners. Classical Inflationary and Ekpyrotic Universes in the No-Boundary Wavefunction. *Phys. Rev.*, D91(8):083525, 2015.

[237] G. Lemaître. Un Univers homogène de masse constante et de rayon croissant rendant compte de la vitesse radiale des nébuleuses extra-galactiques. *Annales de la Société Scientifique de Bruxelles*, 47:49–59, 1927.

[238] G. Lemaître. Expansion of the universe, A homogeneous universe of constant mass and increasing radius accounting for the radial velocity of extra-galactic nebulae. *Monthly Notices of the Royal Astronomical Society, Vol. 91, p.483-490*, 91:483–490, 1931.

[239] Georges Lemaitre. Republication of: The beginning of the world from the point of view of quantum theory. *Nature*, 127:706, 1931. [Gen. Rel. Grav.43,2929(2011)].

[240] N. I Libeskind, R. van de Weygaert, M. Cautun, B. Falck, E. Tempel, T. Abel, M. Alpaslan, M. A. Aragoon-Calvo, J. E. Forero-Romero, R. Gonzalez, S. Gottloober, O. Hahn, W. A. Hellwing, Y. Hoffman, B. J. T. Jones, F. Kitaura, A. Knebe, S. Manti, M. Neyrinck, S. E. Nuza, N. Padilla, E. Platen, N. Ramachandra, A. Robotham, E. Saar, S. Shandarin, M. Steinmetz, R. S. Stoica, T. Sousbie, and G. Yepes. Tracing the cosmic web. *ArXiv e-prints*, May 2017.

[241] Andrei D. Linde. Quantum Creation of the Inflationary Universe. *Lett. Nuovo Cim.*, 39:401–405, 1984.

[242] D. R. Lorimer, M. Bailes, M. A. McLaughlin, D. J. Narkevic, and F. Crawford. A Bright Millisecond Radio Burst of Extragalactic Origin. *Science*, 318(5851):777, Nov 2007.

[243] M. A. H. MacCallum and G. F. R. Ellis. A class of homogeneous cosmological models: II. Observations. *Communications in Mathematical Physics*, 19:31–64, March 1970.

[244] Robert Main, I. Sheng Yang, Victor Chan, Dongzi Li, Fang Xi Lin, Nikhil Mahajan, Ue-Li Pen, Keith Vanderlinde, and Marten H. van Kerkwijk. Pulsar emission amplified and resolved by plasma lensing in an eclipsing binary. *Nature*, 557(7706):522–525, May 2018.

[245] Juan Maldacena. Vacuum decay into Anti de Sitter space. *arXiv e-prints*, page arXiv:1012.0274, Dec 2010.

[246] J.P. May. *A Concise Course in Algebraic Topology*. Chicago Lectures in Mathematics. University of Chicago Press, 1999.

[247] J. Mayer, K. Khairy, and J. Howard. Drawing an elephant with four complex parameters. *American Journal of Physics*, 78:648–649, June 2010.

[248] S. Mazzucchi. *Mathematical Feynman Path Integrals and Their Applications*. World Scientific, 2009.

[249] Paul McFadden and Kostas Skenderis. Holography for Cosmology. *Phys. Rev.*, D81:021301, 2010.

[250] D. B. Melrose and P. G. Watson. Scintillation of Radio Sources: The Role of Caustics. *Astrophysical Journal*, 647(2):1131–1141, Aug 2006.

[251] D. Michilli, A. Seymour, J. W. T. Hessels, L. G. Spitler, V. Gajjar, A. M. Archibald, G. C. Bower, S. Chatterjee, J. M. Cordes, K. Gourdji, G. H. Heald, V. M. Kaspi, C. J. Law, C. Sobey, E. A. K. Adams, C. G. Bassa, S. Bogdanov, C. Brinkman, P. Demorest, F. Fernand ez, G. Hellbourg, T. J. W. Lazio, R. S. Lynch, N. Maddox, B. Marcote, M. A. McLaughlin, Z. Paragi, S. M. Ransom, P. Scholz, A. P. V. Siemion, S. P. Tendulkar, P. van Rooy, R. S. Wharton, and D. Whitlow. An extreme magneto-ionic environment associated with the fast radio burst source FRB 121102. *Nature*, 553(7687):182–185, Jan 2018.

[252] C. W. Misner. Feynman Quantization of General Relativity. *Reviews of Modern Physics*, 29:497–509, July 1957.

[253] C. W. Misner. Mixmaster Universe. *Physical Review Letters*, 22:1071–1074, May 1969.

[254] Charles W. Misner. Mixmaster universe. *Phys. Rev. Lett.*, 22:1071–1074, 1969.

