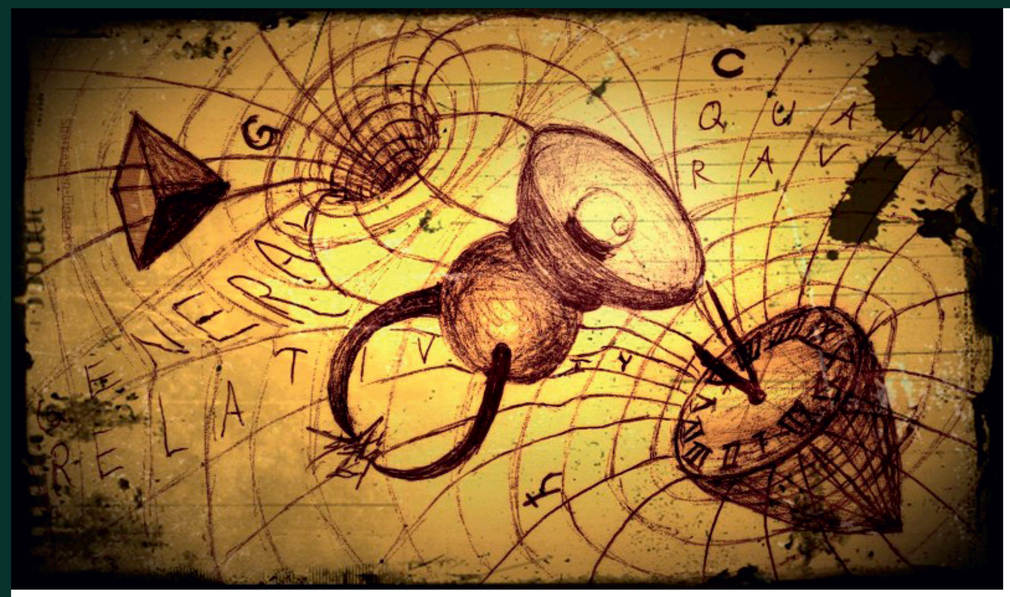
Nonlinear dynamics and the instability of Anti-de Sitter space

Dimitrakopoulos, F.

Nonlinear dynamics and the instability of Anti-de Sitter space

F. V. Dimitrakopoulos

Nonlinear dynamics and the instability of Anti-de Sitter space



Fotios V. Dimitrakopoulos

UNIVERSITY OF AMSTERDAM

Nonlinear Dynamics and the Instability of Anti-de Sitter Space

This work has been accomplished at the Gravity and AstroParticle Physics in Amsterdam (GRAPPA) and the Institute for Theoretical Physics (ITFA) of the University of Amsterdam (UvA) and is part of the Delta ITP consortium, a program of the Netherlands Organisation for Scientific Research (NWO) that is funded by the Dutch Ministry of Education, Culture and Science (OCW).

ISBN: 978-94-028-0675-5

Cover illustration by Dywiann Xyara

Nonlinear Dynamics and the Instability of Anti-de Sitter Space

Academisch Proefschrift

ter verkrijging van de graad van doctor

aan de Universiteit van Amsterdam

op gezag van de Rector Magnificus

prof. dr. ir. K.I.J. Maex

ten overstaan van een door het College voor Promoties ingestelde

commissie, in het openbaar te verdedigen in de Agnietenkapel

op donderdag 15 Juni 2017, te 12.00 uur

door

Fotios Dimitrakopoulos

geboren te Athene, Griekenland

Promotiecommissie

Promotor

prof. dr. J. de Boer	Universiteit van Amsterdam

Co-Promotores

dr. B.W. Freivogel	Universiteit van Amsterdam

Overige leden

prof. dr. E. P. Verlinde	Universiteit van Amsterdam
prof. dr. M.P. Decowski	Universiteit van Amsterdam
prof. dr. B.E.L. Craps	Vrije Universiteit Brussel
dr. J.P. van der Schaar	Universiteit van Amsterdam
dr. J. Mas	Universidad de Santiago de Compostela

Faculteit der Natuurwetenschappen, Wiskunde en Informatica

Publications

This thesis is based on the following publications:

[1] F.V Dimitrakopoulos, Ben Freivogel and Juan F. Pedraza
Fast and slow coherent cascades in Anti-de Sitter spacetime
arXiv:1612.04758 [hep-th].

[2] F.V Dimitrakopoulos, Ben Freivogel, Juan F. Pedraza and I-Sheng Yang
Gauge dependence of the AdS instability problem
Phys. Rev D **94**, 124008, arXiv:1607.08094 [hep-th].

[3] F.V Dimitrakopoulos and I-Sheng Yang
Conditionally extended validity of perturbation theory: Persistence of AdS stability islands
Phys. Rev D **92**, 083013, arXiv:1507.02684 [hep-th].

[4] F.V Dimitrakopoulos, Ben Freivogel, Matthew Lippert and I-Sheng Yang
Position space analysis of the AdS (in)stability problem
JHEP 06(2015) **095**, arXiv:1410.1880 [hep-th].

Other publications by the author:

[5] F.V Dimitrakopoulos, Laursen Kabir, Benjamin Mosk, Maulik Parikh and Jan Pieter van der Schaar
Vacua and correlators in hyperbolic de Sitter space
JHEP **1506**, (2015) 095, arXiv:1410.1880 [hep-th].

Preface & Thesis Guide

This dissertation consists of part of my research work conducted in the four years of my doctoral studies. Broadly speaking, my research focused on problems on Classical and Quantum Gravity and it was conducted in terms of different collaborations[1]. This thesis is concentrated on the *classical* part of my work and specifically the mathematical stability of one of the three ground state solutions to Einstein's theory of gravity, General Relativity.

In the first chapter I will give an in depth introduction to the subject, whereas the four subsequent chapters constitute the developments into which I was involved. Each chapter corresponds to a published work. In chapter 2 a position-space analysis of the problem will be presented, complementing previous studies that where concentrated in Fourier-space analysis. Since, all of the analytical studies in the subject rely on perturbative/approximate schemes, in chapter 3 the validity, up to the relevant time scales, of these perturbative schemes is discussed. In chapter 4, we clarify some confusion that arises due to the gauge freedom of the problem at hand, whereas in the next chapter the evolution of a few initial perturbations is discussed with emphasis on the resulting phase- and amplitude-spectra. We conclude with a short summary of the work as well as some considerations for future directions along those lines.

This thesis is written using the pronoun "we" instead of "I". This is both a standard practice in the field of Theoretical Physics as well as it describes the collaborative nature of research.

[1] My individual contribution to each of the published work will be clarified at the end.

Contents

1 Introduction

The career of a young theoretical physicist consists of treating the harmonic oscillator in ever-increasing levels of abstraction.

— Sidney Coleman

In this introductory chapter background material will be discussed regarding the various concepts that are important to the rest of this thesis. We will introduce the problem at hand and describe the state of knowledge before our involvement. More specifically, the model of spherical symmetric perturbations in AdS and the corresponding approximate schemes will be reviewed and both numerical and analytical results will be discussed. Our aim is to keep this section as *reader friendly* as possible, by keeping it short and minimising the amount of technical details. Unless it is completely unavoidable, the story will be conveyed by means of words and figures.

1.1 Overview

Historically, stability considerations and perturbation theory date back to the era of celestial mechanics and the question of stability of the solar system over long time scales. Between 1609 and 1618 Johannes Kepler determined the trajectories of the planets as they revolve around the Sun. Following the work of Copernicus, Kepler placed the Sun at the centre of the universe and based on observations of the famous astronomer of the time Tycho Brache, he succeeded to show that planets move in ellipses around the Sun and at the end of the revolution the planets find themselves back to where they started.

However, this picture of a perfectly stable solar system would be soon challenged. After Isaac Newton developed his theory about gravity, he derived the Keplerian orbits by restricting to the interaction of a planet with the Sun alone. Although this is the leading contribution to the gravitational force exerted to each planet, it

is not the only one. Planets attract each other as well. When these perturbations are taken into account they might lead to small effects which accumulate in the course of time destroying in that way the Keplerian orbits.

The study of the stability of the solar system has led to remarkable discoveries in Physics and Mathematics with the most prominent one being perhaps the celebrated Kolmogorov-Arnold-Moser (KAM) theory in which it was rigorously shown that both stable and unstable orbits exist depending on whether the ratio of the unperturbed frequencies is a rational number.

Newton's theory was superseded when Albert Einstein published in 1915 his theory of gravitation, known as General Relativity (GR). According to Einstein, gravity is not a force but rather the manifestation of the geometry of spacetime in which the masses move. Massive objects curve the spacetime and spacetime back-reacts to the masses by dictating them which paths they should follow. Einstein's equations possess three vacuum solutions, namely three different empty spacetimes depending on whether the cosmological constant of the theory is positive (de Sitter), zero (Minkowski) or negative (Anti-de Sitter). The most important question regarding a vacuum state is whether it is stable under small perturbations.

The stability of the vacuum solutions of GR comes second (perhaps even first) only to the stability of the solar system and has led to one of the greatest developments in mathematical relativity [6]. Of the three vacuum spacetimes the two where proven to be stable long ago [6, 7]. The stability of the third one (AdS) was not even raised, let alone answered, until very recently [8].

Anti-de Sitter (AdS) spacetime plays a prominent role in modern Theoretical Physics mainly due to its role in the only concrete example of a gauge/gravity duality, the AdS/CFT correspondence [5]. In this picture, a Quantum Field Theory (QFT) living on the boundary of AdS is equivalent to a String Theory in the AdS background. Despite the great importance of (asymptotically) AdS spacetime(s), the study of its (nonlinear) stability was initiated only very recently, albeit it was earlier conjectured that AdS would be nonlinearly unstable [9].

The first model that was developed and has been mostly used so far is that of spherically symmetric perturbations in the form of a (massless) scalar field [4, 8, 10–13]. One then tracks the evolution of the perturbation to determine whether it collapses to form a Black Hole or not.

Before we explain the case of AdS, let us start by discussing the fate of small perturbations in Minkowski space. Consider a self-gravitating spherically symmetric shell of a massless scalar field. If the amplitude of the shell is big enough it will collapse to form a black hole. For sufficiently small amplitudes however, the shell scatters and disperses to infinity and as the the gravitational interactions become

progressively weaker there is no black hole formation. Minkowski is therefore *stable* due to dispersion of energy to infinity.

One can already see that the story for *AdS* will be different. Anti-de Sitter space has a conformal boundary at spatial infinity which makes it effectively a confining box[1]. Now the *small* perturbations that do not immediately form a black hole, scatter at the boundary and propagate back to the interior where they have another chance to interact. This model was first studied in detail in the seminal work of [8] and the picture that emerged can be summarized in the figures (1.1) and (1.2).

We will give more details in the subsequent, slightly technical, section, however the story is not very hard to explain even with no equations at all. In [8] the evolution of spherically symmetric Gaussian wavepackets, with their amplitude being described by the small parameter ϵ, was studied numerically. For large values of the amplitude ϵ, the wave packet quickly collapses, which is signalled by the formation of a horizon in the spacetime. As ϵ was further decreased, so did the size of the horizon, until ϵ reaches a critical value ϵ_0 for which the size of the horizon becomes zero. When the amplitude is decreased further below this value of ϵ_0, a horizon will form, but after the wave packet has reflected at the boundary **once**. Decreasing ϵ further, a second critical amplitude ϵ_1 was found, for which again the size of the horizon (which however now forms only after a bounce at the boudary) shrinks to zero. Further decrease of the amplitude below this second critical value leads to a horizon formation after **two** reflections at the boundary until a third critical value ϵ_2 is reached. Keep decreasing ϵ, this picture repeats itself and the initial field profile collapses to a black hole after some number of reflections at the *AdS* boundary, Fig. (1.1). Another very interesting finding is that the time of the collapse coincides with the first nonlinear time scale of the problem ($t \sim \epsilon^{-2}$), Fig. (1.2).

Based on those observations, the authors of [8] conjectured that *arbitrarily small* perturbations will collapse after a number of bounces. In an attempt to give an analytic understanding to their findings, the authors suggested that black hole formation can be understood as a turbulent cascade of energy towards modes of higher frequency.

This picture is not far from the intuition one might have based on the AdS/CFT correspondence[2] and it was also corroborated by some follow up work [12, 14, 15]. However subsequent studies on the subject revealed a much richer dynamics. Many initial data were discovered that do not collapse and an inverse cascade towards modes of lower frequencies was also observed [4, 10, 11, 13]. Those results casted shadows on the initial conjecture and aroused the interest in the thermalisation

[1] When the standard, reflecting boundary conditions are imposed.

[2] It is generically expected that a strongly coupled field theory will thermalise if perturbed.

process on the boundary [4].

Most of this work was performed using Fourier space analysis and it was based on numerical methods, which are very powerful and can provide great intuition but at the same time are plagued with limitations. In stability considerations we ultimately want to address the limit where the perturbation is infinitesimally small. Such a limit is impossible to be captured by numerical methods and one has to resort to analytic approaches (perturbation theory) [3, 4, 8, 12].

Perturbative methods have been performed both in Fourier space [8, 11, 12], and in Position space [4] complementing each other. Both approaches result into an effective/truncated system of partial differential equations which possess an all important scaling symmetry. This scaling symmetry has been invoked to draw conclusions about the, otherwise inaccessible, limit of infinitesimal perturbations. The most important result is that, at this limit, the non-collapsing solutions form a set of non-zero measure [3], namely they are not rare, and hence AdS can not be said to be generically unstable. The fate of collapsing solutions in this limit is harder to address but studies in this direction suggest that they also form a set of non-zero measure [2].

Those results however are based on perturbation series which are truncated to the first nonlinear order and consequently are valid only up to the first nonlinear timescale. What happens at longer timescales is not yet known and definitely is worth further investigation.

Spherically symmetric perturbations, despite the fact they can lead to results of great importance, are far from being generic. To make generic statements regarding the stability of AdS one has to move beyond spherical symmetry, but little work has been made to date in this direction [14–19]. A very intriguing result of these studies is that in the non-spherically symmetric case, the perturbative scheme doesn't seem to possess the scaling symmetry of the spherically symmetric problem. However, this could be just due to the fact that the non-spherically symmetric model is far more complicated and a more sophisticated perturbative method might be required to faithfully represent the full system, especially when such a perturbation method is performed in Fourier space as in [14–19].

In the rest of this thesis we will describe in more detail the above story.

1.2 Spherically Symmetric Perturbations in AdS

In this section we will present the infalling spherically symmetric massless scalar field in AdS_{d+1} dressed with some technical details. The model is described by

the following action:

$$S = \int d^{d+1}x\sqrt{g}\left[\frac{1}{16\pi G_{(d+1)}}\left(R - 2\Lambda\right) + \frac{1}{2}\partial_\mu\phi\partial^\mu\phi\right], \tag{1.1}$$

with $\Lambda = -\frac{d(d-1)}{2R^2_{AdS}}$. Here, R_{AdS} denotes the radius of the spacetime, but in most of this thesis we will use units where $R_{AdS} = 1$ and $8\pi G = d - 1$. In AdS_{d+1} the resulting equations of motions are just the Einstein's equations with a stress energy tensor due to the scalar field and the wave equation for ϕ in the asysmptoticaly AdS background:

$$\begin{aligned} G_{\mu\nu} - \frac{d(d-1)}{2R^2_{AdS}}g_{\mu\nu} &= 8\pi G\left(\partial_\mu\partial_\nu\phi - \frac{1}{2}g_{\mu\nu}(\partial\phi)^2\right) \\ g^{\mu\nu}\nabla_\mu\phi\nabla_\nu\phi &= 0. \end{aligned} \tag{1.2}$$

This system of equations can be solved either numerically, or perturbatively. In the following we will present developments that have been achieved in both directions.

Asymptotically AdS spacetimes can be parametrised using the following ansatz:

$$ds^2 = \frac{R^2_{AdS}}{\cos^2 x}\left(A^{-1}dx^2 - Ae^{-2\delta}dt^2 + \sin^2 x d\Omega^2_{d-1}\right), \tag{1.3}$$

where the functions $A(t,x)$ and $\delta(t,x)$ as well as the scalar field $\phi(t,x)$ depend only on the radial coordinate x and time t due to spherical symmetry of the problem. Empty AdS corresponds to $A = 1$ and $\delta = 0$. One thing to notice here, that will also be discussed in more detail in chapter 3, is that there is still some gauge freedom in the above ansatz. This can be fixed be choosing a specific normalization for the function $\delta(t,x)$. The two most ubiquitous choices in the literature are:

$$\begin{aligned} \delta(t,0) &= 0, \quad t \text{ corresponds to the proper time at the center of the spacetime,} \\ \delta(t,\frac{\pi}{2}) &= 0, \quad t \text{ corresponds to the proper time at the boundary of the spacetime.} \end{aligned} \tag{1.4}$$

It is customary to introduce the auxiliary variables $\Phi = \phi'$ and $\Pi = A^{-1}e^{\delta}\dot{\phi}$, with the help of which the equations of motion read:

$$\dot{\Phi} = \left(Ae^{-\delta}\Pi\right)', \qquad \dot{\Pi} = \frac{1}{\mu}\left(\mu A e^{-\delta}\Phi\right)',$$
$$A' = \frac{\nu'}{\nu}(A-1) - \mu\nu\left(\Phi^2+\Pi^2\right)A, \qquad \delta' = -\mu\nu\left(\Phi^2+\Pi^2\right),$$
$$\dot{A} = -2\mu\nu A^2 e^{-\delta}\Phi\Pi. \tag{1.5}$$

For ease of notation we have adopted the convention $8\pi G = d-1$ and we have introduced the variables :

$$\mu(x) = (\tan x)^{d-1}, \qquad \nu(x) = \frac{1}{\mu'(x)} = \frac{\sin x \cos x}{(\tan x)^{d-1}}. \tag{1.6}$$

1.2.1 Numerical results

In several works, this system of equations has been solved numerically. In this section we will present these results.

In the seminal work of [8], the problem was studied in $3+1$ dimensions and for different classes of scalar field profiles. In figures (1.1) and (1.2) are presented the results for Gaussian initial profiles of the form:

$$\Phi(0,x) = 0, \qquad \Pi(0,x) = \frac{2\epsilon}{\pi}\exp\left(-\frac{4\tan^2 x}{\pi^2\sigma^2}\right). \tag{1.7}$$

The width was fixed at the value $\sigma = 1/16$ and the *amplitude* ϵ was varied. For large values of the amplitude ϵ, the wave packet quickly collapses, which is signalled by the formation of a horizon at x_H, determined by the vanishing of the function $A(t,x)$. As ϵ is further decreased, so does the horizon radius, until it becomes zero for some critical amplitude ϵ_0. When the amplitude is decreased further below the value of ϵ_0, a horizon will form, after the wave packet has reflected at the boundary once. Decreasing ϵ further, we find a second critical amplitude ϵ_1, for which $x_H = 0$. Further decrease of the amplitude below this value leads to a horizon formation after two reflections at the boundary until a third critical value ϵ_2 is reached. Keep decreasing ϵ this picture repeats itself and the initial field profile collapses to a black hole after some number of reflections at the AdS boundary, Fig. (1.1) . Based on those observations, the authors of [8] conjectured that *arbitrarily small* perturbations will collapse after a number of bounces.

A very important result of this study, is that the time of collapse scales as $t_H \sim \frac{1}{\epsilon^2}$, with the amplitude of the perturbation ϵ. This is depicted in Fig. (1.2), where

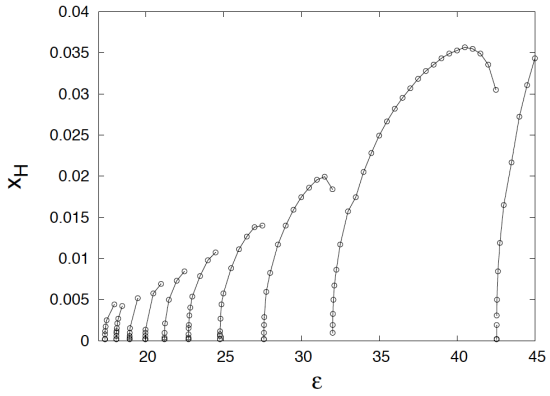


Figure 1.1: *The radius of the horizon x_H vs the amplitude of the perturbation ϵ for the gaussian data (1.7). For the given values of ϵ, the number of bounces at the boundary of AdS varies from 0 to 9 [8].*

the Ricci scalar at the origin $R(t,0) = -2\Pi^2(t,0)/R_{AdS}^2 - 12/R_{AdS}^2$ is presented[3]. This quantity actually oscillates and what is presented is only the upper envelope. As we will see in the next section this is the first nonlinear time scale, namely the earliest time scale where nonlinearities become important.

Another very interesting result of [8], that will also be explained in the next section, is that the *1-mode* data, namely scalar field profiles for which initially only one of the modes of the spectrum is excited, do not collapse

Further studies on the subject have revealed however a much richer and more interesting dynamics. More initial data were found for which their evolution does not lead to a black hole formation. For example, the authors of [10] studied numerically the same gaussian data, eq. (1.7), for different values of σ and they showed that for small values of σ this data indeed collapses, but for $\sigma \geq 0.4$ the collapse is avoided. This can be seen in Fig. (1.3) in which the time of the collapse vs the amplitude of the perturbation is depicted. For values of $\sigma \sim 0.5$ we see an abrupt change in t_c as ϵ decreases.

Based on those results, the authors of [10] conjectured that Gaussian Data with $\sigma \geq 0.4$ are stable. However, later studies reported that the collapse is restored for values $\sigma \sim 8$ [11]. In [10] a *complex* scalar field, minimally coupled to gravity, was studied and stable solutions were found in that system as well, dubbed as *Boson Stars*.

In [20] *time periodic* solutions were constructed in the Einstein *AdS*-massless scalar

[3]This quantity consists of a good indicator for the onset of the instability

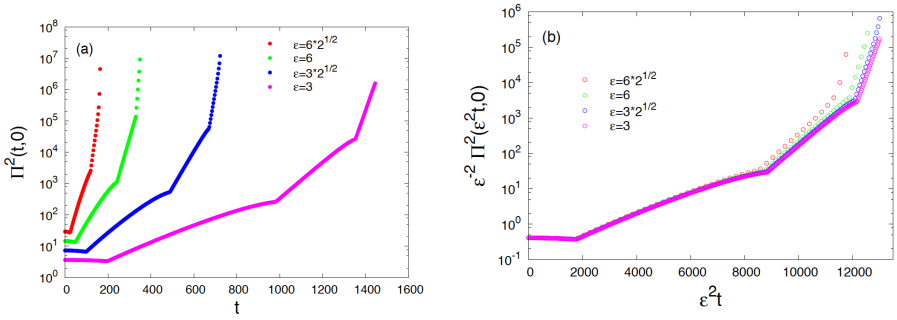


Figure 1.2: *(a) The time evolution of the upper envelope of* $\Pi^2(t,0)$ *for four different values of* ϵ *for the gaussian data (1.7) and (b) the evidence that the time of the collapse scales as* ϵ^{-2} *[8].*

field model for any $d \geq 2$, whereas in [14,15] similar solutions (*geons*) where found in pure gravity. Continuing in the direction of finding stable solutions, the work of [12, 13] was also very fruitful. In those works a class of *quasi-periodic* solutions was constructed and it was shown to be nonlinearly stable. Those solutions are finely tuned, such that the net energy flow in each mode is zero. The ansatz for the amplitudes of those solutions is of the form[4] $\alpha_j(\tau) = A_j e^{-iB_j\tau}$.

Summarizing, the above results lead to a very interesting picture of the phase space of initial perturbations of AdS, which is entirely different than the corresponding phase space of asymptotically flat spacetimes. In AdS there seems to be a part of phase space which is unstable and a part for which there is a threshold ϵ_0, below of which there is no black hole formation.

1.2.2 Weakly nonlinear perturbation theory

One can try and solve this system analytically in perturbation theory, expanding the fields in powers of the amplitude ϵ of the initial perturbation

$$\phi = \sum_{j=0}^{\infty} \epsilon^{2j+1}\phi_{2j+1}(t,x), \quad A = 1 - \sum_{j=0}^{\infty} \epsilon^{2j} A_{2j}(t,x), \quad \delta = \sum_{j=0}^{\infty} \epsilon^{2j}\delta_{2j}(t,x). \tag{1.8}$$

Inserting this ansatz into the equations of motion and collecting terms of the same order of ϵ we obtain a set of linear equations which can be solved order by order. To first order, we merely have a scalar filed propagating in the AdS background

$$\ddot{\phi}_1 + L\phi_1 = 0. \tag{1.9}$$

[4] See section (1.2.2) for more details.

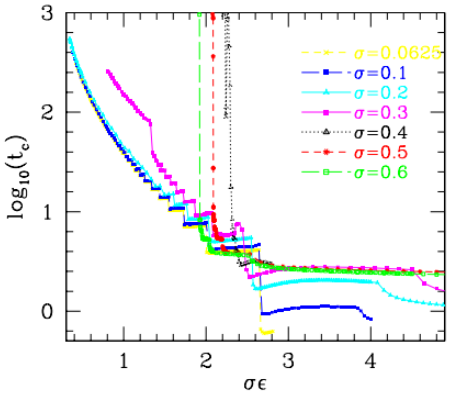


Figure 1.3: *The collapse time of various initial data of the form (1.7) vs the rescaled amplitude $\sigma\epsilon$ [10].*

Here, $L = -\frac{1}{\mu(x)}\partial_x\left(\mu(x)\partial_x\right)$ is the Laplacian of AdS with eigenvalues $\omega_j = (2j+d)$ and eigenfunctions

$$e_j = d_j \cos^d x P_j^{\frac{d}{2}-1,\frac{d}{2}}, \qquad d_j = \frac{2\sqrt{j!(j+d-1)}}{\Gamma(j+\frac{d}{2})}. \tag{1.10}$$

The eigenfunctions are normalized such that $(e_m, e_n) = \delta_{mn}$ in the inner product $(f,g) := \int_0^{\frac{\pi}{2}} f(x), g(x)\mu(x)dx$ and the positivity of the eigenvalues implies that AdS is linearly stable. One can use the eigenfunctions of L to expand the field ϕ, where at this order is simply:

$$\phi_1(t,x) = \sum_{j=0}^{\infty} c_j^{(1)}(t) e_j(x) = \sum_{j=0}^{\infty} \left(\alpha_j e^{i\omega_j t} + \bar{\alpha}_j e^{-i\omega_j t}\right) e_j(x). \tag{1.11}$$

To second order in ϵ we obtain the back reaction on the metric

$$A_2(t,x) = -\nu(x)\int_0^x \left(\dot{\phi}_1(t,y)^2 + \phi'(t,y)^2\right)\mu(y)dy, \tag{1.12}$$

$$\delta_2(t,x) = \begin{cases} -\int_0^x \left(\dot{\phi}_1(t,y)^2 + \phi'(t,y)^2\right)\nu(y)\mu(y)dy, & \text{for } \delta(t,0) = 0 \\ \int_x^{\pi/2} \left(\dot{\phi}_1(t,y)^2 + \phi'(t,y)^2\right)\nu(y)\mu(y)dy, & \text{for } \delta(t,\pi/2) = 0. \end{cases} \tag{1.13}$$

while the first non–trivial dynamics appear at the third order. One obtains the inhomogeneous equation

$$\ddot{\phi}_3 + L\phi_3 = S\left(\phi_1, A_2, \delta_2\right), \tag{1.14}$$

where $S = -2\ddot{\phi}_1\left(A_2+\delta_2\right) - \dot{\phi}_1\left(\dot{A}_2 + \dot{\delta}_2\right) - \phi_1'\left(A_2' + \delta_2'\right)$. We can again expand the field as $\phi_3(t,x) = \sum_{j=0}^{\infty} c_j^{(3)}(t) e_j(x)$, and projecting o the basis $\{e_m\}$, we obtain

an infinite set of decoupled forced harmonic oscillators for the Fourier coefficients

$$\ddot{c}_j^{(3)}(t) + \omega_j^2 c_j^{(3)}(t) = S_j, \quad S_j \equiv (S, e_j)\,. \tag{1.15}$$

Due to the fact that the linear spectrum is highly commensurate, since the frequencies are integers, there can be many resonant terms contained in S_j. Specifically, for every triad (j_1, j_2, j_3)[5] such that $\omega_j = \omega_{j_1} + \omega_{j_2} - \omega_{j_3}$ there is a resonant term in S_j.

Let us digress a bit here and describe the issue of resonances in time dependent perturbation theory in a slightly simpler system, the *anharmonic oscillator*. The exact equation of motion is:

$$\ddot{q} + q + \epsilon q^3 = 0, \tag{1.16}$$

where ϵ is a measure of the strength of nonlinearity. Let us assume the following, simple, initial conditions

$$q(0) = C, \qquad \dot{q}(0) = 0. \tag{1.17}$$

We seek a solution in the form of perturbative series:

$$q(t) = q_0(t) + \epsilon q_1(t) + \epsilon^2 q_2(t) + \ldots = \sum_{i=0}^{\infty} \epsilon^i q_i(t). \tag{1.18}$$

By plugging this ansatz into the equation of motion for q and equating terms of the same power in ϵ, we obtain the following (infinite) system of coupled differential equations:

$$\begin{aligned} \ddot{q_0} + q_0 &= 0 \\ \ddot{q_1} + q_1 &= -q_0^3 \\ \ddot{q_2} + q_2 &= -3q_0^2 q_1 \\ &\ldots \end{aligned} \tag{1.19}$$

This set of equations has the property that if we know $q_0(t)$ we can solve for $q_1(t)$, then we can find $q_2(t)$ and so on so forth. Therefore we can proceed order by order to find a power series expansion of the solution. In practice, one has to truncate this perturbative expansion to some order.

[5] S_j is cubic in a_j at this order.

With the given initial conditions, the solution to the $\mathcal{O}(\epsilon^0)$ term is simply $q_0(t) = C\cos t$. Then we can use this result to solve for $q_1(t)$. In the second equation of the system (1.19), $q_0(t)$ acts a (periodic) driving force for q_1. One can solve this equation by applying the *Green's function method*, which for the case of the harmonic oscillaltor yields

$$\begin{aligned} q_1(t) &= \int_0^t \sin(t-t')q_0(t')dt' \Rightarrow \\ q_1(t) &= C^3\left(-\frac{3}{8}t\sin t - \frac{1}{32}\cos t + \frac{1}{32}\cos 3t\right). \end{aligned} \tag{1.20}$$

However, the solution (1.20) is wrong! We know that the motion is periodic, yet the solution we obtained diverges linearly with time (the first term in (1.20)). Another issue is that the energy is not conserved for this solution. The problem arises because the *driving* term q_0^3 contains a term $\sim \cos t$, which acts as a *resonant* driving term leading to non-periodic solutions in q_1. Such diverging terms are often dubbed *secular* terms.

A way out of this conundrum was found by Lindstedt and Poincare as they realised that one is attempting an expansion of a function with varying period in terms of functions with fixed periodicity 2π. Their proposal was to allow for a change of the independent variable simultaneously with the perturbative expansion. This would allow for the period to change with the amplitude. To be more precise let us define the *new* independent variable s:

$$\begin{aligned} s &\equiv \omega t, \\ \omega &\equiv 1 + \epsilon\omega_1\epsilon^2\omega_2 + \dots. \end{aligned} \tag{1.21}$$

Then we have the power series expansion

$$q(s) = \sum_{i=0}^{\infty} \epsilon^i q_i(s). \tag{1.22}$$

The trick of this method is that due these extra constants ω_i we can remove order by order the secular terms. By changing independent variable $t \to s$, the equation of motion for q reads:

$$\omega^2\ddot{q} + q + \epsilon q^3 = 0, \qquad \dot{q} \equiv \frac{dq}{ds}. \tag{1.23}$$

Plugging now both the series expansions (1.21) and (1.22) and equating again the coefficients of the same powers of ϵ we obtain the following system of coupled equations:

$$\begin{aligned} \ddot{q}_0 + q_0 &= 0 \\ \ddot{q}_1 + q_1 &= -q_0^3 + 2q_0\omega_1 \\ \ddot{q}_2 + q_2 &= -3q_0^2 q_1 + 2(q_1 + q_0^3)\omega_1 + q_0(2\omega_2 - 3\omega_1^2) \\ &\dots \end{aligned} \tag{1.24}$$

The solution to the first equation is, as before, $q_0(s) = C\cos s$ and inserting this into the second one we obtain for q_1:

$$\ddot{q}_1 + q_1 = \left(2C\omega_1 - \frac{3}{4}C^3\right)\cos s - \frac{C^3}{4}\cos 3s. \tag{1.25}$$

No it is obvious that we can choose $\omega_1 = \frac{3}{8}C^2$ and cancel the secular term $\sim \cos s$! This would lead to the solution:

$$\begin{aligned} q_1 &= -\frac{C^3}{32}(\cos s - \cos 3s) \\ \omega &= 1 + \epsilon\frac{3}{8}C^2. \end{aligned} \tag{1.26}$$

There are no *secular* terms now, but the frequency has been shifted by an amount that depends on the amplitude of the nonlinearity ϵ.

This procedure could be seen a bit differently as well. Let us go back to eq. (1.20) for q_1. The solution for $q(t)$ (up to first order) would thus be:

$$\begin{aligned} q(t) &= q_0(t) + \epsilon q_1(t) = C\cos t + \epsilon C^3\left(-\frac{3}{8}t\sin t - \frac{1}{32}\cos t + \frac{1}{32}\cos 3t\right) \\ &= C\cos\left[\left(1 + \frac{3}{8}\epsilon C^2\right)t\right] - \epsilon\frac{C^3}{32}\left[\cos\left(\left(1 + \frac{3}{8}C^2\right)t\right) - \cos\left(3\left(1 + \frac{3}{8}C^2\right)t\right)\right] \end{aligned} \tag{1.27}$$

In the second line we have realised and resummed the Taylor expansion (to first order in ϵ):

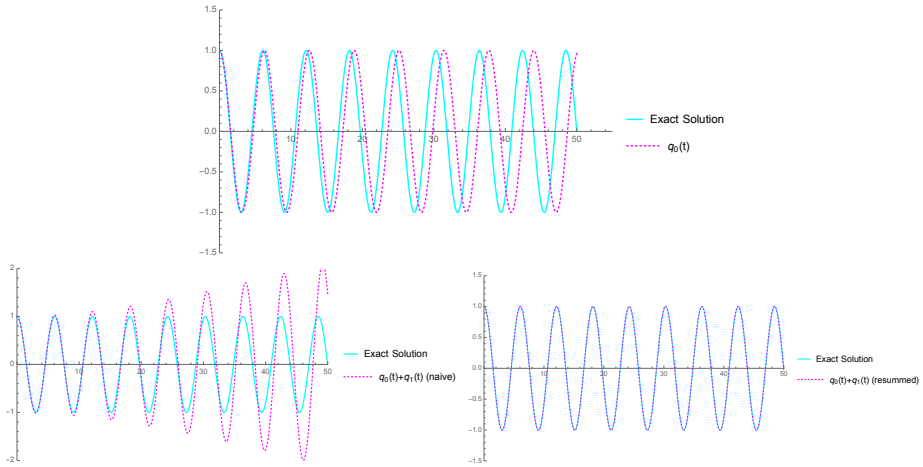


Figure 1.4: *Comparison of the* exact *solution to eq. (1.19), with* $\epsilon = 0.1$ *and* $C = 1$*, obtained numerically, with the different solutions obtained from perturbation theory.* **Up:** *The zeroth order perturbative solution.* **Down***: The perturbative solution up to first order,* naive *(***left***) and* resummed *(***right***).*

$$\cos(t + \epsilon\frac{3}{8}C^2 t) \simeq \cos t - \frac{3}{8}\epsilon C^2 t \sin t. \tag{1.28}$$

To illustrate the success of this resummed perturbation theory we contrast in Fig. (1.4) the different approximate solutions to eq. (1.16), the *exact* solution obtained by numerical methods.

The case of *AdS* is not very different from the toy model of the anharmonic oscillator, only a bit more intricate. A first treatment of the *AdS* resonances appeared in [8] and a more sophisticated technique, the Two Time Framework (TTF) [12, 21], was later developed and allows for a systematic procedure. The basic idea behind TTF, is to allow for an additional, *slow time* dependence, of the fields. In that case, the perturbative expansion of the fields, would become

$$\begin{aligned} \phi &= \epsilon\phi_1(t,\tau,x) + \epsilon^3\phi_3(t,\tau,x) + \mathcal{O}(\epsilon^5), \\ A &= 1 - \epsilon^2 A_2(t,\tau,x) + \mathcal{O}(\epsilon^4), \\ \delta_2 &= \epsilon^2\delta_2(t,\tau,x) + \mathcal{O}(\epsilon^4), \end{aligned} \tag{1.29}$$

where $\tau = \epsilon^2 t$ denotes the slow time. One could go to higher orders by introducing additional slow time variables. Substituting now the above expansion, eq. (1.29), into the equations of motion, we obtain again a system of equations which can be solved order by order in ϵ. The difference now, is that the expansion for the

linearized field can be written as

$$\phi_1(t,\tau,x) = \sum_{j=0}^{\infty} \left(\alpha_j(\tau)e^{-i\omega_j t} + \bar{\alpha}_j(\tau)e^{-i\omega_j t}\right) e_j(x), \tag{1.30}$$

where the slow time τ accounts for the energy transfer among the normal modes (due to non–linearities), while the fast time t accounts for the oscillations of the normal modes. To second order we have again the constraints for A_2 and δ_2 and the first non-trivial dynamics appear to third order in ϵ

$$\partial_t\phi_3 + L\phi_3 + 2\partial_t\partial_\tau\phi_1 = S(\phi_1, A_2, \delta_2), \tag{1.31}$$

where the source term reads $S(\phi_1, A_2, \delta_2) = (\dot{A}_2 - \dot{\delta}_2)\dot{\phi}_1 - 2(A_2 - \delta_2)L\phi_1 + (A_2' - \delta_2')\phi_1'$. The overdots here, represent derivatives with respect to the fast time t, as usual. Projecting again on the basis $\{e_m\}$, and substituting eq. (1.30) we obtain

$$\left(e_j, \partial_t^2\phi_3 + \omega_j^2\phi_3\right) - 2i\omega_j\left(\partial_\tau\alpha_j(\tau)e^{-i\omega_j t} - \partial_\tau\bar{\alpha}_j(\tau)e^{i\omega_j t}\right) = (e_j, S)\,. \tag{1.32}$$

Now we can treat the resonances separately, due to the presence of terms proportional to $e^{\pm i\omega_j t}$ on the left hand side. Setting

$$-2i\omega_j\partial_\tau\alpha_j(\tau) = (e_j, S)\,[\omega_j] = \sum_{klm} S_{jklm}\bar{\alpha}_k\alpha_l\alpha_m, \tag{1.33}$$

we may cancel off the resonant terms from the rest of the eq. (1.31). These are now entirely captured by eq. (1.33), rendering the solution ϕ_3 bounded, and hence of little interest to us. The result is an infinite set of coupled first order ODEs for the Fourier coefficients of the linear expansion, which are known by the name TTF *equations*. The sum, in eq. (1.33), contains only terms for which the *resonance condition* $j + k = l + m$ is satisfied.[6] The TTF equations govern the evolution of both the amplitudes and the phases of the complex coefficients a_j and sometimes can be more convenient to rewrite them using the *amplitude–phase representation* $\alpha_j(\tau) = A_j(\tau)e^{iB_j(\tau)}$, in terms of which the TTF equations can be expressed as

$$\begin{aligned} 2\omega_j\frac{dA_j}{d\tau} &= \sum_{\substack{j+k=l+m \\ j\neq l, j\neq m}} S_{jklm}A_kA_lA_m\,\sin\left(B_j + B_k - B_l - B_m\right) \\ 2\omega_j\frac{dB_j}{d\tau} &= T_jA_j^2 + \sum_{i\neq j} R_{ij}A_i^2 + A_j^{-1}\sum_{\substack{j+k=l+m \\ j\neq l, j\neq m}} S_{jklm}A_kA_lA_m\,\cos\left(B_j + B_k - B_l - B_m\right). \end{aligned} \tag{1.34}$$

[6] In principle, all the resonant *channels*, $j = \pm k \pm j \pm l$, should appear but only the one described here remains [21, 22].

Here $T_j = S_{jjjj}$, $R_{ji} = S_{ijji} + S_{jiji}$, for $i \neq j$. The TTF system possesses an all important scaling symmetry, $\alpha_j(\tau) \to \epsilon\alpha_j(\tau/\epsilon^2)$, which means that the evolution of a perturbation of amplitude A_j for time τ, will be the same as the evolution of the perturbation of amplitude ϵA_j, but for longer time τ/ϵ^2. As mentioned earlier, this allows us, for as long as the TTF approximations are valid [3], to draw conclusions for the vanishing amplitude limit, from results obtained in finite ϵ.

This system of equations has been extensively studied, and will be of great interest in the subsequent chapters, revealing a very interesting phenomenology with a phase space of initial conditions containing both finite sets of stable and unstable perturbations. However, since this is only an approximation to the full system, many questions naturally arise. For example, up to what point are we allowed to trust our approximations, what happens in the evolution of the perturbations after our approximate methods are invalidated and can we remove all the resonances by similar multi-scale methods?

2

Position space analysis of the AdS (in)stability problem

Nothing happens until something moves.

— Albert Einstein

In the previous chapter we analyzed the problem of the evolution of perturbations in *AdS* in *Fourier space*. Now we will present an analysis directly in *position space* that nicely complements the momentum space analysis.

2.1 Introduction

As we mentioned earlier, given a spherically symmetric perturbation of arbitrarily small initial amplitude ϵ, two dramatically different behaviors have been observed at the timescale $\sim \epsilon^{-2}$, the earliest time on which interactions can have a significant effect [8, 10–12, 15, 16, 20, 23–28]. Sometimes a black hole forms around this time; sometimes a long-lived quasi-periodic behavior emerges and gravity does not become strong. This is a great puzzle concerning both the gravitational dynamics in the bulk and the corresponding thermalization process in the holographic boundary theory.

In this chapter we will focus on the bulk perspective and on the simple case of a free massless scalar field coupled to gravity. We treat the system classically and impose spherical symmetry. In the limit of small amplitude ϵ, the energy density is proportional to ϵ^2 and controls the strength of gravitational effects. Therefore, behavior at the time scale ϵ^{-2} is sensitive to the leading-order effects of gravitational interactions. One framework to study this is to analyze the nonlinear couplings between the linearized modes induced by the gravitational interactions. Linearized modes in AdS space all have frequencies which are integer multiples

of the AdS scale. A mode that is initially unexcited can be resonantly driven by the excited modes, which allows for the possibility of efficient transfer of energy. Such efficient energy transfer between modes generically leads to the breakdown of naïve perturbation theory, since the true solution does not remain close to the solution in the non-interacting theory. This resonance effect was argued to be the cause of an energy cascade—energy spreads out into more and higher modes—in order to explain black hole formation and the power-law spectrum observed during such processes [8, 14, 15, 23]. It was also argued that since the AdS spectrum is resonant, such an instability should be the generic outcome of small perturbations.

Counter-examples to the above claim in the form of the stable, quasi-periodic solutions were initially viewed as being special. It was conjectured in [15] that these stable solutions will shrink to a set of measure zero in the small ϵ limit, and the term "stability island" was used to describe their existence in the generically unstable sea of phase space. However, more recent evidence suggests that such a conclusion is too strong. Numerical evidence suggests that, at finite ϵ, the stable and unstable solutions both have nonzero measure in the space of initial conditions [10–12].[1] One can then apply a simple scaling argument, described in more detail in Sec. 2.2, to show that in the $\epsilon \to 0$ limit, the stable solutions persist. However, the same argument fails for unstable solutions. The open question now becomes whether there are "instability corners." Namely, in the $\epsilon \to 0$ limit, do the unstable solutions shrink to a set of measure zero, or do they also continue to have finite measure?

There are some important misconceptions and misunderstandings in the current literature regarding the status of the AdS (in)stability problem, due in part to three points of confusion, which we would like to clarify here. First of all, an energy cascade is not identical to, nor does it guarantee black hole formation. This distinction has not been made clear enough. Both have been frequently used interchangeably and referred to as the "instability of AdS space." Black hole formation requires energy to be focused into a small spatial region. According to the uncertainty principle, energy flowing to high momentum is certainly a necessary condition for that, but it is not sufficient. It is entirely possible for even unboundedly high momentum modes to be populated, but for the energy distribution to stay roughly spatially homogeneous.

Therefore, here, **AdS instability** strictly refers to **black hole formation** only. Because the AdS geometry changes dramatically in this case, such nomenclature

[1] Note that we are always discussing stability on the interaction time scale ϵ^{-2} in this work. The question of the behavior on longer time scales is a fascinating one that touches on issues of ergodicity, Arnold diffusion, and the KAM theorem. We do not know how to attack questions on these longer time scales analytically or numerically.

aligns with a more gravity point of view.[2] This also allows us to study its implications on the boundary CFT. When we refer to a solution or initial condition as stable or unstable, we will always be indicating whether it collapses to form black hole or not.

The second point of confusion is the use of term "generic." Numerical evidence suggests that, at finite ϵ, the stable and unstable solutions both form sets of nonzero measure in the space of initial conditions [10–12].[3] We are interested in the $\epsilon \to 0$ limit, and in this work we use the following definition:

1. "Generic instability" means the set of stable initial conditions (not forming black holes) shrinks to measure zero.
2. "Generic stability" means the set of unstable initial conditions (forming black holes) shrinks to measure zero.
3. "Mixed" means that both sets have nonzero measure as $\epsilon \to 0$.

Until recently, references in the literature did not clearly distinguish between (1) and (3). For example, it was conjectured in [15] that "stability islands" shrink to a set of measure zero in the small-ϵ limit, which is certainly arguing for only (1). However, the numerical evidence in [8] showing that black holes continue to form as ϵ is reduced is consistent with both (1) and (3). Since these are three physically different cases, we think such a clear distinction is needed.

Finally, when addressing the question of instability, one needs to specify a time scale. In this work, we will only discuss the time scale that goes to infinity as ϵ^{-2} in the $\epsilon \to 0$ limit.[4] Indeed a naïve perturbation analysis shows that something interesting can happen at this time scale. The physical question we will address is whether that "something interesting" is generically black hole formation? In the end, we will try to relate the answer to the boundary CFT: Does the boundary system thermalize at this time scale?

[2] From the hydrodynamic point of view, the existence of an energy cascade might be a suitable definition of instability. Indeed, this is the perspective taken by some authors, and we wish the reader to see the distinction clearly.

[3] Strictly speaking, numerical results only cover discrete choices of initial conditions. So, it is therefore impossible on numerical grounds alone to prove that any such set has either zero or nonzero measure. This fact holds equally for both stable and unstable solutions. Nevertheless, if either set really were measure-zero, unless the numerical code secretly enforced extra symmetries, it would be extremely unlikely to find such a result even once in simulations. Thus, despite the numerical controversy over some of the stable solutions [29], we still interpret the current evidence that stable and unstable solutions both have nonzero measures.

[4] Behaviors at shorter time scales are somewhat trivial. For example, for a given fixed time, black hole forming solutions disappear as $\epsilon \to 0$, so the system is generically stable, case (2). The behavior at longer time scales is a very deep problem that touches on issues of ergodicity, Arnold diffusion, and the KAM theorem. We do not know how to attack those questions analytically or numerically.

After making all these definitions clear, in Sec. 2.2 we briefly review the recent progress on this topic. We then present a very simple scaling argument which shows that possibility (1) defined above, "generic instability", is the most unlikely given by existing evidence. This directly argues against the "stability island" conjecture [15]. The remaining question then is whether AdS space is generically stable (2) or mixed (3).

In Sec. 2.3 we set up our perturbative method for studying gravitational self-interaction. This position-space approach is more directly relevant than the usual momentum-space analysis to the question of whether or not black holes form.[5] If energy gets focused into a smaller region, then the solution is evolving toward a black hole. If energy is defocused into a larger region, then the solution is evolving away from a black hole. We explicitly demonstrate that in the $\epsilon \to 0$ limit, the focusing/defocusing dynamics depend only on the gravitational self-interaction near the origin of AdS, when the energy of the perturbation is maximally concentrated. The propagation through the rest of asymptotic AdS space plays no dynamical role.

In Sec. 2.4 we prove a one-to-one correspondence between focusing and defocusing energy in the near-center dynamics. Heuristically, our result is shown in Fig. (2.1): A shell of massless scalar field will become narrower, its energy focused, if it is denser in the front. On the other hand, if it is denser in the tail, it will become wider and energy will defocus.[6] More generally, the leading-order dynamics of focusing and defocusing are related by time reversal, so a local maximum of energy density is also equally likely to grow or diminish within the time scale $\lesssim \epsilon^{-2}$.

In Sec. 2.5 we present the new intuition our method provides and propose a conjecture on the structure of the phase space. Based on the symmetry between focusing and defocusing dynamics, the stable, quasi-periodic solutions can be understood as trajectories that alternate between the two. As a result, they may form quasi-closed loops in phase space. In fact, some unstable solutions are also known to exhibit this alternating behavior while in the weak-gravity regime. Based on this understanding of the dynamics, we propose a conjecture on how to visualize the phase space of small perturbations in AdS space.

We also discuss how these gravitational calculations can shed light on the general concepts of thermalization in a closed system. In particular, contrary to conven-

[5] In principle, one can include all the information about relative phases in the spectral analysis to achieve the same result. Our position-space approach is simply more direct. In addition, it technically circumvents the subtlety that the gravitational interaction imposes significant additional constraints on possible resonances [21].

[6] The shell profile will change in other ways, but all changes are suppressed by ϵ^2. The focusing or defocusing behavior will last for a time scale $\lesssim \epsilon^{-2}$, so it is the dynamics we are interested in here.

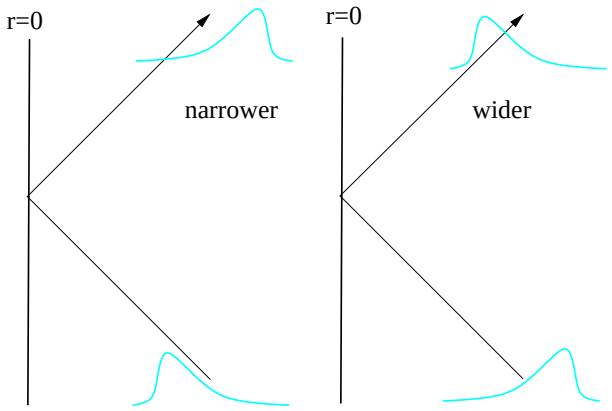


Figure 2.1: *A thin shell that has higher energy density in its front will come out narrower after gravitational self-interactions as it bounces through $r = 0$. A shell with higher energy density in its tail will come out wider after the bounce.*

tional wisdom, black hole formation at the ϵ^{-2} time scale is not necessarily the holographic dual of thermalization in the boundary field theory. If the thermal gas phase is the equilibrium state, then black hole formation describes prethermalization that significantly delays true thermalization [30, 31].

In Sec. 2.6, we provide a quick summary of six major points of this chapter. In Appendices 2.A and 2.B, we provide the computational details of our method and numerical examples to demonstrate how the shape of the profile determines whether its energy is focused or defocused.

2.2 Stability islands or instability corners?

We are interested in the perturbative stability of global AdS space. We will work in $(3+1)$ dimensions and employ the following metric for vacuum AdS_4:

$$ds^2_{AdS_4} = -\left(1+\frac{r^2}{R^2_{\rm AdS}}\right)dt^2 + \frac{dr^2}{1+\frac{r^2}{R^2_{\rm AdS}}} + r^2 d\Omega_2^2 \tag{2.1}$$

where $R_{\rm AdS}$ is the AdS radius and $d\Omega_2^2$ is the round metric on S^2.[7] Our perturbations will take the form of a real, massless scalar field ϕ minimally coupled to Einstein gravity with a negative cosmological constant:

$$S = \int d^4x\,\sqrt{g}\left(\frac{1}{16\pi}R + \frac{6}{R^2_{\rm AdS}} - \frac{1}{2}\partial_\mu\phi\partial^\mu\phi\right)\ . \tag{2.2}$$

[7] Our radial coordinate r is related to the radial coordinate x used in [8] by $r = \tan x$.

The Planck scale has been set to one. We will consider only spherically symmetric solutions, so both the scalar ϕ and the metric functions g_{tt} and g_{rr} will only depend on t and r.

Here we will review some existing evidence and argue that a careful interpretation strongly supports the following conclusion for spherically symmetric perturbations of a massless scalar field in AdS space:

In the $\epsilon \to 0$ limit, at the $T \sim \epsilon^{-2}$ time scale, AdS space is either generically stable, or stable and unstable perturbations are equally generic.