[255] P. J. Morrison. Hamiltonian description of the ideal fluid. *Reviews of Modern Physics*, 70:467–521, April 1998.

[256] M. Morse. Relations between the critical points of a real function on n independent variables. *Transactions of the American Mathematical Society*, 27:345–396, 1925.

[257] V. Mukhanov and S. Winitzki. *Introduction to Quantum Effects in Gravity*. Cambridge University Press, January 2007.

[258] T. T. Nakamura and S. Deguchi. Wave Optics in Gravitational Lensing. *Progress of Theoretical Physics Supplement*, 133:137–153, Jan 1999.

[259] D. Nelson, A. Pillepich, S. Genel, M. Vogelsberger, V. Springel, P. Torrey, V. Rodriguez-Gomez, D. Sijacki, G. F. Snyder, B. Griffen, F. Marinacci, L. Blecha, L. Sales, D. Xu, and L. Hernquist. The illustris simulation: Public data release. *Astronomy and Computing*, 13:12–37, November 2015.

[260] J. v. Neumann. Die Eindeutigkeit der Schrödingerschen Operatoren. *J. Math. Ann.*, 104, 1931.

[261] I. Newton. *Philosophiae Naturalis Principia Mathematica*. 1687.

[262] M. C. Neyrinck. Origami constraints on the initial-conditions arrangement of dark-matter caustics and streams. *ArXiv e-prints*, February 2012.

[263] J.D. North. *Cosmos: an illustrated history of astronomy and cosmology*. University of Chicago Press, 2008.

[264] J. R. Oppenheimer. On the Theory of Electrons and Protons. *Physical Review*, 35(5):562–563, Mar 1930.

[265] P. J. E. Peebles. *The large-scale structure of the universe*. Princeton University Press, 1980.

[266] Ue-Li Pen. The nature of fast radio bursts. *Nature Astronomy*, 2:842–844, Oct 2018.

[267] Ue-Li Pen and Lindsay King. Refractive convergent plasma lenses explain extreme scattering events and pulsar scintillation. *Mon. Not. R. Astron. Soc.*, 421(1):L132–L136, Mar 2012.

[268] Ue-Li Pen and Lindsay King. Refractive convergent plasma lenses explain extreme scattering events and pulsar scintillation. *Mon. Not. R. Astron. Soc.*, 421(1):L132–L136, Mar 2012.

[269] Ue-Li Pen and Yuri Levin. Pulsar scintillations from corrugated reconnection sheets in the interstellar medium. *Mon. Not. R. Astron. Soc.*, 442(4):3338–3346, Aug 2014.

[270] A. A. Penzias and R. W. Wilson. A Measurement of Excess Antenna Temperature at 4080 Mc/s. *Astrophysical Journal*, 142:419–421, July 1965.

[271] A. Peres. On Cauchy's problem in general relativity - II. *Il Nuovo Cimento*, 26:53–62, October 1962.

[272] Michael E. Peskin and Daniel V. Schroeder. *An Introduction to Quantum Field Theory*. 1995.

[273] Planck Collaboration. Planck 2018 results. VI. Cosmological parameters. *arXiv e-prints*, July 2018.

[274] Planck Collaboration, P. A. R. Ade, N. Aghanim, M. Arnaud, F. Arroja, M. Ashdown, J. Aumont, C. Baccigalupi, M. Ballardini, A. J. Banday, and et al. Planck 2015 results. XVII. Constraints on primordial non-Gaussianity. *Astron. Astrophys.*, 594:A17, September 2016.

[275] E. Platts, A. Weltman, A. Walters, S. P. Tendulkar, J. E. B. Gordin, and S. Kandhai. A Living Theory Catalogue for Fast Radio Bursts. *arXiv e-prints*, page arXiv:1810.05836, Oct 2018.

[276] D. Pogosyan, C. Pichon, C. Gay, S. Prunet, J. F. Cardoso, T. Sousbie, and S. Colombi. The local theory of the cosmic skeleton. *Mon. Not. R. Astron. Soc.*, 396:635–667, June 2009.

[277] A. M. Polyakov. De Sitter space and eternity. *Nucl. Phys.*, B797:199–217, 2008.

[278] A. M. Polyakov. Infrared instability of the de Sitter space. 2012.

[279] A. Pontzen and A. Challinor. Bianchi model CMB polarization and its implications for CMB anomalies. *Mon. Not. R. Astron. Soc.*, 380:1387–1398, October 2007.

[280] T. Poston and I.N. Stewart. *Catastrophe Theory and Its Applications*. Pitman, 1978.