The first part of our argument is based on ample numerical evidence at small but finite ϵ. The initial conditions that lead to black hole formation (unstable) and those that lead to quasi-periodic solutions (stable) both form open sets in the phase space of nonzero measure. Note that the phase space is infinite dimensional, so no numerical evidence can prove that any set is really open. Nevertheless, whatever extrapolations are being made should be applied equally to both stable and unstable solutions, and the existing numerical evidence is quite sufficient to show that they are on equal footing. More specifically, numerical tests can scan a one-parameter family of initial conditions, corresponding to a line in phase space. It has been clearly demonstrated that for a few such lines, the initial conditions that lead to stable and unstable solutions both form finite segments [10–12]. We will pragmatically take this as evidence that both stable and unstable sets in phase space have nonzero measure at small but finite ϵ.

In particular, within the set of stable solutions, one can identify a subset for which "gravity never becomes strong" during the $\sim \epsilon^{-2}$ time scale; that is,

$$\exists\ \phi(\epsilon, r, t)\ ,\ \text{such that}\ \left(\dot{\phi}^2 + \phi'^2\right) < \delta \ll 1 \text{ for } 0 \le t \le T \sim \epsilon^{-2}. \tag{2.3}$$

Our next step is to show that in the $\epsilon \to 0$ limit, these stable solutions cannot disappear. We can use the scaling behavior observed in [8, 12], which was trustworthy to leading order in ϵ. We will demonstrate that in the $\epsilon \to 0$ limit, this scaling behavior is exact for stable, weak-gravity solutions.

The spectrum of a massless field in the AdS background is given by integer multiples of the AdS energy scale $R_{\rm AdS}^{-1}$, meaning that the field profile is exactly periodic in time. Heuristically, a spherical wavefront shrinks toward the origin $r = 0$, passes through it, expands again to infinity, and finally bounces off the boundary back to the original position.[8] It is natural to describe the dynamics as a function of

[8]The periodicity of geodesics in AdS is $2\pi R_{\rm AdS}$, and in that time they pass through the origin twice. However, a shell of massless scalar field with Dirichlet boundary conditions at

the "number of bounces" $N = \frac{t}{\pi R_{\rm AdS}}$ instead of the microscopic time t:

$$\phi(r, N+1) \equiv \phi(r, t + \pi R_{\rm AdS}) = \phi(r, N) \equiv \phi(r, t) \ . \tag{2.4}$$

Now, introducing gravitational self-interaction, as long as the field amplitude (and therefore the resulting back-reaction) is small, we have a small correction to the above exactly periodic solution,

$$\phi(r, N+1) - \phi(r, N) = A[\phi, \dot{\phi}] + \mathcal{O}(\phi^5) \ . \tag{2.5}$$

The functional A describes the small, leading-order changes to the profile, which we will analyze further in the following sections. Here we only need to know that it scales like ϕ^3. It is convenient to introduce the rescaled field, $\bar{\phi} \equiv \phi/\epsilon$, whose evolution is given by

$$\bar{\phi}(r, N+1) - \bar{\phi}(r, N) = A[\bar{\phi}, \dot{\bar{\phi}}]\epsilon^2 + \mathcal{O}(\epsilon^4) \ , \tag{2.6}$$

Although the value of N is discrete, in the $\epsilon \to 0$ limit, the change due to each bounce goes to zero. We can therefore take the continuum limit, in which eq. (2.6) becomes

$$\frac{d\bar{\phi}}{d(\epsilon^2 N)} = A \ . \tag{2.7}$$

Thus, the scaling behavior is exact:

$$\bar{\phi}_\epsilon(r, N) = \bar{\phi}_{\frac{\epsilon}{\alpha}}(r, \alpha^2 N) \ . \tag{2.8}$$

Reducing the amplitude of the fluctuation simply slows down the dynamics by α^2: if ϵ is reduced by a factor of α, it takes α^2 more bounces to reach the same configuration. Therefore, if there is a stable solution at some finite ϵ and for a time $T \sim \epsilon^{-2}$ during which gravity never becomes strong, this must also be a stable solution at any smaller ϵ, all the way to the $\epsilon \to 0$ limit.[9]

Interestingly, this same argument is not applicable to unstable solutions. In order to form a black hole, the scalar field profile must first evolve to have large energy density somewhere, $\left(\dot{\phi}^2 + \phi'^2\right) \sim 1$. In other words, gravity must become strong, at which point the higher order terms in eq. (2.6) cannot be ignored. In those cases the scaling behavior is lost. A collapsing solution at some small but finite ϵ might escape that fate if we reduce ϵ further [34].

At this point, we are left with two possibilities:

the boundary is actually periodic in half that time, $\pi R_{\rm AdS}$, during which the wavefront passes through the origin only once.

[9] We should note that the expansion in powers of ϵ is most likely asymptotic [32], but its leading-order result has been accurate for many similar applications [33].

- Stable and unstable perturbations are both generic, in that they both occupy sets of nonzero measure in the phase space.
- AdS space is perturbatively stable generically, but there are special "instability corners", which shrink to measure zero in the limit $\epsilon \to 0$.

Finally, recall that we have so far limited ourselves to spherical symmetry. Intuitively, spherical symmetry arranges for matter to converge at the origin, which is helpful for gravitational collapse. So, even if the first of the above possibilities holds within spherical symmetry, it may be that without spherical symmetry the second is instead the case.

2.3 Weak gravitational self-interaction in position space

2.3.1 The two-region approximation

We now present our approach to explicitly calculating the functional A in eq. (2.5). Our result, a precise expression for A, is given in eq. (2.29). Many of its properties will help us to better understand the dynamics and the possibility of instability corners. Our calculation will be in position space. The advantage for this approach is easily seen if we first picture the evolution of a thin shell of total energy $E \sim \epsilon^2$, thickness w and initial size r_0, such that $r_0 \gg w$. This corresponds to an initial field profile that is roughly given by

$$\phi_0(r,t)|_{t\sim t_i} \sim \frac{-\epsilon\sqrt{w}}{r}\ f\left(-\frac{r-r_0+t-t_i}{w}\right)\ . \tag{2.9}$$

We will take the profile $f(x)$ to be some function that peaks at $x = 0$ and has compact support an order-one range around around this peak (*i.e.* $f(x) = 0$ for $|x| \gtrsim 1$).[10] Note that we have carefully chosen the dependence on w such that it does not affect the total mass, which is controlled solely by ϵ. The small-perturbation limit then corresponds to $\epsilon \to 0$.

Other works studying similar scenarios choose various different initial conditions for the scalar field perturbation. Some authors take initial conditions that place the energy near $r = 0$. In other cases, the perturbation originates from a quench in the boundary CFT and appears as a wavefront coming in from $r = \infty$ [35].

[10] We choose the profile to have compact support only to make the subsequent calculations somewhat cleaner. The shell only needs a narrow, well-defined width. Alternately, one could take f to have, for example, Gaussian tails without affecting the results.

Remember that in the small-ϵ limit, the leading-order behavior is the same as in empty AdS space; that is, the radiation shell simply bounces back and forth between $r = 0$ and $r = \infty$. Therefore, all of these initial positions of the shell are related by a shift in time on the order of R_{AdS}. Since we are interested in the outcome at longer time scales, they are all equivalent for our purposes.

One advantage of our position-space approach is that we can choose an r_0 which implements the following "two-region" approximation:

- For $r < r_0$, we will ignore that the background is AdS space and consider only the back-reaction of the scalar field on Minkowski space.
- For $r > r_0$, we will ignore the scalar field back-reaction and treat the geometry as empty AdS space.

In order to justify this simplification, we first recall the general form of the Schwarzschild-AdS metric:

$$ds^2_{SAdS_4} = -\left(1 - \frac{2M(r)}{r} + \frac{r^2}{R^2_{\text{AdS}}}\right) dt^2 + \frac{dr^2}{1 - \frac{2M(r)}{r} + \frac{r^2}{R^2_{\text{AdS}}}} + r^2 d\Omega_2^2 \tag{2.10}$$

where $M(r)$ is the total mass located inside the sphere of radius r.

For $r < r_0$ we will ignore the r^2/R^2_{AdS} terms in g_{tt} and g_{rr} responsible for the AdS asymptotics and calculate $M(r)$ due to the back-reaction of the radiation shell. This effect is strongest when the shell is near the origin and its energy is concentrated in a small region within $r < w$. We find that $M(r) \sim \epsilon^2$.

At $r = r_0$, we will start including the AdS terms and ignoring the back-reaction terms, such that for $r > r_0$ the metric is just that of empty AdS space. Naïvely, this is allowed if the metric at r_0 is approximately that of Minkowski space; that is, the corrections due to both AdS and back-reaction must be small:

$$\frac{r_0^2}{R^2_{\text{AdS}}} \ll 1 \quad \text{and} \quad \frac{\epsilon^2}{r_0} \ll 1 \ . \tag{2.11}$$

However, we should really ask for a stronger condition.

Our perturbative back-reaction calculation will be organized as an expansion in powers of ϵ^2/w, and we will work up to some power n using the Minkowski background. In order to be able to trust our results up to that order, we cannot allow the transition at r_0 to have a competing effect, meaning

$$\frac{r_0^2}{R^2_{\text{AdS}}} \ll \left(\frac{\epsilon^2}{w}\right)^n \quad \text{and} \quad \frac{\epsilon^2}{r_0} \ll \left(\frac{\epsilon^2}{w}\right)^n \ . \tag{2.12}$$

For any $R_{\rm AdS}$, we can choose the shell small enough and thin enough to accommodate the hierarchy of scales

$$R_{\rm AdS} \gg r_0 \gg w \gg \epsilon^2 \ , \tag{2.13}$$

which satisfies eq. (2.12) for any choice of n.

The two-region approximation provides a very simple picture. In the $\epsilon \to 0$ limit, the dynamical evolution is totally controlled by the central Minkowski region. For the AdS instability problem, the only meaningful calculation is the gravitational self-interaction of a thin-shell when it passes through $r < r_0$. The propagation in the $r > r_0$ region is just propagation in an empty AdS space; the shell simply travels out, reflects off the boundary, and repeats the gravitational evolution near the origin. Since the profile is modified by a small fraction $\sim (\epsilon^2/w)$ during each bounce, we expect on the time scale $\sim \epsilon^{-2}$ an order-one change to accumulate.

For example, the self-interaction might make the shell thinner after each bounce, meaning that the gravitational effect becomes stronger, since more energy is squeezed into a smaller region. If that behavior persists, then eventually the energy will be compressed during a bounce into a region near the origin smaller than its Schwarzschild radius. At this point, the weak-gravity approximation will break down, and it is very likely that in the ϵ^{-2} time scale, the shell will evolve into a black hole. On the other hand, it is also possible that the shell becomes wider after each bounce, and energy is dispersed into a larger region. In this case, there is no particular reason why gravitational effects would necessarily become strong and no indication that a black hole would form in the ϵ^{-2} time scale. The main goal here is to set up a calculation that can capture these two different behaviors.

Before moving on, we need to address the applicability of the thin-shell approximation. A full dynamical picture should accommodate energy distributions of all thicknesses. However, when $w \sim R_{\rm AdS}$, there is no clean way to separate the self-interaction from the effects of the AdS space. Nevertheless, our main interest is the instability in AdS toward black hole formation. In the small-ϵ limit, the energy must become concentrated into thin shells to even have a chance of eventually forming a black hole. Note that not all of the energy needs to be in one thin shell. But, the evolution toward a black hole is determined by the shell with the highest radial energy density, which is dominated by its self-interaction, so we can ignore the influence of other energy distribution outside the shell.

2.3.2 Near-Minkowski expansion

According to our approximation scheme, we can adopt the weak-gravity expansion in Minkowski space [33]:

$$\phi = \epsilon\phi_0 + \epsilon^3\phi_1 + ... \tag{2.14}$$

$$g_{\mu\nu} = g^0_{\mu\nu} + \epsilon^2 g^1_{\mu\nu} + ... \tag{2.15}$$

At zeroth order in ϵ, the background is empty Minkowski space,

$$\sim \mathcal{O}(\epsilon^0) \ , \qquad ds^2 = -dt^2 + dr^2 + r^2 d\Omega^2 \ , \tag{2.16}$$

into which we put the initial shell profile. To first order, the equation of motion for ϕ is just that of a free field,

$$\sim \mathcal{O}(\epsilon) \ , \qquad \ddot{\phi}_0 - \phi_0'' - \frac{2}{r}\phi_0' = 0 \ . \tag{2.17}$$

At the next order, gravity responds to the stress-energy tensor of the first-order profile. We therefore must solve the Einstein equation $G_{\mu\nu} = 8\pi T_{\mu\nu}$ to leading order in small perturbations around empty Minkowski space. Spherical symmetry excludes dynamical degrees of freedom in the metric, so we only need to solve constraint equations. The tt and rr components suffice to provide the full answer, and the solution is parametrized by two intuitive quantities: enclosed mass M and gravitational potential V.

$$ds^2 = -[1 + 2\epsilon^2 V(r,t)]dt^2 + \left[1 + \frac{2\epsilon^2 M(r,t)}{r}\right] dr^2 + r^2 d\Omega^2 \ . \tag{2.18}$$

Note that we have also explicitly extracted the ϵ scaling from M and V, which are given in terms of the leading-order fields:

$$\sim \mathcal{O}(\epsilon^2) \ , \qquad \frac{2M'}{r^2} = 8\pi \frac{\dot{\phi}_0^2 + \phi_0'^2}{2} \ , \tag{2.19}$$

$$\frac{2}{r}\left(-\frac{M}{r^2} + V'\right) = 8\pi \frac{\dot{\phi}_0^2 + \phi_0'^2}{2} \ , \tag{2.20}$$

with boundary conditions $M(0,t) = 0$ and $V(\infty,t) = 0$. Finally, the leading nontrivial dynamics comes at the next order—the change in geometry back-reacts on the field profile.

$$\sim \mathcal{O}(\epsilon^3) \ , \qquad \ddot{\phi}_1 - \phi_1'' - \frac{2}{r}\phi_1' = C\left(\ddot{\phi}_0 + \phi_0'' + \frac{2}{r}\phi_0'\right) + \dot{C}\dot{\phi}_0 + C'\phi_0' \ . \tag{2.21}$$

Here we have abbreviated $C = (V - M/r)$. We see that the field at this order obeys the same wave equation as in the previous order with the addition of a nontrivial source term.

The radial wave equation can be rewritten as a $(1+1)$-dimensional wave equation by introducing $u = r\phi$:

$$r\left(\ddot{\phi} - \phi'' - \frac{2}{r}\phi'\right) = \ddot{u} - u'' \ . \tag{2.22}$$

This implies that the initial shell profile given in eq. (2.9) is really just the left-moving part of an exact, leading-order solution,

$$r\phi_0(r,t) = u_0(r,t) = \sqrt{w}\left[f\left(\frac{r-t}{w}\right) - f\left(\frac{-r-t}{w}\right)\right] \ . \tag{2.23}$$

We remind the reader that in eq. (2.14), the ϵ dependence has been extracted explicitly for ϕ_0, therefore also for u_0. We have taken the liberty to choose the initial time $t_i = -r_0$ to simplify this expression. This allows us to start this calculation once the shell enters the $r < r_0$ region, and the center of the shell reflects off the origin at $t = 0$. Later, we will be interested in corrections to the profile at $t_f = r_0$, when the shell is leaving the central Minkowski region.

Rewriting the system in terms of the (1+1)-dimensional function u is essentially employing a method of images; we extend the range of r into the unphysical $r < 0$ region. To implement boundary conditions at $r = 0$ such that all physical quantities are finite and smooth, we require $u(r,t)$ to be antisymmetric. Similarly, we can extend the definition of M to negative r,

$$M(r,t) = \int_0^r d\tilde{r}\ \frac{\dot{\phi}_0^2 + \phi_0'^2}{2} 4\pi\tilde{r}^2 \ , \tag{2.24}$$

which is naturally an odd function of r. The same extrapolation shows that V is an even function of r.

In terms of these new variables, the problem of a shrinking shell has been mapped to the problem of two wavepackets colliding at $r = t = 0$. Note that this picture is more realistic than it seems; antipodal points of the shell do indeed collide with each other. When the shell is far from the origin, even the leading-order radial energy density is approximately equal to the naive definition of energy in this $(1+1)$-dimensional simplification:[11]

$$\rho_0 = 4\pi r^2 \frac{\dot{\phi}_0^2 + \phi_0'^2}{2} \ \approx 4\pi \frac{\dot{u}_0^2 + u_0'^2}{2} \ . \tag{2.25}$$

To leading order, the colliding shells simply pass through each other. Our goal is to solve the next-order nontrivial effect of such a collision by solving eq. (2.21),

[11] Note that the total energy $E \equiv M(\infty,t) = \int_0^\infty \rho(r)dr$ is in fact equal to the naive $(1+1)$-dimensional energy $\int_{-\infty}^{\infty} 2\pi \frac{\dot{u}_0^2 + u_0'^2}{2}\ dr$.

which in terms of u is simply

$$\ddot{u}_1 - u_1'' = C\left(\ddot{u}_0 + u_0''\right) + \dot{C}\dot{u}_0 + C'\left(u_0' - \frac{u_0}{r}\right) \equiv S(r,t) \ . \tag{2.26}$$

This description has a striking resemblance to soliton collisions [36, 37]. The key to this type of problem is that, before solving the equations, we should already anticipate the physical meaning of the answer. At $t_f = r_0$, after the collision, the leading-order solution implies that an out-going shell of the opposite sign reaches exactly $r = r_0$. On top of that, we can organize the next-order correction into the following form:

$$u_0 + \epsilon^2 u_1 = u_0 - \epsilon^2\left(\frac{\partial u_0}{\partial r}\Delta r + \frac{\partial u_0}{\partial w}\Delta w + ...\right) \tag{2.27}$$

We have again extracted the ϵ dependence explicitly. The shell is actually shifted by $\epsilon^2\Delta r$ from its expected position, its width has changed by $\epsilon^2\Delta w$, and there will be other changes orthogonal to these.

The function u_1 at $t_f = r_0$ can be solved from eq. (2.26) by integrating the retarded Green's function:

$$u_1(r, r_0) = \frac{1}{2}\int_{-r_0}^{r_0} dt \int_{r-r_0+t}^{r+r_0-t} dr' \ S(r', t) \ . \tag{2.28}$$

Note that the lower limit of r' can be negative, which is allowed due to our method of images. The result, however, is the same if we replace the lower bound of the integration range by its absolute value.

Note that this u_1 is only the difference between the incoming shell at $t = -r_0$ and the out-going shell at $t = r_0$, both at position $r = r_0$. Nevertheless, as we have argued that the propagation further to $r = \infty$, the reflection, and the propagation back to $r = r_0$, can all be taken as trivial. This allows us to directly relate u_1 to the functional A from eq. (2.5), which gives the leading-order change due to one bounce.

$$A\left[\frac{u_0}{r}, \frac{\dot{u}_0}{r}\right] = -\frac{u_1(\tilde{r}, t)}{\tilde{r}} \ , \tag{2.29}$$

where $\tilde{r} = 2r_0 - r$ is the spatial reflection of r around r_0. The extra minus sign and changing to this "flipped" position are due to the trivial propagation to and from $r = \infty$.

The full procedure to calculate u_1 and extract physical information like Δr and Δw are tedious but straightforward. We will present the analytical and numerical details in Appendices 2.A and 2.B. Here we highlight two relevant features of the results:

1. u_1 has a $\sim \log r_0$ contribution, which comes entirely from the position shift,

$$\Delta r = -\frac{\int u_1 \partial_r u_0 \, dr}{\int (\partial_r u_0)^2 \, dr} \,, \tag{2.30}$$

   which has a clear physical meaning. The leading-order profile u_0 follows the $t = |r|$ trajectory, but the next-order correction to the metric modifies the null geodesics. The shell will therefore return to $r = r_0$ not exactly when $t = r_0$. However, this shift is irrelevant to the pertinent question of whether energy gets focused.[12]

2. The change in the shell's width is given by

$$\Delta w = \frac{\int u_1 \partial_w u_0 \, dr}{\int (\partial_w u_0)^2 \, dr} \,. \tag{2.31}$$

   Since we have already scaled out the ϵ dependence, Δw only depends on the shape of the shell (i.e. the function f one chooses in eq. (2.23)), and it is independent of both ϵ and w.

In particular, our main result is that Δw is just as likely to be positive as negative. Specifically, when we flip the profile of the incoming shell, $f(x) \to f(-x)$, then $\Delta w \to -\Delta w$. As a special case, a symmetric profile with $f(x) = f(-x)$ will result in no first order Δw during one bounce. This demonstrates that the gravitational self-interaction in AdS is not biased toward focusing energy, and the collapse of small perturbations into black holes is probably not the generic behavior, at least not on time scales $\lesssim \epsilon^{-2}$.

As a complementary calculation, we also investigate how the maximum radial energy density ρ_{Max} of the shell behaves under the same $f(x) \to f(-x)$ transformation. Like the width w, we find that if for a given profile ρ_{Max} increases with each bounce, then for the flipped profile it decreases. This provides another indication that the weak-gravity dynamics are biased neither toward nor against focusing energy.

In the next section, we will give general proofs of these statements. We will also present numerical examples in Appendix 2.B.

[12] This position shift is related to a shift in frequency in the momentum space analysis observed in other works [12].

2.4 Focusing and defocusing

2.4.1 Antisymmetry of the field correction

As a preliminary step in proving the statements of the previous section, we need to determine how the first-order correction u_1 responds to a spatial flip of the initial profile u_0 such that $f(x) \to \tilde{f}(x) = f(-x)$. We find that

$$u_1(r, r_0) \to \tilde{u}_1(r, r_0) \simeq -u_1(\tilde{r}, r_0), \tag{2.32}$$

where $\tilde{r} = 2r_0 - r$ is again the spatial reflection of r around r_0. Note that this is an approximate statement; for a shell of width w, the error in eq. (2.32) is of order w^2/r_0. As we argued in Sec. 2.3, in the $\epsilon \to 0$ limit, we can choose r_0 to make this error arbitrarily small.

The quantities that enter the expression (2.70) for u_1 are u_0 and its derivatives and C and its derivatives. So, let us first see how these quantities transform under the flip. From eq. (2.23), we can see that:

$$u_0(r, t) \to \tilde{u}_0(r, t) = -u_0(r, -t). \tag{2.33}$$

Then, simply by differentiating the two sides of the equation (either with respect to r or t), we obtain the same transformation behavior for the derivatives of u_0. Now, to see how C transforms, all we need is to determine the transformation of M, defined in eq. (2.24):

$$\begin{aligned} M(r,t) \quad &\to \quad \tilde{M}(r,t) = 2\pi \int_0^r dr' \left(\dot{\tilde{u}}_0^2(r',t) + \tilde{u_0}'^2(r',t) + \left(\frac{\tilde{u_0}(r',t)}{r'} \right)^2 \right. \\ &- \quad \left. 2\frac{\tilde{u_0}(r',t)\tilde{u_0}'(r',t)}{r} \right) \\ &= \quad 2\pi \int_0^r dr' \left(\dot{u_0}^2(r,-t) + u_0'^2(r,-t) + \left(\frac{u_0(r,-t)}{r'} \right)^2 \right. \\ &- \quad \left. 2\frac{u_0(r',-t)u_0'(r,-t)}{r'} \right) \\ &= \quad M(r,-t). \end{aligned} \tag{2.34}$$

Since V has the same behavior as M, then $C(r,t) = V - \frac{M}{r}$ transforms under the flip as:

$$C(r,t) \to \tilde{C}(r,t) = C(r,-t). \tag{2.35}$$

Again, a similar relation holds for the derivatives of C. Combining the above results, we see that the source term $S(r,t)$, defined in eq. (2.26), behaves as

$$S(r,t) \to \tilde{S}(r,t) = -S(r,-t) \tag{2.36}$$

under flipping of the initial profile. Also, by demanding regularity at the origin $r = 0$, the initial profile is antisymmetric in r, which in turn implies that $M(r,t)$ is also antisymmetric in r; hence $C(r,t)$ is symmetric. These properties imply the antisymmetry of $S(r,t)$ in its first argument, $S(r,t) = -S(-r,t)$.

Now we are ready to prove eq. (2.32), starting from the integral expression eq. (2.70) for u_1. The integration regions are illustrated in Fig. (2.2).

We first make an approximation to eq. (2.70). The upper limit of the r' integral is $r + r_0 - t$. Instead, we will extend the region of integration up to $r' = \infty$. Because the wavepacket has compact support only over a region of width w, the error introduced by this approximation comes just from the yellow shaded triangle in Fig. (2.2). The area of this added triangle is $\mathcal{O}(w^2)$ and, since $C(r,t) \sim \frac{1}{r}$, the integrand is of order $\frac{1}{r_0}$. Hence, the error is suppressed by a factor of $\frac{w^2}{r_0}$.

A similar, and perhaps even more physical, approximation, albeit with more cumbersome limits of integration, can be made by considering the area of integration denoted by the red lines together with the orange line in Fig. (2.2). In that case, instead of adding the extra contribution from the yellow triangle at the top, we would subtract the area of the green triangle at the bottom. However, the results would be the same.

After this approximation, we have:

$$u_1(r,r_0) \simeq \frac{1}{2}\int_{-r_0}^{r_0} dt \int_{|r-r_0+t|}^{\infty} dr'\ S(r',t)\ . \tag{2.37}$$

Now, flipping the initial profile we get:

$$\tilde{u}_1(r,r_0) \quad \simeq \quad \frac{1}{2}\int_{-r_0}^{r_0} dt \int_{|r-r_0+t|}^{\infty} dr'\ \tilde{S}(r',t)\ . \tag{2.38}$$

Using the flipping property of $S(r,t)$, as discussed above, we can write:

$$\tilde{u}_1(r,r_0) \quad \simeq \quad -\frac{1}{2}\int_{-r_0}^{r_0} dt \int_{|r-r_0+t|}^{\infty} dr'\ S(r',-t)\ . \tag{2.39}$$

We can now change the dummy integration variable t to $-t$ and use the relation $r = 2r_0 - \tilde{r}$ to rewrite the lower integration limit, giving

$$\tilde{u}_1(r,r_0) \quad \simeq \quad -\frac{1}{2}\int_{-r_0}^{r_0} dt \int_{|(2r_0-\tilde{r})-r_0+t|}^{\infty} dr'\ S(r',t)\ . \tag{2.40}$$

Comparing this expression to eq. (2.37), we obtain

$$\tilde{u}_1(r,r_0) = -u_1(\tilde{r},r_0)\ . \tag{2.41}$$

which is our desired result.

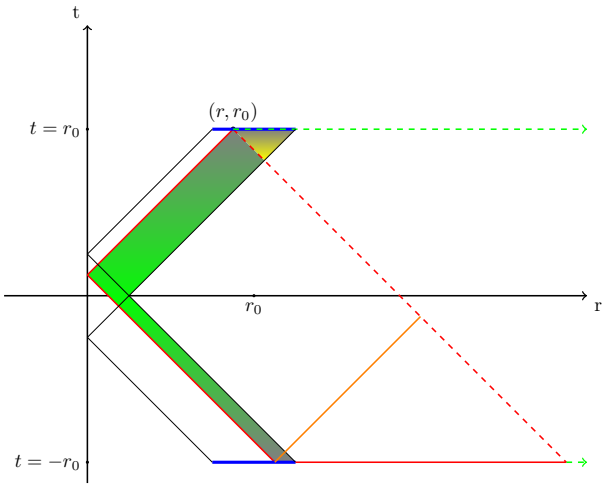


Figure 2.2: *The areas of integration for the computation of $u_1(r, r_0)$. The blue solid line represents the source at times $t = \pm r_0$. The red lines (solid and dashed) indicate the actual area of integration in eq. (2.70) (i.e. the integral $\int_{-r_0}^{r_0} dt \int_{|r-r_0+t|}^{r+r_0-t} d\tilde{r}$), and the green shaded region indicates where the integrand is nonzero. The solid red lines together with the dashed green lines correspond to the region of integration $\int_{-r_0}^{r_0} dt \int_{|r-r_0+t|}^{\infty} d\tilde{r}$ used in our approximate eq. (2.37). The yellow triangle at the top shows the extra nonzero contribution included in the second integral, which is suppressed by $\frac{w^2}{r_0}$. Alternatively, the integral can be approximated by using the orange line instead of the horizontal red line.*

2.4.2 Shell width

In this subsection we will prove eq. (2.31); that is, under a spatial flip $f(x) \to \tilde{f}(x) = f(-x)$, the leading-order correction to the width is antisymmetric:

$$\Delta w \to \Delta \tilde{w} = -\Delta w \tag{2.42}$$

We assume the profile u_0 has compact support within $r_0 - w/2 < r < r_0 + w/2$, and evaluate Δw at late time $t_f = r_0$, well after the collision, at which point the left-moving and the right-moving wavepackets are far away from $r = 0$ and do not interfere with each other. In that case, when computing Δw from eq. (2.31), we can just integrate over the right-moving wavepacket; integrating over both wavepackets would just double both the numerator and denominator in eq. (2.31), yielding the same result. The expression for the change in width of the flipped profile is then

$$\Delta \tilde{w} = \frac{\int_{r_0-w/2}^{r_0+w/2} dr\ \tilde{u}_1(r, r_0) \partial_w \tilde{u}_0(r, r_0)}{\int_{r_0-w/2}^{r_0+w/2} dr\ (\partial_w \tilde{u}_0)^2} \ . \tag{2.43}$$

It is convenient to define $y = r - r_0$, such that the spatial flip of the initial profile is given by

$$u_0(r_0 + y, r_0) \to \tilde{u}_0(r_0 + y, r_0) = u_0(r_0 - y, r_0) \ . \tag{2.44}$$

Starting with the numerator of eq. (2.43) and using the properties of u_0 and u_1 under the flip, we can write

$$\begin{aligned} \int_{r_0-w/2}^{r_0+w/2} dr\ \tilde{u}_1(r, r_0)\partial_w \tilde{u}_0(r, r_0) &= \int_{-w/2}^{w/2} dy\ \tilde{u}_1(r_0 + y, r_0)\partial_w \tilde{u}_0(r_0 + y, r_0) \\ &= -\int_{-w/2}^{w/2} dy\ u_1(r_0 - y, r_0)\partial_w u_0(r_0 - y, r_0) \\ &= -\int_{-w/2}^{w/2} dy\ u_1(r_0 + y, r_0)\partial_w u_0(r_0 + y, r_0) \end{aligned} \tag{2.45}$$

In the third line, we changed the dummy integration variable from y to $-y$. The flip therefore changes the sign of the numerator. Following these same steps with the denominator of eq. (2.43), we can see that it is invariant under the flip. Putting these two statements together yields the desired result, $\Delta\tilde{w} = -\Delta w$.

2.4.3 Energy density

A similar argument holds for the leading-order change in the energy density $\Delta\rho$ at time $t_f = r_0$ due one bounce through the origin. Specifically, for $f(x) \to \tilde{f}(x) = f(-x)$, we find

$$\Delta\tilde{\rho}(r, r_0) \simeq -\Delta\rho(\tilde{r}, r_0) \ . \tag{2.46}$$

where recall $\tilde{r} = 2r_0 - r$. The full radial energy density far from the origin is approximately the (1+1)-dimensional expression, as in eq. (2.25). Expanding it to the next order, we find

$$\rho_0 + \epsilon^2\Delta\rho = 4\pi\frac{\dot{u}_0^2 + {u_0'}^2}{2} + 4\pi\epsilon^2\left(\dot{u}_0\dot{u}_1 + u_0'u_1'\right) \ . \tag{2.47}$$

We kept our principle of always extracting ϵ explicitly. The first term is the initial energy density given in eq. (2.25), which is actually ϵ^{-2} times the actual physical energy density. The second term is the leading change due to a single bounce.

The formula (2.32) we found for the behavior of u_1 under the flip holds at the specific time $t = r_0$, and it is not straightforward to see that the same relation holds for $\dot{u}_1$. An alternative way to proceed is to include the explicit expression

for the derivatives of u_1 at $t_f = r_0$:

$$u_1'(r, r_0) = \frac{1}{2}\int_{-r_0}^{r_0} dt\ [S(r + r_0 - t) - S(r - r_0 + t)] \tag{2.48}$$

$$\dot{u}_1(r, r_0) = \frac{1}{2}\int_{-r_0}^{r_0} dt\,[S(r + r_0 - t) + S(r - r_0 + t)]\ . \tag{2.49}$$

Using these formulae and the expression for u_0 eq. (2.23), and omitting irrelevant constants, we can write down the explicit expression for $\Delta\rho$.

$$\Delta\rho(r, r_0) = \int_{-r_0}^{r_0} dt\ \left[f^{'}(-r - r_0)S(r + r_0 - t, t) - f^{'}(r - r_0)S(r - r_0 + t, t)\right] \tag{2.50}$$

Since the function $f(x)$ has compact support of width w around $x = 0$ and we consider values of r on the order of r_0, then the first term in the integrand vanishes. We can now determine how $\Delta\rho$ behaves under $f(x) \to f(-x)$:

$$\Delta\tilde{\rho}(r, r_0) = -\int_{-r_0}^{r_0} dt f^{'}(-r + r_0)S(r - r_0 + t, -t)\ . \tag{2.51}$$

Changing the dummy integration variable t to $-t$ and substituting with $r = 2r_0 - \tilde{r}$, we obtain

$$\Delta\tilde{\rho}(r, r_0) = -\int_{-r_0}^{r_0} dt f^{'}(\tilde{r} - r_0)S(-\tilde{r} + r_0 - t, t).$$

From the antisymmetry of S in its first argument, we get

$$\begin{aligned}\Delta\tilde{\rho}(r, r_0) &= +\int_{-r_0}^{r_0} dt f^{'}(\tilde{r} - r_0)S(\tilde{r} - r_0 + t, t)\\ &= -\Delta\rho(\tilde{r}, t),\end{aligned} \tag{2.52}$$

which is indeed eq. (2.46).

eq. (2.46) relates the change in energy density at an arbitrary point r and its image $\tilde{r}$ under the flip. However, we are particularly interested in how the change in the maximum energy density is affected by the flip.

The energy density at the position of the maximum, r_{Max}, after one bounce, can be expanded as

$$\rho_{\mathrm{Max}} \equiv \rho(r_{\mathrm{Max}}) = \rho^{(0)}(r_{\mathrm{Max}}) + \epsilon^2\Delta\rho(r_{\mathrm{Max}})\ . \tag{2.53}$$

It might be tempting to directly identify this with the change of maximum energy density. However, we should remember that in addition, the location of the maximum r_{Max} is also, in general, affected by the bounce, receiving corrections at the same order, $r_{\mathrm{Max}} = r^{(0)}_{\mathrm{Max}} + \epsilon^2\Delta r_{\mathrm{Max}}$. So, expanding to order ϵ^2, we find

$$\rho^{(0)}_{\mathrm{Max}} + \epsilon^2\Delta\rho_{\mathrm{Max}} = \rho^{(0)}\left(r^{(0)}_{\mathrm{Max}}\right) + \epsilon^2\rho^{(0)'}\left(r^{(0)}_{\mathrm{Max}}\right)\Delta r_{\mathrm{Max}} + \epsilon^2\Delta\rho\left(r^{(0)}_{\mathrm{Max}}\right)\ . \tag{2.54}$$

However, $r_{\rm Max}$ is an extremum of $\rho^{(0)}$, and so $\rho^{(0)\prime}\left(r^{(0)}_{\rm Max}\right) = 0$. To leading-order, the change in the maximum is due only to a change in ρ and not a shift in the location of the maximum:

$$\Delta\rho_{\rm Max} = \Delta\rho\left(r^{(0)}_{\rm Max}\right) \ . \tag{2.55}$$

Now, under a flip, $r^{(0)}_{\rm Max}$ is mapped to $\tilde{r}^{(0)}_{\rm Max} = 2r_0 - r^{(0)}_{\rm Max}$, the location of the maximum of $\tilde{\rho}$; that is,

$$\Delta\tilde{\rho}_{\rm Max} = \Delta\tilde{\rho}\left(\tilde{r}^{(0)}_{\rm Max}\right) \ . \tag{2.56}$$

From eq. (2.46), we can see that

$$\Delta\tilde{\rho}_{\rm Max} = -\Delta\rho\left(r^{(0)}_{\rm Max}\right) = -\Delta\rho_{\rm Max} \ . \tag{2.57}$$

Therefore, flipping the profile reverses the direction of the change in the maximum energy density. This result nicely complements our result regarding the the change in width eq. (2.42). Both of these results indicate that there is no bias in the weak-gravity dynamics toward either increasing or decreasing the energy concentration.

2.5 Discussion

2.5.1 Phase space diagram

In Sec. 2.3, we provided the recipe to compute the change in a field profile after one bounce, and it contained all the information about the functional A in eq. (2.5). In principle, one can add the resulting $\bar{\phi}_1$ to the original $\bar{\phi}_0$ to make a new initial condition, and calculate the result of the next bounce. Choosing a small ϵ and reiterating this process $\sim \epsilon^{-2}$ times is equivalent to solving eq. (2.7). In principal, this will directly reproduce the long term evolution. Unfortunately, there is one technical difficulty that we have not been able to overcome.

Our method has one disadvantage: energy conservation is by definition an approximation. We basically "turn on" a self-gravitational potential when a shell shrinks below r_0, let energy flow between it and the field kinetic terms, then turn it off when the shell expands over r_0. The amount of potential energy we turn on and off differs by $\sim \epsilon^4 w/r_0^2$. Although this is suppressed by an extra factor of w/r_0 from the quantities we care about in Sec. 2.4 and so it does not invalidate our results, it is technically difficult to control. We could naïvely go to a larger r_0 for better energy conservation, which would increase the integration range required

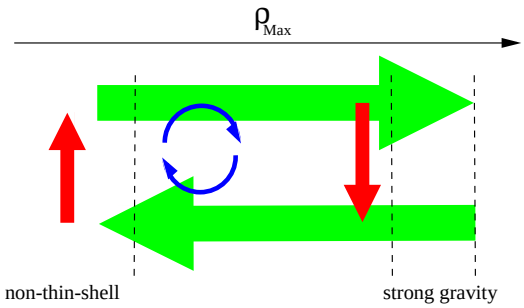


Figure 2.3: *A two-dimensional projection within a constant-energy slice of the phase space. The horizontal axis is the peak energy density, and the big, green arrows toward left and right represent the focusing and defocusing flow due to gravitational self-interaction. Together with the upward and downward flows represented by the small, red arrows, the phase space has a circular flow pattern. The blue loop represents quasi-periodic solutions that stay within the center of this circular flow.*

to solve eq. (2.26). However, more numerical resources would then be necessary in order to proceed.

Nevertheless, we might have learned enough about what happens in a single bounce to make a reliable extrapolation. We will attempt to do so by drawing a phase-space diagram. Since there are no gravitational degrees of freedom within spherical symmetry, the phase space of perturbations is given by all possible scalar field profiles. Due to energy conservation, we can focus on one fixed-energy, co-dimension-one surface in this infinite dimensional phase space. Within this surface, we can draw a two-dimensional projection, Fig. (2.3) and understand its structure based on our knowledge of the dynamics of one bounce.

One guiding principle of this diagram is that during one bounce, the profile changes by an infinitesimal amount $\sim \epsilon^2$, which is also an infinitesimal distance in the diagram. Within the weak-gravity time scale $\sim \epsilon^{-2}$, the evolution trajectory covers a finite distance of the diagram. In this way, the diagram directly represents the dynamical evolution in the rescaled time as given by eq. (2.7).

The horizontal axis of this two-dimensional diagram represents "how close is this profile to becoming a black hole". More technically, it is quantified by the maximum radial energy density at $r = 0$ that is reached during one AdS time. In the small-ϵ limit, the profile is basically freely propagating, so this is a well-defined quantity. Heuristically, this maximum is reached when the highest "peak" goes through $r = 0$, and its value depends on the height of this peak, $\rho_{\rm Max}$.

Note that throughout this chapter, we have been referring to ρ as the rescaled energy density. In our conventions, the actual physical energy density is given by

$\epsilon^2\rho$. It is still convenient to consider the rescaled density here, since $\rho_{\rm Max}$ quantifies how much higher this peak is than the average, namely the relative concentration of energy. Its value increases toward the right-hand-side of Fig. (2.3). For any finite ϵ, there is a finite value of $\rho_{\rm Max} \sim \epsilon^{-2}$ that represents a density high enough to become a black hole. It can be drawn as a vertical line. Some finite distance to the left of this line, we have another line signifying that the energy density is high enough to make gravity too strong to be described by the weak-gravity expansion.

To the left of this second line, gravity is weak enough that our analysis applies. Somewhere even further to the left, our approximation starts to fail for a different reason: we can no longer describe this peak as an isolated thin shell satisfying the hierarchy in eq. (2.13). In the small-ϵ limit, the region to apply our method always exists. This left boundary is not a very clear line. Nevertheless, in this diagram we can roughly picture it as $\rho_{\rm Max} R_{\rm AdS} \sim 1$, that the maximum peak density is comparable to the average density. Clearly, energy is too evenly distributed in the entire AdS space that nothing could be treated as an isolated thin shell.

The vertical axis of this diagram is not intended to represent any particular parameter of the field profile. It is merely reflecting the fact we established in Sec. 2.4 that focusing and defocusing dynamics are equally generic in one bounce. This means one can always find some parameter such that the middle region is divided into two halves: in the upper half, the evolution makes the peak grow higher and moves closer to forming a black hole, and in the lower half, the peak gets lower and moves away from forming a black hole. In App. 2.B, we give specific numerical examples and argue the parameter controlling focusing and defocusing is the asymmetry of energy distribution: focusing occurs when the shell is denser in the leading edge, and defocusing when it is denser in the tail.

In addition to focusing and defocusing which correspond to flowing horizontally in Fig. (2.3), what tendencies to flow in the vertical direction can we identify? Generally speaking, when $\rho_{\rm Max}$ is large, on the left side of Fig. (2.3), the system will tend to flow upwards. This is because a shrinking peak cannot remain the highest peak forever: a growing peak with smaller initial height will eventually take over. If that were the only vertical motion, it would lead to only two possible trajectories: starting in the upper half, $\rho_{\rm Max}$ would keep increasing and flow directly toward black hole collapse; alternatively, staring in the lower region, $\rho_{\rm Max}$ would first decrease, then bounces back to become a black hole. This directly violates many numerical results, so there must be a downward flow somewhere in Fig. (2.3).

The two numerical examples in App. 2.B provide tentative evidence for a downward flow. What we see is that given a symmetric profile on the boundary between the upper and lower region, after one bounce it picks up an asymmetry similar to the profiles in the lower region—its energy becomes denser in the tail. Of course, we

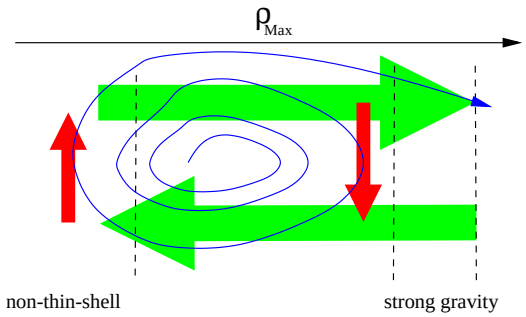


Figure 2.4: *The same circular flow in the phase space, but the blue trajectory now represents an unstable solution. Though initially it follows the circular flow, it fluctuates to larger radius, eventually enters the strong gravity region, and collapses into a black hole.*

studied only two one-parameter slices through an infinite dimensional phase space, so better numerical and/or analytical investigation is required to verify this. Here, we will simply conjecture that such a downward flow exists, because the resulting circulatory flow, shown in Fig. (2.3), explains existing numerical results very well:

- The quasi-periodic solutions stay within the circular flow near the center, as in Fig. 2.3.
- The unstable solutions initially stay within the circular flow, but their radii vary wildly and eventually these solutions enter the strong gravity regime, as in Fig. (2.4).

Note that the actual motion in the true phase space is still very complicated. In this two-dimensional projection, evolution trajectories are allowed to cross each other. Nevertheless, this circular flow allows us to better visualize the dynamics in the phase space.

We can also repeat the argument in Sec. 2.2 in a more pictorial manner. As one reduces ϵ, most of this diagram does not change. Due to the scaling behavior, all trajectories to the left of the strong gravity line remain the same, and so most of the stable solutions remain stable. The trajectories for unstable solutions must cross the strong gravity line to form black holes, so they potentially can change.

Actually, the location of the strong gravity line shifts to the right when ϵ decreases. As the total energy is reduced, it needs to be increasingly focused in order for gravity to become strong, and a collapsing solution must therefore evolve further, across the new weak-gravity regime. In the $\epsilon \to 0$ limit, black hole formation is equivalent to a weak-gravity evolution in which $\rho_{\rm Max}$ goes all the way to infinity.

Although the trajectory for a black hole collapse appears divergent, it does not mean that we can immediately rule out such an evolution. In fact, this divergence is an artifact of the parameter choice, and we should appreciate that $\rho_{\text{Max}} = \infty$ is not an infinite distance away in phase space. Recall that Δw in one bounce is independent of w, so the change in width need not be a small fraction of the total. It is certainly possible to have a profile such that after ϵ^{-2} bounces, Δw is negative and order one, leading to a diverging ρ_{Max}.

The real advantage of this picture is that it recasts the $\epsilon \to 0$ limit of the stability problem into the global regularity problem of determining whether ρ_{Max} diverges. Interestingly, analyzing the regularity of AdS_3 perturbations at finite ϵ below the black hole mass gap is similarly a question of determining whether the energy density diverges. In that case, there is already strong evidence to support regularity [38, 39]. One might hope to reproduce this AdS_3 result in higher dimensions in order to confirm that the instability corners indeed shrink to measure zero. We should again caution that spectral analysis can only provide necessary conditions; it can rule out an instability, but it cannot provide equally strong evidence to support one. If the power spectrum of perturbations agrees with a diverging ρ_{Max}, a long-time evolution of eq. (2.6) in position space is still required for the final answer to the AdS stability problem.

2.5.2 Holographic thermalization

One motivation for studying the stability properties of AdS is to try to learn something about the non-equilibrium dynamics of closed systems. This is due to the AdS/CFT correspondence, which relates this classical gravitational system to the dynamics of a strongly-coupled quantum system. Most investigations of holographic thermalization study the Poincaré patch of AdS, which has an infinite boundary (see, for example, [40–43]). In these cases, any nonzero energy density in the bulk will collapse into a black hole, corresponding to thermalization on the boundary.

Here, instead, we are considering global AdS which has a closed, spherical boundary and therefore a very different thermalization behavior. Other studies of global AdS, such as [35], implicitly assume the connection between forming a black hole in the bulk and thermalization of the boundary system. Although that is valid in some cases, we would like to highlight other possibilities. What are the possible holographic dual descriptions of the bulk story presented here?

One caveat is that explicit examples of the AdS/CFT correspondence usually contain compact extra dimensions whose sizes are comparable to R_{AdS}, for example in $AdS_5 \times S_5$. In the $\epsilon \to 0$ limit, the five-dimensional AdS-Schwarzschild black

hole has a horizon radius much smaller than $R_{\rm AdS}$ and is therefore too small to represent the most typical states; a ten-dimensional black hole of the same total energy, which breaks the symmetry of the S^5, has even higher entropy. Therefore, the gravitational stability problem of AdS_5 does not directly translate to the thermalization problem in the boundary system. This might be an interesting direction for future work, but we will set this concern aside for now. Let us take a very optimistic point of view that the AdS/CFT correspondence can work with extra dimensions arbitrarily smaller than $R_{\rm AdS}$, or even without them.

After limiting our attention to the AdS space and treating our classical field theory as a limit of a quantum gravity theory, the $\epsilon \to 0$ limit leads to a different issue. Recall the well known Hawking-Page transition [44]: A black hole does not always dominate the micro-cannonical ensemble; given low enough energy, thermal gas is the most typical state. In this case, forming a black hole does not imply thermalization. This is the main issue we wish to clarify.

First of all, this issue highlights the importance of our position space approach. Focusing on the power spectrum, initial conditions occupying only low frequency modes must propagate to higher frequency in order to approach either a black hole or a thermal gas state. This type of turbulent cascade is a necessary condition for thermalization. However, without differentiating between the final states toward which the system could be evolving, one cannot argue unequivocally for or against thermalization.

Next, let us analyze under what circumstances the black hole or the thermal gas state will dominate the ensemble. For simplicity, we will work via dimensional analysis and ignore any order-one factors. First, note that the $\epsilon \to 0$ limit is actually the weak-gravity limit, corresponding to

$$\beta \equiv \frac{R_{\rm AdS}}{R_{\rm BH}} \gg 1 \ . \tag{2.58}$$

Namely, the Schwarzschild radius of the black hole made by collapsing the scalar field energy is much smaller than the AdS size. On the other hand, the most straightforward standard for trusting classical gravity is

$$\gamma \equiv \frac{R_{\rm BH}}{l_{\rm Planck}} \gg 1 \tag{2.59}$$

where we have restored the Planck scale $l_{\rm Planck} = \sqrt{G\hbar}$, which has been set to one in the rest this work. This condition implies that, at the very least, if a black hole forms, it could be described by classical gravity. For the limit we have been considering, both β and γ have been taken to infinity. We will see that whether the black hole or the thermal gas dominates depends on the details of how that limit is taken.

The entropy of a black hole with energy E is[13]

$$S_{\rm BH} \sim (l_{\rm Planck} E)^2 \sim \left(\frac{R_{\rm BH}}{l_{\rm Planck}}\right)^2 \ . \tag{2.60}$$

The entropy of a thermal gas in AdS space with the same total energy is

$$S_{\rm gas} \sim \left(\frac{R_{\rm BH} R_{\rm AdS}}{l_{\rm Planck}^2}\right)^{3/4} \ . \tag{2.61}$$

Thus, black hole states dominate the micro-canonical ensemble when

$$\left(\frac{R_{\rm BH}}{l_{\rm Planck}}\right) > \left(\frac{R_{\rm AdS}}{R_{\rm BH}}\right)^{3/2} \quad ; \quad \gamma > \beta^{3/2} \ . \tag{2.62}$$

This condition is equivalent to comparing the thermal wavelength $\lambda_{\rm th}$ of the gas to the black hole radius; the black hole dominates the ensemble if

$$R_{\rm BH} > \lambda_{\rm th} \ . \tag{2.63}$$

We can see that whether the condition in eq. (2.62) is satisfied depends on how the limits of large β and large γ are taken. A classical and small-ϵ limit does not restrict the system to being dominated by either the thermal gas or black hole states.

Note that whether, and for how long, classical evolution is a good approximation depends on more details of the state. For example, even if a black hole forms which is classical according to eq. (2.59), the process by which it formed might not be. A simple rule of thumb for the validity of the classical limit is that the occupation numbers in the modes of interest have to be large. If the system is in a state where the energy is roughly equipartitioned between a number of modes up to some maximum frequency $\omega_{\rm max}$, we require

$$\text{energy per mode} \gg \omega_{\rm max} \ . \tag{2.64}$$

The thermal gas states can never satisfy this condition because modes with frequency of order the temperature have occupation numbers of order one, yet contribute a significant fraction of the entropy of the gas. Independent of the limiting procedure and which states dominate, the thermal gas final state is never compatible with a classical description.[14]

[13] Note that we are assuming here that the spacetime is effectively AdS_4 at distance scales of order the size of the black hole; in string constructions, such as $AdS_4 \times S_7$, black holes whose radius is small compared to the AdS radius would be eleven-dimensional rather than four-dimensional, leading to different formulas.

[14] Nevertheless, from the position-space viewpoint, classical evolution may still describe the "process of approaching" a thermal gas state, at least distinguishing it from approaching a black hole. In the latter case energy becomes more concentrated, but in the former case it does not.

On the other hand, a spherically symmetric collapse into a black hole can often be completely classical. Such a process only needs to excite the radial modes from the longest wavelength $\sim R_{\rm AdS}$ to the shortest wavelength $\sim R_{\rm BH}$, thus

$$\text{number of modes} = \frac{R_{\rm AdS}}{R_{\rm BH}} \ . \tag{2.65}$$

The condition on occupation numbers, eq. (2.64), becomes

$$\beta \ll \gamma^2 \ . \tag{2.66}$$

Comparing this to eq. (2.62), we see that eq. (2.66) is always true when the black hole dominates the ensemble, but it can still be true even if thermal gas dominates.[15] *Thus, the specific stability problem within classical gravity investigated in this chapter, namely a spherically symmetric collapse into a black hole, is a valid dual to some boundary system, independent of whether such a process is equivalent to a efficient thermalization or not.*[16]

Furthermore, when the thermal gas dominates, if a black hole really forms in the time scale we investigated,

$$T_{\text{weak gravity}} = R_{\rm AdS} \frac{R_{\rm AdS}}{R_{\rm BH}} \propto \epsilon^{-2} \ , \tag{2.67}$$

it could represent a significant delay to thermalization. In order to confirm this, we need to compare the naïve thermalization time $T_{\text{weak gravity}}$ to the black hole lifetime, given by the the evaporation time scale,

$$T_{\text{evaporation}} = \frac{R_{\rm BH}^3}{l_{\rm Planck}^2} \ . \tag{2.68}$$

When $T_{\text{evaporation}} > T_{\text{weak gravity}}$, which requires

$$\beta < \gamma \ , \tag{2.69}$$

the system thermalizes only after forming a long-lived black hole, which eventually evaporates. This process of thermalization via a quasi-stationary thermal-like state is known as prethermalization and has been observed in finite-sized, isolated quantum systems [30, 31]. Note that eq. (2.69) is compatible with thermal gas domination and a classical collapse.