[281] Nesar S. Ramachandra and Sergei F. Shandarin. Multi-stream portrait of the cosmic web. *Monthly Notices of the Royal Astronomical Society*, 452(2):1643–1653, 2015.

[282] R. Michael Range. Complex analysis: A brief tour into higher dimensions. *The American Mathematical Monthly*, 110:89–108, 2003.

[283] B. J. Rickett. Interstellar scattering and scintillation of radio waves. *Annual review of astronomy and astrophysics*, 15:479–504, Jan 1977.

[284] A. G. Riess, A. V. Filippenko, P. Challis, A. Clocchiatti, A. Diercks, P. M. Garnavich, R. L. Gilliland, C. J. Hogan, S. Jha, R. P. Kirshner, B. Leibundgut, M. M. Phillips, D. Reiss, B. P. Schmidt, R. A. Schommer, R. C. Smith, J. Spyromilio, C. Stubbs, N. B. Suntzeff, and J. Tonry. Observational Evidence from Supernovae for an Accelerating Universe and a Cosmological Constant. *Astrophysical Journal*, 116:1009–1038, September 1998.

[285] H. P. Robertson. Kinematics and World-Structure. *Astrophysical Journal*, 82:284, 1935.

[286] J. Rotman. *An Introduction to Algebraic Topology*. Graduate Texts in Mathematics. Springer New York, 1998.

[287] V. A. Rubakov. Quantum cosmology. In *NATO Advanced Study Institute: Structure Formation in the Universe Cambridge, England, July 26-August 6, 1999*, 1999.

[288] R. Salmon. Hamiltonian fluid mechanics. *Annual Review of Fluid Mechanics*, 20:225–256, 1988.

[289] J. Schaye, R. A. Crain, R. G. Bower, M. Furlong, M. Schaller, T. Theuns, C. Dalla Vecchia, C. S. Frenk, I. G. McCarthy, J. C. Helly, A. Jenkins, Y. M. Rosas-Guevara, S. D. M. White, M. Baes, C. M. Booth, P. Camps, J. F. Navarro, Y. Qu, A. Rahmati, T. Sawala, P. A. Thomas, and J. Trayford. The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *Mon. Not. R. Astron. Soc.*, 446:521–554, January 2015.

[290] J. Schaye, R. A. Crain, R. G. Bower, M. Furlong, M. Schaller, T. Theuns, C. Dalla Vecchia, C. S. Frenk, I. G. McCarthy, J. C. Helly, A. Jenkins, Y. M. Rosas-Guevara, S. D. M. White, M. Baes, C. M. Booth, P. Camps, J. F. Navarro, Y. Qu, A. Rahmati, T. Sawala, P. A. Thomas, and J. Trayford. The EAGLE project: simulating the evolution and assembly of galaxies and their environments. *Mon. Not. R. Astron. Soc.*, 446:521–554, January 2015.

[291] P.A. Schilpp. *Albert Einstein: Philosopher-Scientist*. Number v. 1 in The Library of living philosophers. Open Court Press, 1970.

[292] B. P. Schmidt, N. B. Suntzeff, M. M. Phillips, R. A. Schommer, A. Clocchiatti, R. P. Kirshner, P. Garnavich, P. Challis, B. Leibundgut, J. Spyromilio, A. G. Riess, A. V. Filippenko, M. Hamuy, R. C. Smith, C. Hogan, C. Stubbs, A. Diercks, D. Reiss, R. Gilliland, J. Tonry, J. Maza, A. Dressler, J. Walsh, and R. Ciardullo. The High-Z Supernova Search: Measuring Cosmic Deceleration and Global Curvature of the

Universe Using Type IA Supernovae. *Astrophysical Journal*, 507:46–63, November 1998.

[293] E. Schrödinger. Quantisierung als Eigenwertproblem. *Annalen der Physik*, 384(4):361–376, Jan 1926.

[294] R. Scoccimarro. Gravitational Clustering from χ^2 Initial Conditions. *Astrophysical Journal*, 542:1–8, October 2000.

[295] Marco Serone, Gabriele Spada, and Giovanni Villadoro. The power of perturbation theory. *Journal of High Energy Physics*, 2017(5):56, May 2017.

[296] S. Shandarin, S. Habib, and K. Heitmann. Cosmic web, multistream flows, and tessellations. *Phys. Rev. D*, 85(8):083005, April 2012.

[297] S. F. Shandarin and R. A. Sunyaev. The conjecture of the cosmic web. Commentary on: Zel'dovich Ya. B., 1970, A&A, 5, 84. *A&A*, 500:19–20, June 2009.