[15] Note that spherical symmetry is very important here. Without it, the number of modes would have been $\left(\frac{R_{\rm AdS}}{R_{\rm BH}}\right)^3$, and with that many modes, the black hole collapse would have failed to remain classical in the thermal gas-dominated regime.

[16] Here, "efficient" means that thermalization happens in the shortest time scale allowed by the dynamics, $\sim \epsilon^{-2}$. One should not confuse this with, for example, the much faster thermalization in the Poincaré patch of AdS, where, within one AdS time, perturbations cross the horizon, form a planar black hole, and appear to thermalize.

To summarize, spherically symmetric black hole formation within $T_{\text{weak gravity}}$ can have two different holographic interpretations:

- When $\gamma^2 > \beta^3$, it represents efficient thermalization of the boundary system.
- When $\beta < \gamma < \gamma^2 < \beta^3$, it represents prethermalization, which delays true thermalization (to thermal gas[17]) at a time scale $\gtrsim T_{\text{evaporation}} > T_{\text{weak gravity}}$.

For the remaining possibility, when $\gamma^2 < \beta^3$ but $\beta > \gamma$, the implication of black hole formation is inconclusive from a thermalization point of view. Black holes decay too fast to be quasi-stationary intermediate states, but their evaporation cannot guarantee reaching the thermal gas state either.

2.6 Summary

- By combining existing numerical data with our analysis, we have argued that for a massless scalar field in AdS space, in the small-amplitude $\epsilon \to 0$ limit, solutions remaining stable up to the interaction time scale $T \sim \epsilon^{-2}$ form an open set. This improves similar observations in finite-ϵ numerical simulations [10,11] and argues against the conjecture that the weakly turbulent instability occurs in all but a set of measure zero in the space of initial conditions [8,14,15].
- One important difference between our approach and previous work is that we analyzed the problem in position space. We pointed out that only position space properties can provide necessary and sufficient conditions for the collapse into a black hole. Any analysis of the power spectrum can at most provide necessary conditions for black hole formation.
- In the position space analysis, we exploited the small-amplitude $\epsilon \to 0$ limit and argued that the only relevant dynamics are the gravitational self-interactions near $r = 0$. This argument requires a hierarchy of scales given in eq. (2.13), which is difficult to reach in realistic numerical simulations.
- We showed that gravitational self-interaction near $r = 0$ obeys an exact antisymmetry under time reversal. As a result, it is equally generic for interactions to focus or defocus energy. This equality is consistent with

[17] Note that since thermal gas is never classical, we do not know exactly when will it really form. We only know that within $T_{\text{evaporation}}$, the systems was too busy forming a black hole and then remaining as one, so it cannot reach the thermal gas state yet.

existing numerical results.[18] We remind the readers that gravity can be effectively repulsive: tidal forces tend to pull things apart. The possible defocusing of a radiation shell is due to such tidal effect.

- By making use of scaling symmetry, we simplified the stability problem in the $\epsilon \to 0$ limit into a global regularity problem within a finite rescaled time. The evolution was recast as a simple, first-order differential equation. We hope that this point of view, combined with the other techniques in this work and the existing literature, will allow a rigorous analysis of stability in the vanishing amplitude limit.
- Even if black holes do form in the ϵ^{-2} time scale, we point out that it does not always represent efficient thermalization of the boundary theory via AdS/CFT duality. In some cases, black hole formation describes prethermalization, and actual thermalization is delayed until this black hole evaporates.

2.A Analytical details

In this appendix, we will clarify some analytical details omitted in Sec. 2.3. There we showed how to reach a simple differential equation for u_1, eq. (2.26), which can be solved simply by integrating the Green's function:

$$\begin{aligned} u_1(r,t_f) &= \frac{1}{2}\int_{-r_0}^{r_0} dt \int_{r-r_0+t}^{r+r_0-t} dr'\ S(r',t) \\ &= \frac{1}{2}\int_{-r_0}^{r_0} dt \int_{r-r_0+t}^{r+r_0-t} dr'\ \left(C(\ddot{u}_0+u_0'')+\dot{C}\dot{u}_0+C'\left(u_0'-\frac{u_0}{r'}\right)\right) \end{aligned} \tag{2.70}$$

Here we should be careful about our method of images. A physical solution ϕ_1 is only given by a u_1 that is an odd function of r, and it is not obvious that the u_1 given by the above integral will have this property. Another potentially worrisome observation is that the lower limit of the r' integral can be negative for some positive r, but a physical answer should only invoke an integration over the physical space $r > 0$ where the quantity $C = V - M/r$ is naturally defined.

In this case, these concerns about the method of images can be easily resolved. As explained in Sec. 2.3, we can generalize the definition of V and M to include the $r < 0$ region. We will find that M is an odd function of r and V is even.

[18] More specifically, one could take any numerical simulation and pause it at a moment when gravity is still weak. If one keeps the field profile but reverses the time derivative at this moment, the simulation will literally evolve backward toward the original initial profile, up to the numerical error and higher-order effects (which are small if the hierarchy of scales in eq. (2.13) is satisfied during the process).

Together with the fact that u_0 is odd, we see that the integrand in eq. (2.70) is odd. Any integration over negative r is canceled by an equal region with positive r, so effectively the lower limit of the r' integral is $|r-r_0+t|$. eq. (2.70) is effectively only integrating over the physical range. It is, however, more convenient to keep working in this form and avoid the confusion of taking an absolute value. An odd integrand here also guarantees that u_1 is an odd function which leads to a physical ϕ_1.

The form of eq. (2.70) clearly suggests some integrations by parts.

$$
\begin{aligned}
u_1(r,t_f) &= -\frac{1}{2}\int_{-r_0}^{r_0} dt \int_{r-r_0+t}^{r+r_0-t} dr'\ C'\frac{u_0}{r'} \ +\ \frac{1}{2}\int_{-r_0}^{r_0} dt \int_{r-r_0+t}^{r+r_0-t} dr'\ (Cu_0''+C'u_0') \\
&\quad +\frac{1}{2}\int_{r-2r_0}^{r+2r_0} dr' \int_{-r_0}^{r_0-|r-r'|} dt\ (C\ddot{u}_0+\dot{C}\dot{u}_0) \qquad (2.71)\\
&= -\frac{1}{2}\int_{-r_0}^{r_0} dt \int_{r-r_0+t}^{r+r_0-t} dr'\ C'\frac{u_0}{r'} \\
&\quad +\frac{1}{2}\int_{-r_0}^{r_0} dt\ C\Big[(r+r_0-t),t\Big]u_0'\Big[(r+r_0-t),t\Big] \\
&\quad -\frac{1}{2}\int_{-r_0}^{r_0} dt\ C\Big[(r-r_0+t),t\Big]u_0'\Big[(r-r_0+t),t\Big] \\
&\quad +\frac{1}{2}\int_{r-2r_0}^{r} dr'\ C\Big[r',(r_0-r+r')\Big]\dot{u}_0\Big[r',(r_0-r+r')\Big] \\
&\quad +\frac{1}{2}\int_{r}^{r+2r_0} dr'\ C\Big[r',(r_0-r'+r)\Big]\dot{u}_0\Big[r',(r_0-r'+r)\Big] \\
&\quad -\frac{1}{2}\int_{r-2r_0}^{r+2r_0} dr'\ C(r',-r_0)\dot{u}_0(r',-r_0) \\
&= \frac{1}{2}\int_{-r_0}^{r_0} dt\ C\Big[(r-r_0+t),t\Big]\Big(\dot{u}_0\Big[(r-r_0+t),t\Big]-u_0'\Big[(r-r_0+t),t\Big]\Big) \\
&\quad +\frac{1}{2}\int_{-r_0}^{r_0} dt\ C\Big[(r+r_0-t),t\Big]\Big(\dot{u}_0\Big[(r+r_0-t),t\Big]+u_0'\Big[(r+r_0-t),t\Big]\Big) \\
&\quad -\frac{1}{2}\int_{-r_0}^{r_0} dt \int_{r-r_0+t}^{r+r_0-t} dr'\ C'\frac{u_0}{r'}-\frac{1}{2}\int_{r-2r_0}^{r+2r_0} dr'\ C(r',-r_0)\dot{u}_0(r',-r_0)\ .
\end{aligned}
$$

In the above equation, we first isolated two terms which should be integrated by parts, and for one of them we interchange the order of integration so it can be done with respect to t instead of r'. The integration by parts produces five boundary terms as line integral along five segments which we explicitly write down. Finally, two pairs of segments can combine with each other and be expressed as time integrals. We collect the remaining space integral and the only non-boundary term which cannot be integrated by parts in the end.

Note that up to this step, we have not used any approximations. We did not even use the property that C is sourced by ϕ_0. In other words, this expression could describe the change in the field profile under the influence of any other spherically symmetric gravitational effects, either apart from or on top of its self-interaction.

Our next step is to plug in eq. (2.23) and use our assumption that it represents a thin shell: we assume that u_0 is only nonzero within two narrow packages around $r = t$ and $r = -t$. This significantly simplifies eq. (2.71) to

$$\begin{aligned} u_1(r, t_f) \quad = \quad & -\frac{\epsilon}{\sqrt{w}}\, f'\left(\frac{r - r_0}{w}\right) \int_{-r_0}^{r_0} dt\; C\Big[(r - r_0 + t), t\Big] \\ & +\frac{\epsilon}{\sqrt{w}}\, f'\left(\frac{-r - r_0}{w}\right) \int_{-r_0}^{r_0} dt\; C\Big[(r + r_0 - t), t\Big] \\ & -\frac{1}{2}\int_{-r_0}^{r_0} dt \int_{r-r_0+t}^{r+r_0-t} dr'\; C'\frac{u_0}{r} - \frac{1}{2}\int_{r-2r_0}^{r+2r_0} dr'\; C(r', -r_0)\dot{u}_0(r', -r_0)\ . \end{aligned} \tag{2.72}$$

Note that here the f' means a derivative with respect to the variable of f instead of a r derivative, which should always be clear from the context.

Since in the end, we are only interested in the physical range $r > 0$, we can actually drop the second term because the profile f is zero there. This starts to take the promised form of eq. (2.27), and we can almost identify

$$\Delta r = \int_{-r_0}^{r_0} dt\; C\Big[|r - r_0 + t|, t\Big]\ . \tag{2.73}$$

Note that we have added an absolute value to the first variable in C. This makes no difference since it is even, but it helps to emphasize the fact that the integral can be strictly limited to the physical $r > 0$ region.

The physical meaning of eq. (2.73) now becomes clear. When the metric includes first-order corrections, such as in eq. (2.18), a null ray actually follows

$$\left(1 + \frac{M}{r}\right)|dr| = (1 + V)\; dt\ . \tag{2.74}$$

Thus an incoming null ray starting from $r = r_0$ and $t_i = -r_0$ does not exactly return to $r = r_0$ at $t_f = r_0$; the amount it misses is exactly given by eq. (2.73). The leading-order correction due to gravity, of course, includes the fact that geodesics are changed, and the shell simply follows the new geodesic. A geometric calculation is enough to determine how much a localized object appears to be shifted from the position predicted by the zeroth-order theory.

For any finite-sized source, the gravitational potential at large r is proportional to $1/r$, so the integral in eq. (2.73) actually had a piece proportional to $\log r_0$. Since

we are have taken r_0 to be large, one might have worried that such a term would the ruin perturbation expansion. However, such the $\log r_0$ term is totally expected from the change of geodesics and does not interfere with our goal of computing the change in width or other changes.

One last concern about the position shift in eq. (2.73) is that it is a function of r. This turns out not to be a problem either, since the r-dependent part of Δr is not proportional to $\log r_0$. We can see this by taking a derivative with respect to r:

$$\begin{aligned}\partial_r \Delta r &= \partial_r \left(\int_{-r_0}^{r_0-r} dt\ C\Big[(r_0 - r - t), t\Big] + \int_{r_0-r}^{r_0} dt\ C\Big[(t - r_0 + r), t\Big] \right) \\ &= \left(-\int_{-r_0}^{r_0-r} dt\ C'\Big[(r_0 - r - t), t\Big] + \int_{r_0-r}^{r_0} dt\ C'\Big[(t - r_0 + r), t\Big] \right) \end{aligned} \tag{2.75}$$

According to the Einstein's equation, we have

$$C' = V' - \frac{M'}{r} + \frac{M}{r^2} = \frac{2M}{r^2} + \frac{r}{2}\left(T_{rr} - T_{tt}\right) \ . \tag{2.76}$$

This means that as long as we restrict the matter sources to (1) finite-sized sources that vanish beyond some fixed r and/or (2) radiation in the radial direction, $T_{rr} = T_{tt}$, then the r dependence of Δr will not have a $\log r_0$ (or any other large r_0) dependence. Furthermore, there is no small-r divergence either, since the two terms in eq. (2.75) takes opposite signs and cancel each other near $r = 0$. Pictorially, this means that different infinitesimal segments within the wavepacket "shift" differently from one another by some finite amount.

In the last line of eq. (2.72), the first term is also finite for the same reason as eq. (2.76), and the second term is finite because u_0 has compact support. These terms should be combined and understood as some perturbative deformations of the wavepacket profile. They are cleanly separated from the $\Delta r \sim \log r_0$ overall shift, which is uniquely defined by a projection:

$$\Delta r = -\frac{\int u_1 \partial_r u_0\ dr}{\int (\partial_r u_0)^2\ dr} \ . \tag{2.77}$$

We can simply remove this shift mode from eq. (2.27) and study the other deformations. A more physical way to understand the removal of this shift is letting the wavepacket evolve an extra time $\Delta t = \Delta r$ such that it really reaches position r_0; then it will be fair to compare with the zeroth-order profile at the same position.

In order to eventually form a black hole, we need the energy density to become large. Since the total energy is conserved, the most trivial way to increase the energy density is to narrow the width of the profile. The leading-order change in

width can be extracted from u_1 by the following projection:

$$\Delta w = \frac{\int u_1 \partial_w u_0 \, dr}{\int (\partial_w u_0)^2 \, dr} \, . \tag{2.78}$$

Note that the ϵ dependence was already scaled out in eq. (2.14). Interestingly, ϵ^2 has the unit of length in our conventions, and the physical change in width is $\epsilon^2 \Delta w$. Therefore, Δw is dimensionless. The width w is the only other dimensionful quantity that can potentially affect Δw in the leading order (r_0 affects only the subleading error), and so there is no way it can enter the expression for Δw.

What we really wish is to determine the sign of Δw. $\Delta w < 0$ means that the shell gets narrower, and several bounces later it might form a black hole. On the other hand, $\Delta w > 0$ means that the shell gets wider, and several bounces later the energy will be more diluted, which in some sense is moving away from a black hole in phase space.

One technical point to note here is that, given our shell profile eq. (2.23), the mode $\partial_w u_0$ actually measures the scaling of the profile around some center. If that center is not the center of mass, then this scaling not only changes the width but also shifts the position. A simple projection will be contaminated by the large $\sim \log r_0$ contribution from the position shift. We will avoid this by always defining the profile $f(x)$ to have its center of mass at $x = 0$. This means that, on top of the normalization

$$\int f'(x)^2 \, dx = 1 \, , \tag{2.79}$$

we also demand that

$$\int x f'(x)^2 \, dx = 0 \, . \tag{2.80}$$

This guarantees that the scaling mode $\partial_w u_0$ is orthogonal to the shift mode $\partial_r u_0$.

2.B Numerical examples

2.B.1 The asymmetry-focusing correlation

In this appendix, we numerically evaluate the change to a thin-shell profile after one bounce. Our example will be the following two one-parameter families of

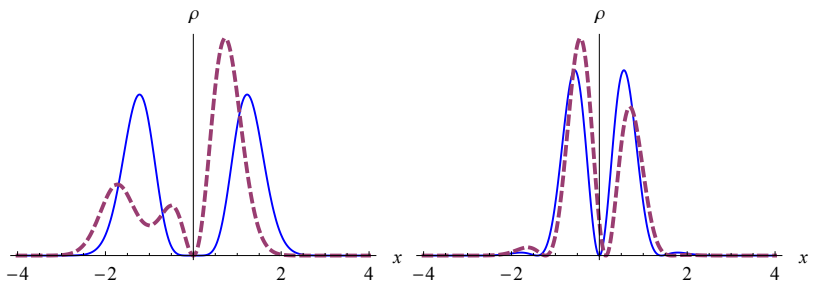


Figure 2.5: *The left panel shows the energy density of the profiles with the "+" sign defined in eq. (2.81), and the right panel shows the profiles with the "−" sign. The blue curves are the symmetric, $a = 0$ profiles. The red (dashed) curves are for $a = 0.5$, which is the maximally "tilted" profile in the range we scanned through. We can see that the family with the "+" sign is more sensitive to the parameter a.*

profiles.

$$g_a(x) = (1 + ax \pm x^2)\, e^{-x^2} \,, \tag{2.81}$$

$$N_a = \sqrt{\int g_a'^2 \, dx} \,, \tag{2.82}$$

$$X_a = N_a^{-2} \int x g_a'^2 \, dx \,, \tag{2.83}$$

$$f_a(x) \equiv N_a^{-1} g_a(x + X_a) \,. \tag{2.84}$$

They are symmetric when $a = 0$, and varying a scans through two different directions of asymmetry. Note that the quadratic term is necessary. Without it, a small a simply means an overall shift in position and the profile will be still symmetric to leading order. Our definition of $f_a(x)$ shifts the center of mass back to $x = 0$ and preserves only the asymmetry generated by a.

We plot some representative profiles in Fig. (2.5). Note that for both families, we have $f_a(x) = f_{-a}(-x)$. Scanning through positive and negative values of a can confirm our analytical proof in Sec. 2.4 that flipping the profiles leads to opposite behaviors within one bounce. It will also provide a better understanding about what physical quantity really affects whether a profile becomes focused or not.

There are infinite ways to be asymmetric, and our parameter a certainly is not the unique parameter to quantify the asymmetry. It also has no reason to be the asymmetry directly responsible for focusing or defocusing the profile. However, for any family of profiles centered around a symmetric one, we can define a natural parameter to quantify the asymmetry, at least for small values of a. Here is how it goes. First of all, g_a has a center of mass shifted by X_a from g_0 by the definition

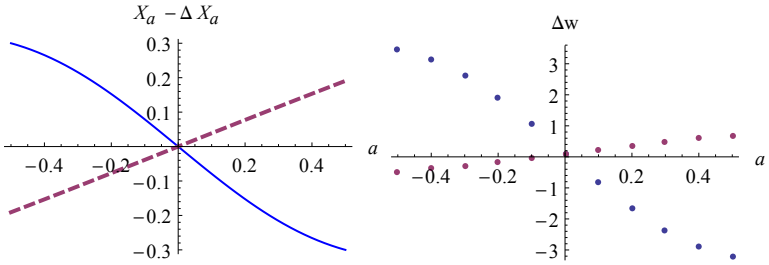


Figure 2.6: *The left panel shows the asymmetry parameter defined as the difference between the center-of-mass shift and the field-profile shift. The blue curve is for the "+" family and the red curve for the "−" family. The right panel shows the change is width after one bounce, which qualitatively agrees with the asymmetry parameter. These are done with $r_0 = 60$ and $w = 1$. Recall that the physical change in width is actually $\epsilon^2 \Delta w$.*

in eq. (2.83). On the other hand, one can also naturally define the shift by a projection to the zero mode, which is exactly the way we defined Δr in eq. (2.77).

$$\bar{X}_a = N_0^{-2} \int [g_0(x) - g_a(x)] g_0'(x)\, dx \ . \tag{2.85}$$

These two shifts already disagree at linear order in a, therefore the amount of their disagreement, $\Delta X_a = (X_a - \bar{X}_a)$, seems to be a reasonable way to quantify the asymmetry.

Given these profiles, we solve eq. (2.18) for M and V, and then we can integrate eq. (2.70). When we scan the parameter a from -0.5 to 0.5, the change in width Δw defined in eq. (2.78) follows a pattern closely resembling the behavior of this asymmetry parameter, ΔX_a. We compare them side-by-side in Fig. (2.6). Note that they are not identical. For example, the relative slopes between the two families near $a = 0$ are not the same. Thus, although we see a rough correlation between them, we cannot claim that our asymmetry parameter directly controls focusing or widening in one bounce.

In our conventions, the profiles are moving toward the right in Fig. (2.5). If we compare their shapes in Fig. (2.5) to their behaviors in Fig. (2.6), we get the following impression:

- When the wavepacket is denser in the front, we get $\Delta w < 0$. The shell gets focused into a smaller region, and gravitational effect will become stronger in the next bounce.
- When the wavepacket is denser in the tail, we get $\Delta w > 0$. It profile gets wider after one bounce, and gravitational effect will become weaker in the

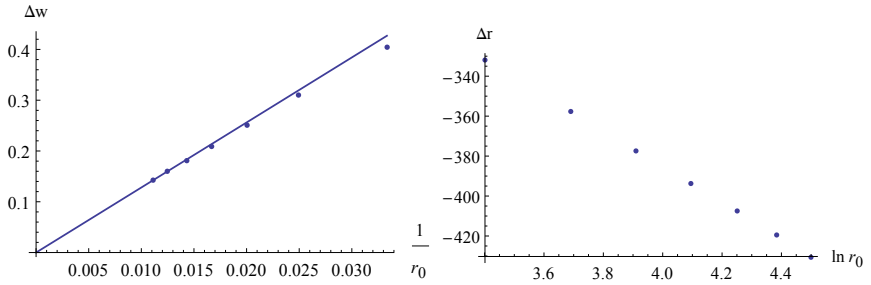


Figure 2.7: *Results with $a = 0$ and varying r_0 from 30 to 90 in steps of 10. The left panel shows Δw, which indeed goes to zero as $1/r_0$. The right panel shows the position shift Δr defined in eq. (2.77) which has the correct $\log r_0$ dependence.*

next bounce.

Since the family of profiles with "+" sign is much more sensitive to the parameter a, we will use it to test other behaviors later in App. 2.B.2.

In Fig. (2.6), one might notice that for the $a = 0$ symmetric profiles, the changes in width are not exactly zero as we argued earlier. This deviation is not physical but simply an artifact of our approximation. That is because although the physical solution is symmetric in time, our technical choice breaks that symmetry by a small amount. We have set the correction to the field profile at the initial time to zero, $u_1(r, -r_0) = 0$. This is a small error since the first-order correction to the metric in eq. (2.18) would have already modified the free field profile at that time, by a small fraction $\sim \epsilon^2 V \sim \epsilon^2/r_0$.

We test this explanation by varying r_0 and verifying that Δw goes to zero in the expected way; see Fig. (2.7). We also verified that the position shift indeed has an r_0-dependent shift $\Delta r \propto \log r_0$, as discussed in App. 2.A.

Finally, with a symmetric profile, one can ask for a prediction for the next bounce. What we observed in these examples is that a symmetric profile will pick up a $\Delta X_a > 0$ in one bounce. This is a very tentative evidence that in the next bounce, they will have $\Delta w > 0$, namely their energy become defocused. We stress again that this is not a proof, but merely two examples we observed. A more thorough investigation is required to support the generic downward flow we conjectured in Sec.2.5.

2.B.2 The effect of another object

Black hole formation does not always involve all of the energy becoming concentrated into a thin shell. For example, an initially smooth field profile might start to develop one or more sharp peaks. It is possible for the energy density in these peaks to become large enough to induce strong gravity and black hole collapse before or even without the average density of the entire profile ever becoming large.

In this section, we will present some evidence that sharp peaks evolve similarly to thin shells. In the perturbative regime, one is free to separate the matter into components in many ways. In particular, we can treat a smooth field profile with a sharp peak as a thin shell propagating in the background of some additional diffuse source of gravity. Our approach is convenient since eq. (2.70) and further analysis about it do not rely on the specific metric ansatz eq. (2.18). As long as the additional source are also spherically symmetric, we can simply repeat the calculation in the previous appendix.

We will start by considering a simple situation in which the additional matter sources are static. In addition to the thin shell with total mass $4\pi\epsilon^2$, we have

$$M_0(r) = 10\epsilon_{\rm star}^2 \tanh\left(\frac{r}{w_{\rm star}}\right)^2 , \tag{2.86}$$

$$P_r \approx 0 . \tag{2.87}$$

This is a star of roughly constant density, radius $w_{\rm star}$ and total mass $10\epsilon_{\rm star}^2$. It does not interact with the massless field which forms the shell in any other way other than gravitationally, so it simply enters by altering the metric in eq. (2.18). We assume the star is stable and supports itself by a negligible amount of radial pressure (but it can have angular pressure), so it does not add any extra term to modify T_{rr}.

According to the momentum space analysis, including this additional gravitational source breaks the AdS resonance structure and should interfere with black hole formation [28]. We show that such an interpretation is not necessary to understand the dynamics of thin shells in one bounce. Remember that for a symmetric shell profile, we argued that there cannot be a change in width due to the time-reversal symmetry. Adding an extra, static source does not break that symmetry, so symmetric profiles again cannot change in size. And, it is straightforward to verify that asymmetry still focuses or defocuses in qualitatively the same way as before.

In Fig. (2.8), we demonstrate that whether the shell gets thinner or thicker has the same dependence on the asymmetry induced by the parameter a. Its magnitude

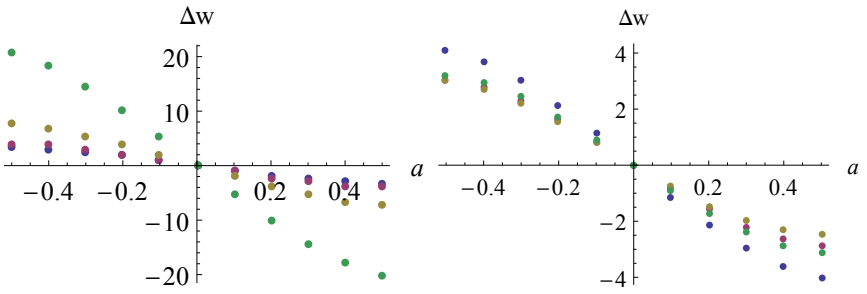


Figure 2.8: *We plot the change in width Δw as a function of the parameter a in eq. (2.81). Both figures are with $r_0 = 30$. The left panel shows the effect of dialing the mass of the additional matter source, $\epsilon^2_{\rm star} = 0, 10, 50, 200$, without changing its size $w_{\rm star} = 1$. The right panel shows the effect of dialing the size while keeping the same density, $w_{\rm star} = 1, 5, 20, 200$. We have removed the errant $1/r_0$ contribution by hand.*

does have an interesting dependence on the additional source. In the first set of data we fix the size of the star to be equal to the shell. The change in width turns out to grow linearly with the additional mass. On the other hand if we keep the same density and increase the size of the star, the change in width is not strongly affected.

3

Conditionally extended validity of perturbation theory: Persistence of stability islands

Beauty is the first test; there is no permanent place in the world for ugly mathematics.

— Godfrey H. Hardy

3.1 Introduction and Summary

So far we have presented two approximate methods for dealing with the problem of small perturbations. One is based on *Fourier space* analysis and the other one is based on a *position space* analysis. The upshot is that they both result in a set of *approximate* equations that possess a very important scaling symmetry. However such approximate schemes break down after some time, which in the problem at hand is the first nonlinear time scale $t \sim \epsilon^{-2}$. Coincidentally, this time scale is the the one at which black hole formation takes place [8]. Therefore one is actually interested in what happens exactly at this time scale and if we can trust our approximate schemes up to this *long* time scale. This is exactly the question that we are engaged with in this chapter.

3.1.1 Truncated Perturbative Expansion

A linear equation of motion $D\phi = 0$ often has close-form analytical solutions. A nonlinear equation, $D\phi = F_{\text{nonlinear}}(\phi)$, on the other hand, usually does not. One can attempt to expand $F_{\text{nonlinear}}$ when ϕ is small. For example,

$$D\phi = F_{\text{nonlinear}}(\phi) = \phi^2 + \mathcal{O}(\phi^3) \ . \tag{3.1}$$

When the amplitude is small, $|\phi| < \epsilon \ll 1$, one can solve the truncated equation of motion that includes the ϕ^2 term as a perturbative expansion of $|\phi^2/\phi| < \epsilon$ from

the linear solutions. For a small enough choice of ϵ, this can be a good enough approximation to the fully nonlinear theory. Unfortunately, this will only work for a short amount of time. After some time $T \sim \epsilon^{-1}$, the correction from the first nonlinear order accumulates and becomes comparable to the original amplitude. Thus the actual amplitude can exceed ϵ significantly to invalidate the expansion.[1]

A slightly more subtle question arises while applying such a truncated perturbation theory. Occasionally, there can be accidental cancellations while solving it. Thus during the process, the amplitude may stay below ϵ for $T \sim \epsilon^{-1}$. *In these occasions, are we then allowed to trust these solutions?*

It is very tempting to directly answer "no" to the above question. When $T \sim \epsilon^{-1}$, not only the accumulated contribution from ϕ^2, which the theory does take into account, modifies ϕ significantly. The ϕ^3 term that the theory discarded also modifies ϕ^2, and so on so forth. Since we have truncated all those even higher order terms which may have significant effects, the validity of the expansion process seems to unsalvageably break down.

The above logic sounds reasonable but it is not entirely correct. In this chapter, we will demonstrate that at exactly $T \sim \epsilon^{-1}$ time scale, the opposite is true. *These "nice" solutions we occasionally find in the truncated theory, indeed faithfully represent similar solutions in the full nonlinear theory.* This idea is not entirely new. We are certainly inspired by the application of the two-time formalism and the renormalization flow technique in the AdS-(in)stability problem, and both of them operate on this same concept [12, 21].[2] However, one may get the impression from those examples that additional techniques are required to maintain the approximation over the long time scale. One of the main points that we want to make here is that the validity of truncated theory extends *trivially* in those occasions. As long as the truncated theory is implemented recursively, which is the natural way to solve any time evolution anyway, it remains trustworthy in those occasions.

In Section 3.2, we will state and prove a theorem that guarantees a truncated perturbative expansion, implemented recursively, to approximate the full nonlinear theory accurately in the long time scale in the relevant occasions. More concretely, this theorem leads to the two following facts:

- If one solves the truncated theory and finds solutions in which the amplitude remains small during the long time scale, then *similar* solutions exist in the

[1] The notion of "time" here is just to make connections to practical problems for physicists. The general idea is valid whenever one tries to solve perturbation theory from some limited boundary conditions to a far-extended domain.

[2] We thank Luis Lehner for pointing out that some Post-Newtonian expansions to General Relativity also shows validity at this long time scale [45].

full nonlinear theory.[3]

- If numerical evolution of the full nonlinear theory provides solutions in which the amplitude remains small during the long time scale, then a truncated theory can reproduce *similar* solutions.

Formally, the meaning of "*similar*" in the above statements means that the difference between two solutions goes to zero faster than their amplitudes in the zero amplitude limit. This theorem provides a two-way bridge between numerical and analytical works. Anything of this nature can be quite useful. For example, numerical results are usually limited to finite amplitudes and times, while the actual physical questions might involve taking the limit of zero amplitude and infinite time. With this theorem, we can start from known numerical results and extend them to the limiting case with analytical techniques.

In Section 3.3, we will prove another theorem which enables us to do just that in the AdS (in)stability problem. The key is that the truncated theory does not need to be exactly solved to be useful. Since its form is simpler than the full nonlinear theory, it can manifest useful properties, such as symmetries, to facilitate further analysis. Since it is a truncated theory, the symmetry might be an approximation itself, and naively expected to break down at the long time scale. Not surprisingly, using a similar process, we can again prove that such symmetry remains trustworthy in the relevant occasions.

It is interesting to note that the conventional wisdom, which suggested an unsalvageable breakdown at $T \sim \epsilon^{-1}$, is not entirely without merits. We can prove that both theorems hold for $T = \alpha\epsilon^{-1}$ for an arbitrarily but ϵ-independent value of α. However, pushing it further to a slightly longer, ϵ-dependent time scale, for example $T \sim \epsilon^{-1.1}$, the proofs immediately go out of the window. The situation for $T \sim \epsilon^{-1}(-\ln\epsilon)$ is also delicate and will not always hold. Naturally, for time scales **longer** than $T \sim \epsilon^{-1}$, one needs to truncate the theory at an even higher order to maintain its validity. A truncated theory up to the ϕ^m term will only be valid up to $T \sim \epsilon^{1-m}$.

3.1.2 The AdS (In)Stability Problem

In Section 3.4, we will apply both theorems to the AdS-(in)stability problem [4, 8, 10–12, 14–16, 20–28, 34, 46]. Currently, the main focus of this problem is

[3]Be careful that sometimes, especially in gauge theories, the full nonlinear theory might impose a stronger constraint on acceptable initial conditions. One should start from those acceptable initial conditions in order to apply our theorem. We thank Ben Freivogel for pointing this out.

indeed the consequence of the nonlinear dynamics of gravitational self-interaction, at the time scale that the leading order expansion should generically break down. Some have tried to connect such break down to the formation of black holes and further advocate that such instability of AdS space is generic. In particular, Dias, Horowitz, Marolf and Santos made the **stability island conjecture** [15]. *Although at finite amplitudes, there are numerical evidence and analytical arguments to support measure-nonzero sets of non-collapsing solutions, they claimed that the sets of these solutions shrinks to measure zero at the zero amplitude limit.*

Since the relevant time scale goes to infinity at the zero amplitude limit, such conjecture cannot be directly tested by numerical efforts. Nevertheless, by the two theorems we prove in this chapter, it becomes straightforward to show that such conjecture is in conflict with existing evidence. The physical intuition of our argument was already outlined in [4], and here we establish the rigorous mathematical structure behind it.

- Theorem I allows us to connect non-collapsing solutions [10–12, 15] to analogous solutions in a truncated theory, both at finite amplitudes.
- Theorem II allows us to invoke the rescaling symmetry in the truncated theory and establish those solutions at arbitrarily smaller amplitudes.
- Using Theorem I again, we can establish those non-collapsing solutions in the full nonlinear theory at arbitrarily smaller amplitudes.

Thus, if non-collapsing solutions form a set of measure nonzero at finite amplitudes as current evidence implies, then they persist to be a set of measure nonzero when the amplitude approaches zero. Since the stability island conjecture states that stable solutions should shrink to sets of measure zero, it is in conflict with existing evidence.

It is important to note that defeating the stability island conjecture is not the end of the AdS (in)stability problem. Another important question is whether collapsing solutions, which likely also form a set of measure nonzero at finite amplitudes, also persist down to the zero-amplitude limit. It is easy to see why that question is harder to answer. Truncated expansions of gravitational self-interaction, at least all those which have been applied to the problem, do break down at a certain point during black hole formation. Thus, Theorem I fails to apply, one cannot establish a solid link between the truncated dynamic to the fully nonlinear one, and the AdS (in)stability problem remains unanswered.

In order to make an equally rigorous statement about collapsing solutions, one will first need to pose a weaker claim. Instead of arguing for the generality of black hole formation, one should consent with "energy density exceeding certain

bound" or something similar. This type of claim is then more suited to be studied within the validity of Theorem I, and it is also a reasonable definition of AdS instability. If arbitrarily small initial energy always evolves to have finite energy density somewhere, it is a clear sign of a runaway behavior due to gravitational attraction.[4]

Finally, we should note that the truncated theory is already nonlinear and may be difficult to solve directly. If one invokes another approximation while solving the truncated theory, such as time-averaging, then the process becomes vulnerable to an additional form of breakdown, such as the oscillating singularity seen in [48]. Even if numerical observations in some cases demonstrate a coincident between such breakdown and black hole formation, the link between them is not yet as rigorous as the standard established in this chapter for non-collapsing solutions.

3.2 Theorem I: Conditionally Extended Validity

Consider a linear space $\mathcal{H}$ with a norm satisfying triangular inequality.

$$||x+y|| \leq ||x|| + ||y|| \ , \quad \text{for all } x,\ y \in \mathcal{H} \ . \tag{3.2}$$

Then consider three smooth functions L, f, g all from $\mathcal{H}$ to itself. We require that $L(x) = 0$ if and only if $x = 0$, and it is "semi-length-preserving".

$$||L(x)|| \leq ||x|| \ . \tag{3.3}$$

Note that this condition on the length is at no cost of generality. Given any smooth function $\bar{L}$ meeting the first requirement, we can always rescale it to make it exactly length-preserving and maintaining its smoothness.

$$L(x) \equiv \frac{||x||}{||\bar{L}(x)||}\bar{L}(x) \ , \quad \text{if} \quad x \neq 0 \ ; \quad L(x) \equiv 0 \ , \quad \text{when} \quad x = 0 \ . \tag{3.4}$$

f and g are supposed to be two functions that within some radius $r < 1$, they are both close to L but even closer to each other.

1. Being close to L: $\forall\ ||x|| < r$,

$$||f(x)-L(x)|| < a||x||^m, \ \ ||g(x)-L(x)|| < a||x||^m, \quad \text{for some } a > 0 \, , \ \ m > 1 \, . \tag{3.5}$$

[4] It is then natural to believe that black hole formation follows, though it is still not guaranteed and hard to prove. For example, a Gauss-Bonnet theory can behave the same up to this point, but its mass-gap forbids black hole formation afterward [47].

And doing so smoothly: $\forall\ ||x||, ||y|| < r$ and some $b > 0$,

$$\begin{aligned} \Big|\Big|[f(x) - L(x)] - [f(y) - L(y)]\Big|\Big| &< b \cdot ||x - y|| \mathrm{Max}(||x||, ||y||)^{m-1} \ , \\ \Big|\Big|[g(x) - L(x)] - [g(y) - L(y)]\Big|\Big| &< b \cdot ||x - y|| \mathrm{Max}(||x||, ||y||)^{m-1} \ . \end{aligned} \tag{3.6}$$

2. Even closer to each other: $\forall\ ||x|| < r$,

$$||f(x) - g(x)|| < c||x||^l \ , \quad \text{for some} \ \ c > 0 \ , \ \ l > m \ . \tag{3.7}$$

Let us make the analogy to the physical problem more transparent by an example. Choose a finite time Δt to evolve the linear equation of motion $D\phi = 0$, L is given by $L[\phi(t)] = \phi(t + \Delta t)$. Similarly, evolving the full nonlinear theory $D\phi = F_{\mathrm{nonlinear}}(\phi)$ leads to a different solution ϕ that defines $f[\phi(t)] = \phi(t + \Delta t)$, and $D\phi = \phi^2$ defines $g[\phi(t)] = \phi(t + \Delta t)$. Furthermore, the norm can often be defined as the square-root of conserved energy in the linear evolution, which satisfies both the triangular inequality and the semi-preserving requirement.

From this analogy, evolution to a longer time scale is naturally given by applying these functions recursively. We will thus define three sequences accordingly.

$$f_0 = g_0 = L_0 = x \ , \quad L_n = L(L_{n-1}) \ , \quad f_n = f(f_{n-1}) \ , \quad g_n = g(g_{n-1}) \ . \tag{3.8}$$

We will prove a theorem which guarantees that after a time long enough for both g_n and f_n to deviate significantly from L_n, they can still stay close to each other.

Theorem I: For any finite $\delta > 0$ and $\alpha > 0$, there exists $0 < \epsilon < r$ such that if $||f_n|| < \epsilon$ for all $0 \le n < \alpha\epsilon^{1-m}$, then $||f_n - g_n|| < \delta\epsilon$.

Since f_n is known to be of order ϵ, thus when its difference with g_n is arbitrarily smaller than ϵ, one stays as a good approximation of the other.

Proof: First of all, we define

$$\Delta_n \equiv c\epsilon^l \sum_{i=0}^{n-1} (1 + b\epsilon^{m-1})^i = c\epsilon^l \frac{(1 + b\epsilon^{m-1})^n - 1}{b\epsilon^{m-1}} \ . \tag{3.9}$$

Within the range of n stated in the Theorem, it is easy to see that

$$\Delta_n \leq \Delta_{\lfloor \alpha\epsilon^{1-m} \rfloor} < c\epsilon^l \frac{(1+b\epsilon^{m-1})^{\alpha\epsilon^{1-m}} - 1}{b\epsilon^{m-1}} < \frac{c\epsilon^{1+l-m}}{b}\left(e^{b\alpha} - 1\right) \ . \tag{3.10}$$

Since $l > m$, there is always a choice of ϵ such that $\Delta_n < \delta\epsilon$. We will choose an ϵ small enough for that, and also small enough such that

$$||f_n|| + a\epsilon^m \ < \ \epsilon + a\epsilon^m < r \ , \tag{3.11}$$

$$||f_n|| + \Delta_n \ < \ \epsilon + \delta\epsilon < r \ . \tag{3.12}$$

Next, we use mathematical induction to prove that given such choice of ϵ,

$$||f_n - g_n|| \leq \Delta_n \ . \tag{3.13}$$

For $n = 0$, this is trivially true.

$$||f_0 - g_0|| = 0 \leq \Delta_0 = 0 \ . \tag{3.14}$$

Assume this is true for $(n-1)$,

$$||f_{n-1} - g_{n-1}|| \leq \Delta_{n-1} \ . \tag{3.15}$$

Combine it with eq. (3.7) and (3.6), we can derive for the next term in the sequence.

$$\begin{aligned} ||f_n - g_n|| \ &\leq \ ||f(f_{n-1}) - g(f_{n-1})|| + ||g(f_{n-1}) - g(g_{n-1})|| \\ &\leq \ c\epsilon^l + (1 + b\epsilon^{m-1})\Delta_{n-1} = \Delta_n \ . \end{aligned} \tag{3.16}$$

Thus by mathematical induction, we have proven the theorem.

Note that although in the early example for physical intuitions, we took f as the full nonlinear theory and g as the truncated theory, their roles are actually interchangeable in Theorem I. Thus we can use the theorem in both ways. If a fully nonlinear solution, presumably obtained by numerical methods, stays below ϵ, then Theorem I guarantees that a truncated theory can reproduce such solution. The reverse is also true. If the truncated theory leads to a solution that stays below ϵ, then Theorem I guarantees that this is a true solution reproducible by numerical evolution of the full nonlinear theory.

Also note that the truncated theory might belong to an expansion which does not really converge to the full nonlinear theory. This is quite common in field theories that a naïve expansion is only asymptotic instead of convergent. Theorem I does not care about whether such full expansion is convergent or not. It only requires that the truncated theory is a good approximation to the full theory up to some

specified order, as stated in eq. (3.7). Divergence of an expansion scheme at higher orders does not invalidate our result.[5]

Finally, if one takes a closer look at eq. (3.9), one can see that if n is allowed to be larger than the ϵ^{1-m} time scale, for example $n \sim \epsilon^{-s}$ with $s > m-1$, then Δ_n fails to be bounded from above in the $\epsilon \to 0$ limit. Since the upperbound we put is already quite optimal, we believe that the truncated theory does break down at any longer time scale. In particular, this does not care about l. Namely, independent of how small the truncated error is, accumulation beyond the ϵ^{1-m} time scale always makes the truncated dynamic a bad approximation for the full theory. Thus, the conventional wisdom only requires a small correction. *Usually, the truncated theory breaks down at the ϵ^{1-m} time scale. Occasionally, it can still hold at exactly this time scale but breaks down at any longer time scale.*

3.3 Theorem II: Conditionally Preserved Symmetry

We will consider an example that the truncated theory has an approximate scaling symmetry. Let $L(x) = x$, $g(x) = L(x) + G(x)$, such that for all $||x||, ||y|| < r$,

$$\begin{aligned} ||G(x)|| &< a||x||^m \ , & (3.17)\\ ||G(x) - G(y)|| &< b \cdot ||x-y||\mathrm{Max}(||x||, ||y||)^{m-1} \ , & (3.18)\\ ||G(x) - N^m G(x/N)|| &< d||x||^p \ , & (3.19) \end{aligned}$$

for a given $p > m$ and any $N > 1$. Namely, the linear theory is trivial that $L_n = x$ does not evolve with n. The only evolution for g_n comes from the function $G(x)$, which is for many purposes effectively "an x^m term". In this case, it is reasonable to expect a rescaling symmetry: reducing the amplitude by a factor of N, but evolve for a time longer by a factor of N^{m-1}, leads to roughly the same result.

Theorem II: For any finite $\delta > 0$ and $\alpha > 0$, there exists $0 < \epsilon < r$ such that if $||g_n(x)|| < \epsilon$ for all $0 \le n < \alpha\epsilon^{1-m}$, then

$$||N g_n(x/N) - (1-\beta) g_{n'}(x) - \beta g_{n'+1}(x)|| < \delta\epsilon \ . \qquad (3.20)$$

Here $n' = \lfloor (nN^{1-m}) \rfloor$ is the largest integer smaller than or equal to (nN^{1-m}), and $\beta = (nN^{1-m}) - n'$. This should be valid for any $N > 1$ and for $0 \le n <$

[5] We thank Jorge Santos for pointing out the importance to stress this point.

$\alpha(\epsilon/N)^{1-m}$.

The physical intuition is the following. Every term in the rescaled sequence stays arbitrarily close to some weighted average of the terms in the original sequence, which exactly corresponds to the appropriate "time" (number of steps) of the rescaling. We will first prove this for a special case, $N = 2^{\frac{1}{m-1}}$. This case is particularly simple, since such rescaling exactly doubles the length of the sequence, and β will be either 0 or 1/2 which leads to two specific inequalities to prove:

$$\left|\left|2^{\frac{1}{m-1}} g_{2n-1}(x/2^{\frac{1}{m-1}}) - \frac{g_{n-1}(x) + g_n(x)}{2}\right|\right| \leq C \cdot \epsilon^q \,, \tag{3.21}$$

$$\left|\left|2^{\frac{1}{m-1}} g_{2n}(x/2^{\frac{1}{m-1}}) - g_n(x)\right|\right| \leq C \cdot \epsilon^q \,, \tag{3.22}$$

for some $C > 0$ and $q > 1$. This will again be done by a mathematical induction.

During the process, it should become clear that the proof can be generalized to any $N > 1$. We will not present such proof, because the larger variety of β values makes it more tedious, although it is still straightforward. However, for self-completeness what we need in the next session is only that N can be arbitrarily large. Through another mathematical induction, we can easily prove Theorem II for $N = 2^{\frac{k}{m-1}}$ for arbitrarily positive integer k. It is still a bit tedious, so we will present that in Appendix 3.A.

Proof for $\mathbf{N = 2^{\frac{1}{m-1}}}$

We start by defining the monotonically increasing function

$$\Delta_n = \left(\frac{d}{b} 2^{m-1} \epsilon^{p-m+1} + \frac{a}{2}\epsilon^m\right)\left[\left(1 + \frac{b}{2}(2\epsilon)^{m-1}\right)^n - 1\right], \tag{3.23}$$

with the properties

$$\begin{aligned} \Delta_n &< \Delta_{\alpha(\epsilon/N)^{1-m}} \\ &= \left(\frac{d}{b} 2^{m-1} \epsilon^{p-m+1} + \frac{a}{2}\epsilon^m\right)\left[\left(1 + \frac{b}{2}(2\epsilon)^{m-1}\right)^{\alpha(\epsilon/N)^{1-m}} - 1\right] \end{aligned} \tag{3.24}$$

$$
\begin{aligned}
&< \left(\frac{d}{b}2^{m-1}\epsilon^{p-m+1}+\frac{a}{2}\epsilon^m\right)\left[e^{2^{m-1}\alpha b/4}-1\right] < C\cdot\epsilon^q \ , \\
\Delta_{n+1} &= \Delta_n\left(1+\frac{b}{2}(2\epsilon)^{m-1}\right)+\frac{d}{2}\epsilon^p+\frac{ab}{4}\epsilon^m(2\epsilon)^{m-1} \ . \qquad (3.25)
\end{aligned}
$$

The meaning of eq. (3.24) is that, in the range we care about, Δ_n is bounded from above by a power of ϵ higher than one, since $q = \text{Min}\{p-m+1, m\}$. Given that, we can always choose ϵ small enough such that

$$
\big|\big|g_n(x)\big|\big| + \Delta_n \quad < \quad \epsilon + C\cdot\epsilon^q < r \ . \qquad (3.26)
$$

Given our choice of ϵ, we can employ mathematical induction to prove that

$$
\begin{aligned}
\Big|\Big|Ng_{2n-1}(x/N) - \frac{g_{n-1}(x)+g_n(x)}{2}\Big|\Big| &\leq \Delta_{2n-1} \qquad (3.27)\\
\Big|\Big|Ng_{2n}(x/N) - g_n(x)\Big|\Big| &\leq \Delta_{2n} \ , \qquad (3.28)
\end{aligned}
$$

which prove eq. (3.21) and (3.22).

First, we observe that for $n = 0$,

$$
\Big|\Big|Ng_0(x/N) - g_0(x)\Big|\Big| = \Big|\Big|n\frac{x}{N} - x\Big|\Big| = 0 < C\cdot\epsilon^q \ , \qquad (3.29)
$$

is obviously true. Then, we assume that eq. (3.28) is true for n in the original sequence and $2n$ in the rescaled sequence. We can prove for the next term, the $(2n+1)$ term in the rescaled sequence.

$$
\begin{aligned}
&\Big|\Big|Ng_{2n+1}(x/N) - \frac{g_n(x)+g_{n+1}(x)}{2}\Big|\Big| \\
= &\Big|\Big|Ng_{2n}(x/N) + NG\big(g_{2n}(x/N)\big) - g_n(x) - \frac{1}{2}G\big(g_n(x)\big)\Big|\Big| \\
= &\Big|\Big|Ng_{2n}(x/N) + NG\big(g_{2n}(x/N)\big) - g_n(x) - \frac{1}{2}G\big(g_n(x)\big) \\
+ &\ NG\big(g_n(x)/N\big) - NG\big(g_n(x)/N\big)\Big|\Big| \\
< &\Big|\Big|Ng_{2n}(x/N) - g_n(x)\Big|\Big| + N\Big|\Big|G\big(g_{2n}(x/n)\big) - G\big(g_n(x)/N\big)\Big|\Big| \\
+ &\ \frac{1}{N^{m-1}}\Big|\Big|G\big(g_n(x)\big) - N^m G\big(g_n(x)/N\big)\Big|\Big| \\
< &\ \Delta_{2n} + \Delta_{2n}\frac{b}{N^{m-1}}\left(\big|\big|\Delta_{2n}\big|\big| + \big|\big|g_n(x)\big|\big|\right)^{m-1} + \frac{d}{2}\epsilon^p \\
< &\ \Delta_{2n} + \Delta_{2n}\frac{b}{2}\left(\Delta_{2n}+\epsilon\right)^{m-1} + \frac{d}{2}\epsilon^p \\
< &\ \Delta_{2n} + \Delta_{2n}\frac{b}{2}(2\epsilon)^{m-1} + \frac{d}{2}\epsilon^p < \Delta_{2n+1}. \qquad (3.30)
\end{aligned}
$$

Similarly we can prove for the $(2n+2)$ term in the rescaled sequence, which is the $(n+1)$ term in the original sequence.

$$\begin{aligned}
& \Big|\Big| N g_{2n+2}(x/N) - g_{n+1}(x) \Big|\Big| \\
= \; & \Big|\Big| N g_{2n+1}(x/N) + NG\big(g_{2n+1}(x/N)\big) - \frac{1}{2} g_{n+1}(x) \\
- \; & \frac{1}{2} g_n(x) - \frac{1}{2} G\big(g_n(x)\big) + NG\left(\frac{g_n(x)+g_{n+1}(x)}{N^m}\right) - NG\left(\frac{g_n(x)+g_{n+1}(x)}{N^m}\right) \\
+ \; & NG\left(\frac{g_n(x)}{N}\right) - NG\left(\frac{g_n(x)}{N}\right) \Big|\Big| \\
< \; & \Big|\Big| N g_{2n+1}(x/N) - \frac{g_{n+1}(x)+g_n(x)}{2} \Big|\Big| \\
+ \; & N \Big|\Big| G\big(g_{2n+1}(x/N)\big) - G\left(\frac{g_n(x)+g_{n+1}(x)}{N^m}\right) \Big|\Big| \\
+ \; & N \Big|\Big| G\left(\frac{g_n(x)+g_{n+1}(x)}{N^m}\right) - G\left(\frac{g_n(x)}{N}\right) \Big|\Big| + N \Big|\Big| G\left(\frac{g_n(x)}{N}\right) - \frac{1}{N^{m-1}} G\big(g_n(x)\big) \Big|\Big| \\
< \; & \Delta_{2n+1} + \Delta_{2n+1} \frac{b}{2} (2\epsilon)^{m-1} + \frac{d}{2}\epsilon^p + \frac{ab}{4}\epsilon^m (2\epsilon)^{m-1} = \Delta_{2n+2} \; . \qquad (3.31)
\end{aligned}$$

This completes the mathematical induction.

eq. (3.25) takes basically the same form as eq. (3.9). Thus, Theorem II also only holds up to exactly the ϵ^{1-m} time scale, but not any longer.