[298] S. F. Shandarin and Y. B. Zel'dovich. The large-scale structure of the universe: Turbulence, intermittency, structures in a self-gravitating medium. *Reviews of Modern Physics*, 61:185–220, April 1989.

[299] B. Simon. *Functional Integration and Quantum Physics*. AMS Chelsea Publishing Series. AMS Chelsea Pub., American Mathematical Society, 2005.

[300] Rafael D. Sorkin. Is the spacetime metric Euclidean rather than Lorentzian? *arXiv e-prints*, page arXiv:0911.1479, Nov 2009.

[301] L. G. Spitler, J. M. Cordes, J. W. T. Hessels, D. R. Lorimer, M. A. McLaughlin, S. Chatterjee, F. Crawford, J. S. Deneva, V. M. Kaspi, R. S. Wharton, B. Allen, S. Bogdanov, A. Brazier, F. Camilo, P. C. C. Freire, F. A. Jenet, C. Karako-Argaman, B. Knispel, P. Lazarus, K. J. Lee, J. van Leeuwen, R. Lynch, S. M. Ransom, P. Scholz, X. Siemens, I. H. Stairs, K. Stovall, J. K. Swiggum, A. Venkataraman, W. W. Zhu, C. Aulbert, and H. Fehrmann. Fast Radio Burst Discovered in the Arecibo Pulsar ALFA Survey. *Astrophysical Journal*, 790(2):101, Aug 2014.

[302] V. Springel, S. D. M. White, A. Jenkins, C. S. Frenk, N. Yoshida, L. Gao, J. Navarro, R. Thacker, D. Croton, J. Helly, J. A. Peacock, S. Cole, P. Thomas, H. Couchman, A. Evrard, J. Colberg, and F. Pearce. Simulations of the formation, evolution and clustering of galaxies and quasars. *Nature*, 435:629–636, June 2005.

[303] E. Stein, J.W. Milnor, M. Spivak, R. Wells, R. Wells, and J.N. Mather. *Morse Theory*. Annals of mathematics studies. Princeton University Press, 1963.

[304] M. H. Stone. Linear Transformations in Hilbert Space. III. Operational Methods and Group Theory. *Proceedings of the National Academy of Science*, 16(2):172–175, Feb 1930.

[305] M. H. Stone. On one-parameter unitary groups in hilbert space. *Annals of Mathematics*, 33(3):643–648, 1932.

[306] Yuya Tanizaki and Takayuki Koike. Real-time Feynman path integral with Picard-Lefschetz theory and its applications to quantum tunneling. *Annals of Physics*, 351:250–274, Dec 2014.

[307] C. Teitelboim. Quantum mechanics on the gravitational field. *Phys. Rev.*, 25:3159–3179, jun 1982.

[308] C. Teitelboim. Causality versus gauge invariance in quantum gravity and supergravity. *Phys. Rev.*, 50:705–708, March 1983.

[309] Claudio Teitelboim. Causality Versus Gauge Invariance in Quantum Gravity and Supergravity. *Phys. Rev. Lett.*, 50:705, 1983.

[310] Claudio Teitelboim. The Proper Time Gauge in Quantum Theory of Gravitation. *Phys. Rev.*, D28:297, 1983.

[311] R. Thom. *Structural Stability And Morphogenesis*. CRC Press, 2018.

[312] Ian Thompson. NIST Handbook of Mathematical Functions, edited by Frank W.J. Olver, Daniel W. Lozier, Ronald F. Boisvert, Charles W. Clark. *Contemporary Physics*, 52(5):497–498, Sep 2011.

[313] Matthias Troyer and Uwe-Jens Wiese. Computational Complexity and Fundamental Limitations to Fermionic Quantum Monte Carlo Simulations. *Physical Review Letters*, 94(17):170201, May 2005.

[314] Edward P. Tryon. Is the universe a vacuum fluctuation. *Nature*, 246:396, 1973.

[315] I. V. Tyutin. Gauge Invariance in Field Theory and Statistical Physics in Operator Formalism. *ArXiv e-prints*, December 2008.

[316] W. G. Unruh and R. M. Wald. Time and the interpretation of canonical quantum gravity. *Physical Review D*, 40:2598–2614, October 1989.

[317] M. v. d. Kamp, P. S. Cannon, and M. Terkildsen. Effect of the ionosphere on defocusing of space-based radars. *Radio Science*, 44(01):1–15, Feb 2009.

[318] J. v. Neumann. Uber einen satz von herrn m. h. stone. *Annals of Mathematics*, 33(3):567–573, 1932.