3.4 Application: Persistence of Stability Islands

First, we review the "stability island conjecture" argued by Dias, Horowitz, Marolf and Santos in [15]. Numerical simulations suggest that given a small but finite initial amplitude $\phi_{\text{init.}} \sim \epsilon$ in AdS space with Dirichlet boundary condition, dynamical evolution can lead to black hole formation at the long time scale $T \sim \epsilon^{-2}$ [8][6]. In the meanwhile, some initial conditions do not lead to black holes at the same time scale. In particular, there are special solutions (set of measure zero) which not only do not collapse, they stay exactly as they are. These especially stable solutions are called geons (in pure gravity) or Boson-stars/Oscillons (scalar field) [10, 14, 20][7]. At finite amplitudes, they are not only stable themselves, but also stabilize an open neighborhood in the phase space, forming stability islands which prevent nearby initial conditions from collapsing into black holes in the $\sim \epsilon^{-2}$ time scale.

[6] Note that for this purpose, $m = 3$, thus ϵ^{-2} is the relevant time scale.

[7] There are also quasi-periodic solutions which do not stay exactly the same but demonstrate a long-term periodic behavior and the energy density never gets large [12].

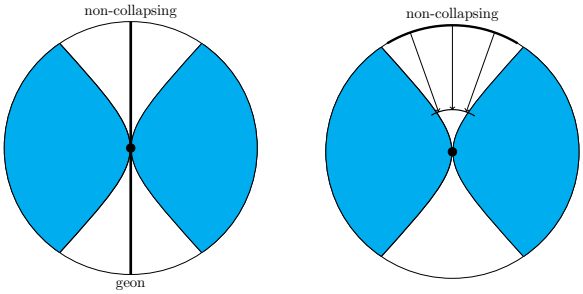


Figure 3.1: *Phase space diagrams of small initial perturbations around empty AdS (central black dot) according to the stability island conjecture. The radial direction represents field amplitude (total energy), and the angular direction represents field profile shape (energy distribution). Initial perturbations in the shaded (blue) region will collapse into black holes at the $\sim \epsilon^{-2}$ time scale, while the unshaded region, around the exactly stable geons (thick black line), will not. The unshaded region is cuspy, showing that according to the conjecture, the angular span of non-collapsing perturbations goes to zero as amplitude goes to zero. The right panel demonstrate the usage of both Theorems we proved in this chapter. We can transport the known, non-collapsing solutions, directly in the radial direction, to an arc of identical angular span at an arbitrarily smaller radius. It is a direct contradiction to the cuspy nature of the unshaded region.*

Dias, Horowitz, Marolf and Santos argued that such stabilization effect can be understood as breaking the AdS resonance.[8] It should lose strength as the geon's own amplitude decreases. Thus, such stability islands go away in the limit of zero amplitude. The easiest way to summarize their conjecture is by the cuspy phase-space diagram in Fig. (3.1). Other than the measure-zero set of exact geons/Boson-stars, non-collapsing solutions at finite amplitude will all end up collapsing as amplitude goes to zero.

Next, we will show that the requirements of both Theorem I and II are applicable to the AdS (in)stability problem. For simplicity, we present the analysis on a massless scalar field in global AdS space of Dirichlet boundary condition. Metric fluctuation in pure gravity will also meet the requirements [15, 16]. We will avoid going into specific details of the AdS dynamics, but only provide the relevant works where those can be found.

- The linear space $\mathcal{H}$ we used to state both Theorems (see the beginning of

[8] In some sense, this argument [15] provides a stronger support for non-collapsing solutions to have a nonzero measure, because it goes beyond spherical symmetry. Current numerical results are limited to spherical symmetry, thus strictly speaking cannot establish a nonzero measure for either collapsing or non-collapsing results. This is why controversies over some numerical results [29,49] should not undermine the belief that non-collapsing solutions form a set of nonzero measure at finite amplitude.

Sec.3.2) contains all smooth functions $\phi(\vec{r})$ on the domain of the entire spatial slice of the AdS space at one global time between.

- The function L evolves one such function forward for one "*AdS* period", namely $T = 2\pi R_{\rm AdS}$ in the explanation right below eq. (3.7), using the fixed background equation of motion. It includes no gravitational self-interaction and is a linear function. Actually, since the AdS spectrum has integer eigenvalues, the evolution is exactly periodic [50, 51]. $L(x) = x$ is trivial, automatically conserves length and also meets the requirement for Theorem II.

- The definition of the norm is trickier. We first evolve x, using the fixed-background evolution, for exactly $2\pi R_{\rm AdS}$, and examine the maximum local energy density ever occurred during such evolution. The norm is defined to be the square-root of this value, $\sqrt{\rho_{\rm Max}}$. The evolution is linear, and the quantity is both a maximum and effectively an absolute value, thus it satisfies the triangular inequality.[9]

- The actual dynamic, including Einstein equations, is clearly nonlinear. When the maximum energy density is always small, the gravitational back-reaction is well-bounded. One can perform a recursive expansion in which the leading order correction to the linear dynamic comes from coupling to its own energy, $\rho\phi \propto \phi^3$ [4, 8, 33]. A theory truncated at this order and the full nonlinear theory can be our f and g, interchangeably, in Theorem I with $m = 3$. [10]

- The ϕ^3 contribution calculated in different approximation methods might be different [4, 8, 33], but they all satisfy the approximate rescaling symmetry required for the function G in Theorem II.

Now we have established the applicability of both Theorems in this chapter, the stability island conjecture can be disproved in three simple steps.

1. At a small but finite amplitude where measure nonzero sets of non-collapsing solutions exist (the outermost thick arc in Fig. (3.1), apply Theorem I to translate them into solutions in the truncated theory.

2. Use Theorem II to scale down the above solutions to arbitrarily small amplitudes. That means projecting radially in Fig. (3.1) into an arc of the same angular span.

[9] The reason why we adopt this tortuous definition of norm is to guarantee that gravitational interaction during one AdS time stays weak when the norm is small, thus we can apply both theorems. Note that defining total energy as the norm would fail such purpose.

[10] Such expansion, continued to higher orders, is likely only asymptotic instead of convergent. As explained in Sec.3.2, that does not cause a problem for our theorems.

3. Use Theorem I again to translate these rescaled solution in the truncated theory back to the full nonlinear theory. This establishes the existence of non-collapsing solutions as a set of measure non-zero (an arc of finite angular span in Fig. (3.1). [11]

Thus, we have established that the measure-non-zero neighborhood stabilized by a geon at finite amplitude, if never evolves to high local energy density during the long time scale, directly guarantees the same measure-non-zero, non-collapsing neighborhood at arbitrarily smaller amplitudes. This directly contradicts the stability island conjecture.

It is interesting to note that the collapsing solutions always have energy density large at a certain point, thus neither theorems we proved here are applicable to them. As a result, one cannot establish their existence at arbitrarily small amplitudes through a similar process. Therefore, the opposite possibility to the stability island conjecture, that collapsing solutions disappear into a set of measure zero at zero amplitude, is still consistent with current evidence.

3.A Arbitrarily small rescaling

In Sec.3.3, we have proven Theorem II for $N = 2^{\frac{1}{m-1}}$. Now, we will generalize it to arbitrary $N' = 2^{\frac{k}{m-1}} = N^k$, for any $k \in \mathbb{N}^+$:

$$\left|\left|N^k g_n(x/N^k) - (1-\beta_k(n))\, g_{\lfloor \frac{n}{2^k} \rfloor}(x) - \beta_k(n) g_{\lfloor \frac{n}{2^k} \rfloor + 1}(x)\right|\right| \leq C' \cdot \epsilon^q, \quad (3.32)$$

where we have written down explicitly the dependence of β on k and n:

$$\beta_k(n) = \frac{2}{n^k} - \left\lfloor \frac{n}{2^k} \right\rfloor, \quad (3.33)$$

which possesses the following properties for positive integers j and l:

$$\beta_{k+1}(2l) = \beta_k(l)\ , \qquad \text{this is always true ;} \quad (3.34)$$

$$\beta_{k+1}(2l+1) = \frac{1}{2}\beta_k(l) + \frac{1}{2}\beta_k(l+1), \quad \text{for } l+1 \neq j \cdot 2^k\ ; \quad (3.35)$$

$$\beta_{k+1}(2l+1) = \frac{1}{2}\beta_k(l) + \frac{1}{2}[1-\beta_k(l+1)], \quad \text{for } l+1 = j \cdot 2^k\ . \quad (3.36)$$

These follow naturally from the properties of the floor function that

$$\left\lfloor \frac{2l+1}{2^{k+1}} \right\rfloor = \left\lfloor \frac{l}{2^k} \right\rfloor , \qquad \text{is always true ,} \quad (3.37)$$

[11] The only works for rescaling down to smaller amplitudes. Rescaling to larger amplitudes can easily exceed the radius of validity of perturbative expansion even at short time scales.

and

$$\left\lfloor \frac{2l+1}{2^{k+1}} \right\rfloor = \left\lfloor \frac{l+1}{2^k} \right\rfloor, \quad \text{when} \quad l \neq j \cdot 2^k - 1 \; ; \tag{3.38}$$

$$\left\lfloor \frac{2l+1}{2^{k+1}} \right\rfloor = \left\lfloor \frac{l+1}{2^k} \right\rfloor - 1, \quad \text{when} \quad l = j \cdot 2^k - 1 \; . \tag{3.39}$$

We now define:

$$F_k \equiv C \cdot \epsilon^q \sum_{i=0}^{k} N^{i(1-q)} = \frac{C \cdot \epsilon^q}{1 - N^{1-q}} \left(1 - N^{k(1-q)}\right) \leq C' \cdot \epsilon^q \; , \tag{3.40}$$

for $C' = C/(1 - N^{1-q})$. This converges as $k \to \infty$, since $1 - q < 0$, and satisfies the recursive relation:

$$F_{k+1} = F_k + C \cdot N^{k(1-q)}. \tag{3.41}$$

Now we will prove eq. (3.32) by induction. We have already shown that it holds for $k = 1$ in Sec.3.3, hence assuming that it holds for arbitrary k, we want to show that it holds for $k + 1$ as well.

It is helpful to split the proof in three parts, one for $n = 2l$, one for $n = 2l + 1$, with $l \neq j \cdot 2^k - 1$ and one for $n = 2l + 1$, with $l = j \cdot 2^k - 1$.

1. *Part 1:* $n = 2l$

$$\begin{aligned}
& \left|\left| N^{k+1} g_{2l}(x/N^{k+1}) - (1 - \beta_{k+1}(2l))\, g_{\lfloor \frac{2l}{2^{k+1}} \rfloor}(x) - \beta_{k+1}(2l) g_{\lfloor \frac{2l}{2^{k+1}} \rfloor + 1}(x) \right|\right| \\
= & \left|\left| N \cdot N^k g_{2l}\left(\frac{x/N^k}{N}\right) - (1 - \beta_k(l))\, g_{\lfloor \frac{2l}{2^{k+1}} \rfloor}(x) - \beta_k(l) g_{\lfloor \frac{2l}{2^{k+1}} \rfloor + 1}(x) \right|\right| \\
< & N^k \left|\left| N g_{2l}\left(\frac{x/N^k}{N}\right) - g_l\left(x/N^k\right) \right|\right| \\
+ & \left|\left| N^k g_l\left(x/N^k\right) - \left(1 - \beta_k(l)\right) g_{\lfloor \frac{l}{2^k} \rfloor}(x) - \beta_k(l) g_{\lfloor \frac{l}{2^k} \rfloor + 1}(x) \right|\right| \\
< & C \cdot N^{k(1-q)} + F_k = F_{k+1}.
\end{aligned} \tag{3.42}$$

2. *Part 2:* $n = 2l + 1$, *with* $l \neq j \cdot 2^k - 1$

$$\begin{aligned}
& \left|\left| N^{k+1} g_{2l+1}(x/N^{k+1}) - (1 - \beta_{k+1}(2l+1))\, g_{\lfloor \frac{2l+1}{2^{k+1}} \rfloor}(x) \right.\right. \\
- & \left.\left. \beta_{k+1}(2l+1) g_{\lfloor \frac{2l+1}{2^{k+1}} \rfloor + 1}(x) \right|\right|
\end{aligned}$$

$$
\begin{aligned}
&= \Big|\Big|N\cdot N^k g_{2l+1}\left(\frac{x/N^k}{N}\right) - \left(1-\frac{1}{2}\beta_k(l)-\frac{1}{2}\beta_k(l+1)\right) g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor}(x)\\
&- \frac{1}{2}\left(\beta_k(l)+\beta_k(l+1)\right) g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor+1}(x)\Big|\Big|\\
&< N^k\Big|\Big|Ng_{2l+1}\left(\frac{x/N^k}{n}\right) - \frac{g_l(x/N^k)+g_{l+1}(x/N^k)}{2}\Big|\Big|\\
&+ \frac{1}{2}\Big|\Big|N^k g_l(x/N^k) - (1-\beta_k(l))\, g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor}(x) - \beta_k(l) g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor+1}(x)\Big|\Big|\\
&+ \frac{1}{2}\Big|\Big|N^k g_{l+1}(x/N^k) - (1-\beta_k(l+1))\, g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor}(x) - \beta_k(l+1) g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor+1}(x)\Big|\Big|\\
&< C\cdot N^{k(1-q)} + 2\frac{1}{2}F_k = F_{k+1}. \qquad (3.43)
\end{aligned}
$$

3. *Part 3:* $n = 2l+1$, *with* $l = j\cdot 2^k - 1$

$$
\begin{aligned}
&\Big|\Big|N^{k+1} g_{2l+1}(x/N^{k+1}) - (1-\beta_{k+1}(2l+1))\, g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor}(x)\\
&- \beta_{k+1}(2l+1) g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor+1}(x)\Big|\Big|\\
&= \Big|\Big|N\cdot N^k g_{2l+1}\left(\frac{x/N^k}{N}\right) - \left(1-\frac{1}{2}\beta_k(l)-\frac{1}{2}\big(1-\beta_k(l+1)\big)\right) g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor}(x)\\
&- \frac{1}{2}\left(\beta_k(l)+\big(1-\beta_k(l+1)\big)\right) g_{\lfloor\frac{2l+1}{2^{k+1}}\rfloor+1}(x)\Big|\Big|\\
&< N^k\Big|\Big|Ng_{2l+1}\left(x/N^k\right) - \frac{g_l(x/N^k)+g_{l+1}(x/N^k)}{2}\Big|\Big|\\
&+ \frac{1}{2}\Big|\Big|N^k g_l(x/N^k) - \beta_k(l) g_{\lfloor\frac{l}{2^k}\rfloor+1}(x) - \big(1-\beta_k(l)\big) g_{\lfloor\frac{l}{2^k}\rfloor}(x)\Big|\Big|\\
&+ \frac{1}{2}\Big|\Big|N^k g_{l+1}(x/N^k) - \beta_k(l+1) g_{\lfloor\frac{l+1}{2^k}\rfloor-1}(x) - \big(1-\beta_k(l+1)\big) g_{\lfloor\frac{l+1}{2^k}\rfloor}(x)\Big|\Big|\\
&< C\cdot N^{k(1-q)} + 2\frac{1}{2}F_k = F_{k+1}. \qquad (3.44)
\end{aligned}
$$

We have used here the fact:

$$\beta_k(l+1) g_{\lfloor\frac{l+1}{2^k}\rfloor-1}(x) = \beta_k(l+1) g_{\lfloor\frac{l+1}{2^k}\rfloor+1}(x), \qquad (3.45)$$

since $\beta_k(j\cdot 2^k) = 0$.

4 Gauge dependence of the AdS (in)stability problem

How wonderful that we have met with a paradox. Now we have some hope of making progress.

— Niels Bohr

4.1 Introduction

The two-time framework (TTF) is a well-established tool that reduces the full gravitational dynamics into the "slow-time" evolution of complex amplitudes of approximate eigenstates. It operates on two approximations:

- The deviation from empty AdS metric is small, so we can keep only the leading order gravitational back-reaction.
- The evolution can be averaged over a "fast" time scale set by the AdS radius, reducing to the dynamics in a "slow" time scale.

One can simply follow the two-time evolution and observe whether the first approximation breaks down. If it does not, then the metric stays near empty AdS and an instability is not triggered. If it does break down, then it implies an order one deviation from empty AdS, thus triggering an instability.

In the previous chapter we saw how one can use the perturbative schemes, such as TTF, to establish the persistence of a non-zero measure of stable solutions in the limit of infinitesimal amplitude $\epsilon \to 0$. One might want to use these approximate systems to do a similar statement for the phase space of collapsing solutions in the same limit. This case is however a little bit trickier, since we have to be careful on how to interpret the breakdown of our approximations. For example if one witnesses a breakdown of TTF, he can not be directly certain whether it is the *weak gravity* or some other approximation that breaks down.

In [48], numerical results suggested that gravitational instability seems to coincide with a breakdown of TTF from an oscillating singularity—the complex amplitudes all start to acquire infinite phases. However, a direct logical link between the two was missing, because the physical interpretation of the oscillating singularity remained unclear. That is because a breakdown of TTF could be due to failure of either one of the two approximations, but only the breakdown of the first approximation has direct implications for the instability.[1] Later, in [52], was suggested that TTF might not suffer from an oscillating singularity if one chooses a different gauge, a fact that was subsequently verified numerically in [53]. Those results appeared to add more confusion.

In this chapter, after we summarize the results of [48] and [52, 53], we point out that their combination actually eliminates the confusion. A diverging difference[2] between the results in two different gauges implies a diverging redshift between two different locations in AdS, which in turn implies a diverging deviation in the metric. Alternatively, one could have used only the result in the boundary gauge where the TTF solutions stays finite [53]. Explicitly calculating its geometric back-reaction demonstrates the same divergence [54].

Note that the actual geometric back-reaction is the TTF result multiplied by the amplitude squared of the initial perturbation. A diverging TTF redshift means that linearized gravity breaks down for arbitrarily small initial amplitude, which triggers a genuine instability of global AdS.

4.2 An oscillating singularity

It was already stated in chapter 1 that in the ansatz eq. (1.3) there is some remaining gauge freedom, which can be eliminated by choosing a normalization condition for the function $\delta(x,t)$. In [48], the *two-mode, equal energy data* in AdS_5 were numerically evolved both in full GR, as well as in the TTF system. The authors of this work chose the normalization $\delta(0,t)=0$, which is commonly dubbed as the *interior time gauge*. With this choice, t corresponds to the proper time at the center of the spacetime.

The initial data evolved are

$$\phi(x,0)=\epsilon\left(\frac{1}{4}e_0(x)+\frac{1}{6}e_1(x)\right),\quad \Pi(x,0)=0, \tag{4.1}$$

[1] Some may have the intuition that the breakdown of the second approximation also can only come from large deviations from the AdS metric, but such statement is never proven explicitly.

[2] We will specify what this means in the subsequent sections.

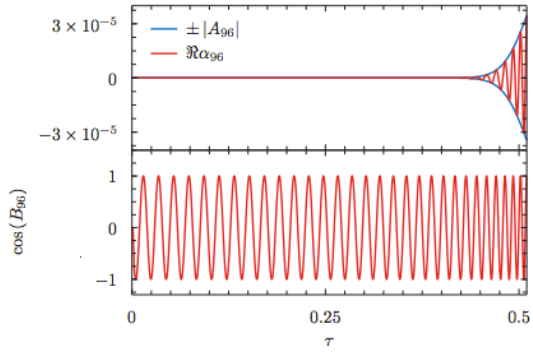


Figure 4.1: *Amplitude spectrum A_n, for different values of τ. Solid lines correspond to a 100 mode truncation, while dashed lines correspond to a 200 mode truncation. [48]*

which they correspond to the following mode amplitudes:

$$\alpha_0(0) = 1/8, \quad \alpha_1(0) = 1/12, \quad \alpha_{n\geq 2}(0) = 0. \tag{4.2}$$

The system was truncated at $n_{max} = 175$ modes and the results are presented in Fig. (4.1). The outcome is that an oscillating singularity was observed. This can be understood better in the so-called *amplitude-phase representation*, $\alpha_n = A_n(t)e^{iB_n(t)}$. Evolving the aformentioned data leads to a logarithmic blow-up of the *slow time* derivative of the phases at some finite slow time $\tau_\star \sim 0.5$.

To give better evidence for the results, we can employ the *analyticity strip method* [57], which uses Fourier asymptotics to diagnose the formation of a singularity. Let us consider an evolution equation with a solution denoted by $\alpha(t, x)$. If we analytically continue this solution to the complex plane, then this analytically continued solution $\alpha(t, z)$ will typically have complex singularities. If a complex singularity hits the real axis in finite time, then $\alpha(t, x)$ becomes singular. This can become more rigorous if we denote the location of the pair of complex singularities closest to the real axis by $z = x \pm i\rho$, so that ρ denotes the width of the *analyticity strip* around the real axis. Using this notation we can say that $\alpha(t, x)$ *becomes singular if* $\rho(t)$ *vanishes in finite time*. The analyticity strip width $\rho(t)$ is encoded in an exponential decay of the asymptotic Fourier coefficients of the solution $\alpha(t, x)$. In the problem at hand, this consists of making the following asymptotic ansatz for the amplitudes

$$A_n(\tau) \sim n^{-\gamma(\tau)} e^{-\rho(\tau)n} \tag{4.3}$$

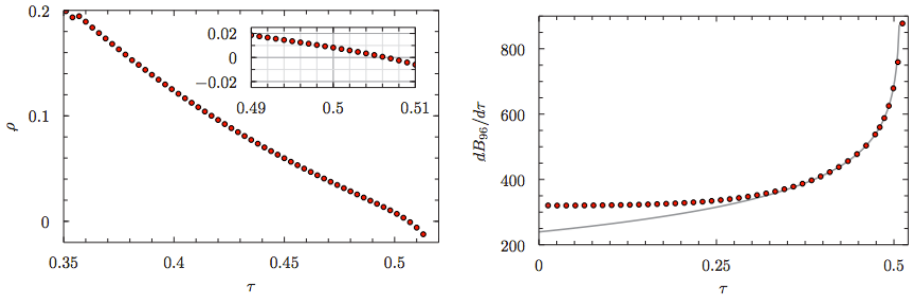


Figure 4.2: *The radius of analyticity $\rho(\tau)$ obtained by fitting eq. (4.3) to the data (***left***) and numerical evidence for the logarithmic blowup of the derivative for the sample phase B_{96} (***right***) [48]*

and fitting this formula to the data we can obtain the time dependence of the *pawer law* $\gamma(\tau)$ and the *analyticity radius* $\rho(\tau)$. The numerical results of [48] suggest that $\rho(\tau)$ tends to zero at finite slow time $\tau_\star$ and the value of γ approaches $\gamma \to 2$ at the same time. Therefore one can assume the following behavior for $\tau \to \tau_\star$

$$A_n(\tau) \sim n^{-2} e^{-\rho_0(\tau_\star - \tau)}. \tag{4.4}$$

Plugging in this result and solving for the phases yields

$$\frac{dB_n}{d\tau} \sim \sum_{j \neq n} R_{jn} A_j^2 \sim n^2 \ln(\tau - \tau_\star), \tag{4.5}$$

which blows up logarithmically at the same time $\tau_\star$ at which the analyticity strip width $\rho(\tau)$ becomes zero. In Fig. (4.2) numerical evidence that supports this analysis is presented.

This result however is **gauge dependent**. As it was first argued in [52] and it was later verified numerically in [53], this oscillating singularity is not present in the *boundary time gauge*, defined by $\delta(\pi/2, t) = 0$.

As can be seen in Fig. (4.3), the derivatives of the phases don't blow up in the boundary gauge. The reason for this different behavior in the two gauges can be traced back in the *slighlty* different behavior of the interaction coefficients S_{ijkl} in the two gauges [52, 55]

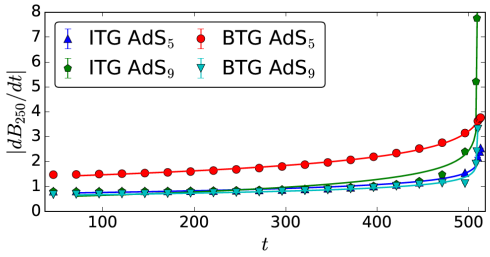


Figure 4.3: *Numerical evidence that the logarithmic divergence of the phases is a* gauge dependent *effect. [53]*

4.3 Comparing the two gauges

In this section we compare the results in the two gauges and we show that within the validity of TTF, they indeed describe the same physical evolution. The relation between the two gauges has also been studied in [56] and some of the results can be found there as well. We will follow similar notations, but our attention lies on oscillating singularities that occur in one gauge and not the other. With some extra care we show what goes wrong as TTF breaks down when such a singularity develops in the central gauge.

The gauge choice should not affect any physical quantities. However, the two different gauges do lead to two different sets of differential equations, which were numerically evaluated to very different results. In [48] the case of the two–mode equal energy data in AdS_5 was studied and an oscillating singularity was reported. Namely, the derivatives of the phases blow up. In [53] it was shown that this singularity does not appear in the boundary gauge and therefore the singular behaviour of the system might be only an artefact of the gauge choice.

On top of just numerical results, one can also see this difference from the asymptotic scaling of the R_{ij} coefficients as was first suggested in [52]. It was shown that for AdS_5 the R_{ij} coefficients scale in the central gauge as $R_{ij}^{CG} \sim i^3 j^2$ and therefore, for a *power-law* spectrum $A_n \sim n^{-2}$ as observed in [48], the sum in the second term of eq. (1.34) diverges logarithmically. On the other hand, the asymptotic scaling of these coefficients in the boundary gauge was shown to be $R_{ij}^{BG} \sim i^2 j^2$, thus although the evolution leads to the same power-law spectrum the same sum converges. One can check that the rest of the sums do not diverge.

Despite this apparent difference, these results do not contradict each other. The oscillating singularity observed in [48], combined with the absence of that in [53], has an obvious physical meaning. It implies an infinite gravitational redshift be-

tween the boundary and the center of the spacetime.