[319] R. van de Weygaert and J. R. Bond. Clusters and the Theory of the Cosmic Web. In M. Plionis, O. López-Cruz, and D. Hughes, editors, *A Pan-Chromatic View of Clusters of Galaxies and the Large-Scale Structure*, volume 740 of *Lecture Notes in Physics, Berlin Springer Verlag*, page 335, 2008.

[320] A. N. Varchenko. Zeta-Function of Monodromy and Newton's Diagram. *Inventiones Mathematicae*, 37:253, Jan 1976.

[321] V. A. Vasil'ev. Asymptotic exponential integrals, Newton's diagram, and the classification of minimal points. *Plenum Publishing Corporation*, 1978.

[322] M. Vergassola, B. Dubrulle, U. Frisch, and A. Noullez. Burgers' equation, Devil's staircases and the mass distribution for large-scale structures. *Astron. Astrophys.*, 289:325–356, September 1994.

[323] A. Vilenkin. Quantum Creation of Universes. *Phys. Rev.*, D30:509–511, 1984.

[324] A. Vilenkin and M. Yamada. Tunneling wave function of the universe. II. The back-reaction problem. *Physical Review D*, 99(6):066010, March 2019.

[325] Alexander Vilenkin. Creation of Universes from Nothing. *Phys. Lett.*, B117:25–28, 1982.

[326] Alexander Vilenkin. The Birth of Inflationary Universes. *Phys. Rev.*, D27:2848, 1983.

[327] Alexander Vilenkin. Boundary Conditions in Quantum Cosmology. *Phys. Rev.*, D33:3560, 1986.

[328] Alexander Vilenkin. The Interpretation of the Wave Function of the Universe. *Phys. Rev.*, D39:1116, 1989.

[329] Alexander Vilenkin. Approaches to quantum cosmology. *Phys. Rev.*, D50:2581–2594, 1994.

[330] Alexander Vilenkin and Masaki Yamada. Tunneling wave function of the universe. *Phys. Rev.*, D98(6):066003, 2018.

[331] M. Vogelsberger, S. Genel, V. Springel, P. Torrey, D. Sijacki, D. Xu, G. Snyder, D. Nelson, and L. Hernquist. Introducing the Illustris Project: simulating the coevolution of dark and visible matter in the Universe. *Mon. Not. R. Astron. Soc.*, 444:1518–1547, October 2014.

[332] M. von Smoluchowski. Zur kinetischen Theorie der Brownschen Molekularbewegung und der Suspensionen. *Annalen der Physik*, 326(14):756–780, Jan 1906.

[333] J. Wainwright and G. F. R. Ellis, editors. *Dynamical Systems in Cosmology*, 1997.

[334] R. M. Wald. Asymptotic behavior of homogeneous cosmological models in the presence of a positive cosmological constant. *Physical Review D*, 28:2118–2120, October 1983.

[335] R. M. Wald. *General relativity*. Chicago, University of Chicago Press, 1984.

[336] R. M. Wald. Quantum field theory in curved space-time. In B. Julia and J. Zinn-Justin, editors, *Gravitation and Quantizations, Session LVII of Les Houches*, page 63, 1995.

[337] A. G. Walker. On Riemannian spaces with spherical symmetry about a line, and the conditions for isotropy in general relativity. *The Quarterly Journal of Mathematics*, 6:81–93, 1935.

[338] X. Wang, A. Szalay, M. A. Aragón-Calvo, M. C. Neyrinck, and G. L. Eyink. Kinematic Morphology of Large-scale Structure: Evolution from Potential to Rotational Flow. *Astrophysical Journal*, 793:58, September 2014.

[339] J. A. Wheeler. Geometrodynamics and the Problem of Motion. *Reviews of Modern Physics*, 33:63–78, January 1961.

[340] J. A. Wheeler. *Superspace and the Nature of Quantum Geometrodynamics*, page 27. Singapore: World Scientific, 1987.

[341] Edward Witten. A New Look At The Path Integral Of Quantum Mechanics. *arXiv e-prints*, page arXiv:1009.6032, Sep 2010.

[342] Edward Witten. Analytic Continuation Of Chern-Simons Theory. *AMS/IP Stud. Adv. Math.*, 50:347–446, 2011.

[343] I. B. Zeldovich, A. V. Mamaev, and S. F. Shandarin. Laboratory observation of caustics, optical simulation of the motion of particles, and cosmology. *Uspekhi Fizicheskikh Nauk*, 139:153–163, January 1983.

[344] Y. B. Zel'dovich. Gravitational instability: An approximate theory for large density perturbations. *A&A*, 5:84–89, March 1970.

[345] Mike Zeng. Physicists Debate Hawking's idea that the universe had no beginning. *Quanta magazine*, 2019.