From the metric (1.3), one can see that the two gauge choices are related as:

$$dt_{BG} = e^{-\delta\left(t_{CG},\frac{\pi}{2}\right)} dt_{CG}. \tag{4.6}$$

Integrating and keeping terms only up to order $\mathcal{O}(\epsilon^2)$ we get :

$$t_{BG} = t_{CG} - \epsilon^2 \int_0^{t_{CG}} dt \delta_2(t,\tau,0) + \mathcal{O}(\epsilon^4) \tag{4.7}$$

Neglecting terms that oscillate in the fast time scale t, we can approximate $\delta_2(t,\tau,0)$ by the *time averaged* quantity $\delta_2(\tau,0)$. For completeness we will present the computation of this quantity in section (4.3.1). We then get:

$$\begin{aligned} t_{BG} &\approx t_{CG} - \epsilon^2 \int_0^{t_{CG}} dt \delta_2(\tau,0) + \mathcal{O}(\epsilon^4) \\ &= t_{CG} + 2\epsilon^2 \int_0^{t_{CG}} dt \sum_j \left(A_{jj} + \omega_j^2 V_{jj}\right) A_j^2 + \mathcal{O}(\epsilon^4), \end{aligned} \tag{4.8}$$

Now, using the fact that the field $\phi(t,\tau,x)$ transforms as a scalar under such a *gauge transformation* one can derive the relation for the complex coefficients $\alpha_j(\tau)$ in the two gauges from eq. (1.11):

$$\begin{aligned} \phi^{CG}(t_{CG}) &= \phi^{BG}(t_{BG}) \Rightarrow \\ \alpha_j^{CG}(\tau_{CG})\, e^{i\omega_j t_{CG}} &= \alpha_j^{BG}(\tau_{BG})\, e^{i\omega_j t_{BG}} \end{aligned} \tag{4.9}$$

The relation of the slow time in the gauges is obtained simply by multiplying eq. (4.8) by ϵ^2 to obtain

$$\tau_{BG} = \tau_{CG} + 2\epsilon^2 \int_0^{\tau_{CG}} d\tau \sum_j \left(A_{jj} + \omega_j^2 V_{jj}\right) A_j^2 + \mathcal{O}(\epsilon^4) \tag{4.10}$$

Substituting in the right hand side of the above equation, Taylor expanding and neglecting terms that are of order $\mathcal{O}(\epsilon^2)$ we obtain:

$$\begin{aligned} \alpha_j^{CG}(\tau_{CG})\, e^{i\omega_j t_{CG}} &\approx \left[\alpha_j^{BG}(\tau_{CG}) + \epsilon^2 \dot{\alpha}^{BG}(\tau_{CG}) \int_0^{\tau_{CG}} \delta_2 d\tau\right] \\ &\times \exp\left(i\omega_j t_{CG} + i\omega_j \int_0^{\tau_{CG}} \delta_2 d\tau\right) \end{aligned} \tag{4.11}$$

Therefore, we find that the complex coefficients in the two gauges are related by:

$$\alpha_j^{CG}(\tau) = \alpha_j^{BG}(\tau) \exp\left(i\omega_j \int_0^{\tau} \delta_2(\tau',0) d\tau'\right) + \mathcal{O}(\epsilon^2) \tag{4.12}$$

as is also explained in [56]. This result can also be expressed in the *amplitude–phase* representation, yielding:

$$A_j^{CG}(\tau) = A_j^{BG}(\tau) \tag{4.13}$$

$$B_j^{CG}(\tau) = B_j^{BG}(\tau) - \omega_j \int_0^\tau d\tau' \sum_i \left(A_{ii} + \omega_i^2 V_{ii}\right) A_i^2(\tau') \tag{4.14}$$

That the amplitudes and the phases are related as above can be directly checked by applying eq. (4.12) to the corresponding evolution equation in the two gauges, eq. (1.34), and recalling that the difference is entirely contained in the coefficients [56]:

$$T_j^{BG} = T_j^{CG} + \omega_j^2 \left(A_{jj} + \omega_j^2 V_{jj}\right), \tag{4.15}$$

$$R_{ij}^{BG} = R_{ij}^{CG} + \omega_j^2 \left(A_{ii} + \omega_i^2 V_{ii}\right). \tag{4.16}$$

In [54] it was shown that a large geometric back–reaction is related to the amplitude spectra and the coherence of the phases, where a phase–coherent cascade is defined by a spectrum of phases that (for large j) is linear in the mode number j:

$$B_j(\tau) = \gamma(\tau) j + \delta(\tau) + \ldots, \tag{4.17}$$

This is an *asymptotic*[3] statement and the ellipsis represent terms that are subleading in j. The reader should be aware here that the function $\delta(\tau)$ in the above equation is not the same function appearing in eq. (1.3). From eq. (4.13) we see that the evolution of the amplitudes is not affected by the choice of the gauge so what remains is to show that phase coherence is also unaffected and hence the physical conclusions will be independent of the choice of the gauge. Starting from eq. (4.17) for the central gauge we have:

$$B_j^{CG}(\tau) \approx \gamma^{CG}(\tau) j + \delta^{CG}(\tau), \tag{4.18}$$

and applying eq. (4.14) we can obtain the corresponding expression for the boundary gauge. This reads:

$$\begin{aligned} B_j^{BG}(\tau) &\quad -\omega_j \int_0^\tau d\tau' \sum_i \left(A_{ii} + \omega_i^2 V_{ii}\right) A_i^2(\tau') \approx \gamma^{CG}(\tau) j + \delta^{CG}(\tau) \Rightarrow \\ B_j^{BG}(\tau) &\quad \approx \left(\gamma^{CG}(\tau) + \int_0^\tau d\tau' \sum_i \left(A_{ii} + \omega_i^2 V_{ii}\right) A_i^2(\tau')\right) j + \delta^{CG}(\tau). \end{aligned} \tag{4.19}$$

[3] Recall also that asymptotically holds $\omega_j \approx j$, a fact that we use in eq.(4.19).

We see that the phase spectrum in the boundary gauge takes the form of eq. (4.17) :

$$B_j^{BG}(\tau) \approx \gamma^{BG}(\tau) j + \delta^{BG}(\tau), \tag{4.20}$$

with the functions $\gamma(\tau)$ and $\delta(\tau)$ in the two gauges being related as:

$$\gamma^{BG}(\tau) = \gamma^{CG}(\tau) + \int_0^\tau d\tau' \sum_i \left(A_{ii} + \omega_i^2 V_{ii}\right) A_i^2(\tau') \tag{4.21}$$

$$\delta^{BG}(\tau) = \delta^{CG}(\tau) \tag{4.22}$$

4.3.1 The oscillating singularity as an infinite gravitational redshift

Having clarified that physical conclusions can not be affected by the choice of the gauge, the next step is to reconcile the two different numerical results in the two gauges. In this section we will argue that the fact that $\dot{B}_j$ diverges in the one gauge and not in the other can be interpreted as an **infinite gravitational redshift** between the boundary and the center of the spacetime. Recall that the gravitational redshift between a source and an observer is given by the formula:

$$1 + z = \sqrt{\frac{g_{tt}(obs)}{g_{tt}(source)}}. \tag{4.23}$$

We can compute this quantity in one of the two gauges. Let us choose the nonsingular *boundary gauge* and compute the redshift between the boundary ($x = \pi/2$) and the center ($x = 0$) of the spacetime. Using the metric (1.3), the normalization $\delta(t, \pi/2) = 0$ and keeping terms only up to the order of $\mathcal{O}(\epsilon^2)$, the quantity under the square root reads:

$$\frac{g_{tt}(t,0)}{g_{tt}(t,\pi/2)} \sim 1 - \epsilon^2 \delta_2(t,\tau,0) + \mathcal{O}(\epsilon^4). \tag{4.24}$$

The expression for $\delta_2(t,\tau,0)$, eq. (1.13), yields[4]:

$$\begin{aligned}
\delta_2(t,\tau,0) &= \int_0^{\pi/2} \left(\dot{\phi}_1(t,x)^2 + \phi_1'(t,x)^2\right) \mu(x)\nu(x) dx \\
&= \int_0^{\pi/2} \sum_{ij} \left(\dot{c}_i^{(1)}(t)\dot{c}_j^{(1)}(t) e_i(x) e_j(x) + c_i^{(1)}(t) c_j^{(1)}(t) e_i(x) e_j(x)\right) \mu(x)\nu(x) dx \\
&= \sum_{ij} \left(\dot{c}_i^{(1)}\dot{c}_j^{(1)} V_{ij} + c_i^{(1)} c_j^{(1)} A_{ij}\right).
\end{aligned} \tag{4.25}$$

[4]For ease of notation we have omitted to write explicitly the slow time dependence in some cases, but it is implicitly assumed.

To go to the second line, we simply used the expansion in eigenmodes $\phi_1(t,x) = \sum_j c_j^{(1)}(t) e_j(x)$ and in the third line we defined the interaction coefficients:

$$A_{ij} \equiv \int_0^{\pi/2} e_i'(x) e_j'(x) \mu(x) \nu(x) dx \tag{4.26}$$

$$V_{ij} \equiv \int_0^{\pi/2} e_i(x) e_j(x) \mu(x) \nu(x) dx. \tag{4.27}$$

The expansion coefficients $c_j^{(1)}$ are related to the complex coefficients α_j as[5]:

$$c_j^{(1)} = \alpha_j e^{i\omega_j t} + \bar{\alpha}_j e^{-i\omega_j t} \tag{4.28}$$

$$\frac{dc_j^{(1)}}{dt} = i\omega_j \left(\alpha_j e^{i\omega_j t} - \bar{\alpha}_j e^{-i\omega_j t} \right). \tag{4.29}$$

Substituting eq. (4.28) in the above expression for $\delta_2(t,\tau,0)$ we will get several terms of the form $e^{i\Omega t}$, where $\Omega = \omega_i \pm \omega_j$. Keeping only terms with $\Omega = 0$, the so called *resonant* terms[6], we finally obtain the following expression:

$$\delta_2(t,\tau,0) \approx 2 \sum_i \left(A_{ii} + \omega_i^2 V_{ii} \right) A_i^2(\tau) \equiv \delta_2(\tau,0). \tag{4.30}$$

By differentiating eq. (4.14), we can see that this quantity, the *time–averaged* δ_2, which was first mentioned in eq. (4.8), is precisely the difference of the slow time derivatives of the phases in the two gauges. Therefore, by comparing the results in the boundary and the central time gauge we can draw conclusions about geometric quantities, and in particular the gravitational redshift. In the case of interest, where the derivatives of the phases diverge in one gauge but not in the other, one concludes that $\delta_2(\tau,0)$ diverges, and so does the redshift, eq. (4.24). This large back-reaction in turn implies the breakdown of linearized gravity. On the other hand if the derivatives are finite in both gauges there is no divergence, while the case is not clear if an oscillating singularity appears both in the boundary as well as in the central gauge. In that case $\delta_2(\tau,0)$ could be either finite or infinite.

4.4 Conclusions

In this chapter we presented an explicit derivation on the anticipated fact that physical results can not be affected by the different gauge choices. We demonstrated that gauge-invariant quantities are related to the amplitude spectrum and

[5] As we have stated below eq. (1.34) these are also related to A_j and B_j coefficients as: $A_j = |\alpha_j|$ and $B_j = Arg(\alpha_j)$.

[6] These are are the terms that are proportional to $e^{\pm i(\omega_i - \omega_j)} \delta_{ij}$. This procedure is equivalent to time–averaging over the *fast time* t.

the coherence of the phases in the TTF solution, and both properties are unaffected by the gauge choice. This result holds even when the difference between the two gauges diverges. Furthermore we established that the oscillating singularity observed in [48] is indeed a physical singularity, by showing that it is related to an infinite redshift between the boundary and the center of the spacetime.

This means that the breakdown of the TTF observed in [48] is due to large gravitational effects which lead to the breakdown of the weak gravity approximation. Such a conclusion cannot be deduced by the observed singularity in the central time gauge alone. In that case is not clear whether the breakdown of the perturbation theory is caused by strong gravity or by the breakdown of other approximations. Therefore, with our analysis we establish that the singular solution is a genuine singular solution of the gravitational problem. Due to the scaling symmetry of the TTF system the solution will survive in the $\epsilon \to 0$ limit, and thus provide a way to address the phase space of initial conditions in this limit.

An interesting thing to point out here is that for this conclusion we need to compare the derivatives of the phases in the two gauges. Therefore, the fact that in higher dimensions a discrepancy between the two gauges has not been observed [53] is rather intriguing. However since in both gauges an oscillating singularity was observed, and actually in the central time gauge this divergence was more prominent than in the boundary time gauge, it might still signal a diverging redshift, since these results are compatible with a diverging $\delta_2(\tau, 0)$, as we explained in Section 4.3.1.

5

Fast and Slow Coherent Cascades in AdS

Physics is about questioning, studying, probing nature. You probe, and, if you're lucky, you get strange clues.

— Lene Hau

5.1 Introduction

One could hope for a resolution of the AdS instability problem in terms of a theorem. A preliminary attempt for such a theorem was presented in [54]. We will discuss the details of this work in the next section, but for now it suffices to say that it conjectures that the cascade of energy towards higher modes takes place in a *coherent way*. Namely, the higher frequency modes are turned on with their phases coherently aligned. The theory of [54] also makes predictions for the resulting *power-law* of this cascade.

In this chapter we study the evolution of the phases and the amplitudes of the Fourier modes, in the Two Time Framework (TTF) approximation, for a variety of initial conditions, and we find that the phases are not excited in a random way, as in the standard theory of weak turbulence, but in a quite coherent way, supporting the coherent phase ansatz of [54]. However, we observe small deviations from a strictly coherent phase spectrum, suggesting that the conjecture [54] should be refined.

We also study the evolution of the two-mode, equal energy initial data, that are considered minimal for the onset of the instability, and we provide a combination of numerical and analytic evidence to support the conjecture that they belong to a class of initial conditions that collapse in **infinite slow time**[1]. We show analytically that solutions dominated by the two lowest modes obey a simple

[1] At the vanishing amplitude limit, $\epsilon \to 0$.

speed limit on the rate of energy transfer to higher frequencies. The speed limit depends on the frequency; we show that in $3+1$ dimensions it takes infinite time to transfer energy to arbitrarily high frequency, while in higher dimensions it takes finite time.

A nice way to track energy transfer in this system is the so-called *analyticity strip method* [38, 48, 57], which diagnoses at what time the amplitude spectrum ceases to decrease exponentially at large mode numbers. We find numerically that, for the two-mode initial data, the width of the analyticity strip goes to zero **exponentially** in time. There has been a long-lasting dispute about the ultimate fate of this initial data [12, 29, 58] and we hope we settle the issue with this work.[2]

We also study initial data that has a Gaussian profile. We find that this data collapses in finite slow time. In addition, we find that at the time of the collapse the amplitudes A_n of the normal modes approach a power-law spectrum of the form $A_n \sim n^{-\gamma}$ with $\gamma \approx 3/2$, which modifies slightly the result $\gamma = 8/5$ reported in [11].[3] This amplitude spectrum corresponds to an energy spectrum $E_n \sim 1/\omega_n$.

5.2 Phase coherent turbulent cascades

In this section we present a minimal introduction to the work of [54] and we will mainly focus on what is necessary for the comprehension of the rest of this chapter.

We have seen so far, that the Fourier analysis of the black hole formation in AdS, points towards a cascade of energy from lower frequency to higher frequency modes. This process is known as *turbulence* and when the cascade is driven by nonlinearities it is dubbed as *weak turbulence*. Despite the striking resemblance with the case of *AdS* it needs to be clarified that in the standard theory of weak turbulence, as developed by Kolmogorov and Zakharov for example, the effect of the phases was not taken into account. The phases of the eigenmodes where drawn from random distributions and therefore their effect averages to zero. An imporant consequence of this choice is that, for a system with quartic interactions[4] there is no energy transfer in the first nonlinear time scale ($t \sim \epsilon^{-2}$) and the first non-trivial dynamics appears at $t \sim \epsilon^{-4}$.

[2] The fact that this data might collapse in infinite slow time was suggested earlier to us by Andrzej Rostworowski, as we understand, based on simulations of full General Relativity in AdS. It is our understanding that A. Rostworowski and collaborators have also obtained interesting results about this $3+1$-dimensional problem, and we look forward to comparing them to our results.

[3] The fact that the exponent might be smaller than $8/5$ was also conveyed to us in a discussion with A. Rostworowski and collaborators.

[4] As it is the case for *AdS*.

The basic idea in [54] is to establish a class of solutions to the TTF, that are stationary and have a power-law spectrum[5]

5.2.1 Geometric deviation and phases

As it was first mentioned in chapter 2, when the energy is very small (approaching zero), the only way to have a large back reaction is focus this energy into a very small small region. This region can then be described locally by the Minkowski metric for which the perturbative expansion for a small back reaction is well known:

$$ds^2 = -\left(1 + \frac{2M}{r^{d-2}} + 4V\right) dt^2 + \left(1 + \frac{2M}{r^{d-2}}\right) dr^2 + r^2 d\Omega_{d-1}^2, \tag{5.1}$$

where M and V are the enclosed mass and the gravitational potential respectively. The enclosed mass is given by

$$M(r) \sim \int_0^r dr' \frac{\dot{\phi}^2 + \phi'^2}{2} \tag{5.2}$$

and if we assume a *power law* spectrum in the form

$$A_n = A_0 (n+1)^{-\alpha}, \tag{5.3}$$

we obtain for M

$$\begin{aligned} M(r) &\sim \int_0^r dr' r'^{d-1} \left(\sum_{n=0}^{\infty} w_n A_n e_n(r') \cos(w_n t + B_n) \right)^2 \\ &\sim \int_0^r dr' r'^{d-1} \left(\sum_{n=r^{-1}}^{\infty} w_n^2 A_n^2 e_n^2(r') + \left(\sum_{n=0}^{r^{-1}} w_n A_n e_n(0) \cos(w_n t + B_n) \right)^2 \right) \end{aligned} \tag{5.4}$$

The importance of the phases in the back-reaction is apparent already in the second term of the above expression. Modes of higher frequency ($n > r^{-1}$) oscillate

[5] In contrast to the quasi periodic solutions [12] that are stationary but with an exponential spectrum.

rapidly within the integration range and thus the cross terms vanish. On the other hand, modes of lower frequency ($n < r^{-1}$) are basically constant within the integration range and if their phases are coherent, then the cross terms have a significant contribution to M.

The phases are considered *coherent* if there is a time during one *AdS* period where all of the modes are in phase. The phase θ_n of a particular mode is related to the *slow* phase B_n as

$$\theta_n(\tau, t) = B_n(\tau) + \omega_n t = B_n(\tau) + (2n + d)t \tag{5.5}$$

Therefore, the phase coherence of two modes n and m is defined in the usual way as

$$\theta_n - \theta_m = 2\pi N_{nm}, \tag{5.6}$$

where N_{nm} are integers which in principle can depend on the modes involved. Coherence requires that we can solve this equation for the *fast time* t over one period $0 < t < 2$, at the same t for all modes. Plugging in the formula for the phases θ_n, we obtain

$$\begin{aligned} B_n(\tau) - B_m(\tau) &= 2\pi N_{nm} + 2(n - m)t \\ &= (n - m)\theta(\tau). \end{aligned} \tag{5.7}$$

To go to the second line, we dropped the the first term, since the phases are only defined mod 2π and for ease of notation we set $\theta(\tau) \equiv 2t$. A solution to the above equation for every n and m is

$$B_n(\tau) = n\gamma(\tau) + \delta(\tau), \tag{5.8}$$

where γ and δ are functions of the slow time τ, but independent of the mode number n. In the problem at hand, one is interested in describing the asymptotic spectrum, namely large values of n. Therefore one should, in principle allow for corrections to the above solution that are subleading in n.

The effect of the phases in the backreaction, for different values of the resulting power-law α, is summarized in Fig. (5.1).

However, this phase coherent ansatz needs to be maintained during the time evolution and according to the results of [54] this happens only when $\alpha = \frac{d}{2}$. This predicted power-law leads to a spectrum of energies of the form $E_n \sim n^{2-d}$. The resulting energy spectrum and the *phase coherent cascade* for a variety of initial conditions will be a part of our studies in the following sections.

d=3	regular	naked curvature singularity	naked redshift singularity	black hole
incoherent phases	$\alpha > \frac{5}{2}$	$\frac{5}{2} \geq \alpha > \frac{3}{2}$	$\alpha \leq \frac{3}{2}$	never
coherent phases	$\alpha > 3$	$3 \geq \alpha > 2$	$\alpha = 2$	$\alpha < 2$
$d > 3$				
incoherent phases	$\alpha > \frac{d+2}{2}$	$\frac{d+2}{2} \geq \alpha > \frac{d}{2}$	$\alpha = \frac{d}{2}$	$\alpha < \frac{d}{2}$
coherent phases	$\alpha > \frac{d+3}{2}$	$\frac{d+3}{2} \geq \alpha > \frac{d+1}{2}$	$\alpha = \frac{d+1}{2}$	$\alpha < \frac{d+1}{2}$

Figure 5.1: *The effect of the phases for different values of the* power-law α *and for different spatial dimensions dimensions d [54].*

5.3 The two–mode data

We will start by looking at two mode–initial data, which is parametrized as follows:

$$A_n = \frac{\epsilon}{3}\left(\delta_n^0 + \kappa\delta_n^1\right). \tag{5.9}$$

This is the case where we initially excite the lowest two modes with no relative phase. Here, ϵ denotes the amplitude of the perturbation, while κ parametrizes the amount of energy in the second lowest mode. For concreteness, we have considered data with $\epsilon = 1$ and different values of κ. However, as discussed before, the TTF equations are invariant under the scaling symmetry $\alpha_n(\tau) \to \epsilon\alpha_n(\tau/\epsilon^2)$. Therefore, the solution for different values of ϵ would be completely equivalent to the case $\epsilon = 1$ but evolved to a different time $\tau_{\max}$.

5.3.1 Numerical results

Equal energy:

Setting $\kappa = 3/5$ corresponds to a situation where the energy is equally distributed between the two modes. This case is very interesting since it has been argued to be in the borderline of one of the stability islands [12, 29, 58]. Here, instead of looking at the value of Π at the origin,[6] we would like to concentrate on the spectrum of the amplitudes and the phases, hoping to shed some light from a different perspective and clarify the fate of this initial condition.

In order to detect the formation of singularities from the spectrum we use the so-called analyticity strip method, introduced in [57], and employed for the first time in the context of the AdS instability problem in [38]. The idea here is to

[6] The value of $\Pi(t, 0)$ at the origin corresponds to the Ricci scalar and therefore is a good indicator of singularity.

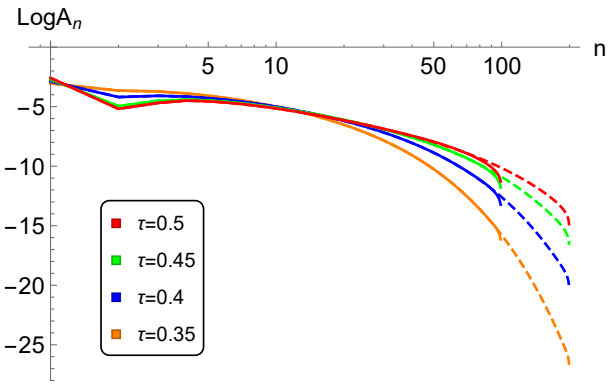


Figure 5.2: *Amplitude spectrum A_n, for different values of τ. Solid lines correspond to a 100 mode truncation, while dashed lines correspond to a 200 mode truncation.*

consider the analytic extension of $\phi(t,z)$ into the complex plane of the radial variable $r \to z \in \mathbb{C}$. The function $\phi(t,z)$ will typically have complex singularities moving in time; if one of these singularities hits the real axis, $\phi(t,r)$ becomes singular. The pair of singularities closest to the real axis are denoted as $z = x \pm i\rho$, so that ρ determines the width of the analyticity strip around the real axis. Thus, if ρ vanishes at some point during the evolution then $\phi(t,r)$ will be singular. Now, ρ is encoded in the exponential decay of the Fourier coefficients $A_n \sim e^{-\rho n}$ (at large n), so it can be obtained from the asymptotics of a given numerical solution.

In order to evolve the system of equations (1.34)-(1.34) we need to fix the gauge and find the coefficients S_{ijkl} up to a maximum number $i = j = k = l = n_{\max}$. We choose to work in the boundary gauge, where $\delta(t, \pi/2) = 0$, since in this gauge the evolution of the phases is regular [2,52,53] (and therefore, the numerical integration is easier). In order to estimate the effect of the cutoff $n_{\max}$ we evolve the system for two different cases, first for $n_{\max} = 99$ and then for $n_{\max} = 199$. In Fig. (5.2) we show the amplitude spectrum in the two cases, for different values of τ. In general they agree over a wide range of n, but differ mildly in the range $n \in (79, 99)$. Next, we fit the amplitudes using the following ansatz

$$A_n(\tau) \sim \alpha(\tau) n^{-\gamma(\tau)} e^{-\rho(\tau) n} \,. \tag{5.10}$$

We exclude the modes $n \in (70, 99)$ for $n_{\max} = 99$ and $n \in (140, 199)$ for $n_{\max} = 199$ to avoid any cutoff effect. Also, since the analyticity strip method applies only asymptotically, we also exclude the first modes, $n \in (0, 29)$ for $n_{\max} = 99$ and $n \in (0, 59)$ for $n_{\max} = 199$. Comparing the two fittings should then provide a good test of our numerics, since the two ranges of n barely intersect with each other.

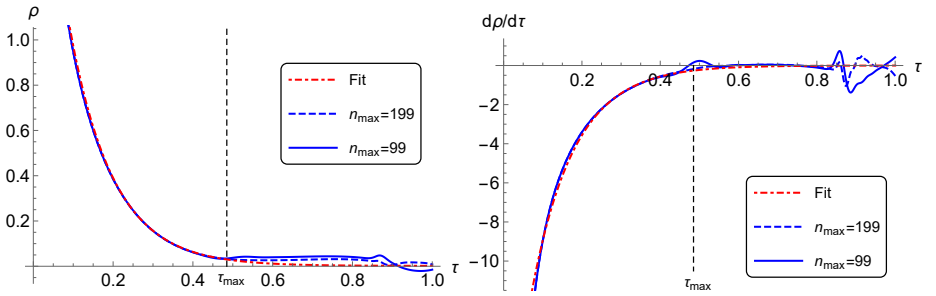


Figure 5.3: **Left:** *Evolution of $\rho(\tau)$ for $n_{max} = 99$ and $n_{max} = 199$. We define $\tau_{max} \sim 0.485$ as the maximum time such that $|\rho_{199} - \rho_{99}| \leq 1/n_{max}$; beyond this point we do not trust our numerical evolution. We have also included the best fit up to this time, using the ansatz $\rho(\tau) = \rho_0 e^{-\alpha\tau} + \rho_\infty$. This yields $\rho_\infty \sim 5 \times 10^{-5} \ll 1/n_{max}$ which strongly supports the idea that $\rho \to 0$ as $\tau \to \infty$.* **Right:** *Evolution of $d\rho/d\tau$ for $n_{max} = 99$ and $n_{max} = 199$. The behavior of the derivative for $\tau < \tau_{max}$ shows that the exponential function is indeed a good fit for ρ.*

In Fig. (5.3) we plot the evolution of ρ and its derivative as a function of τ for the two cases considered, $n_{\rm max} = 99$ and $n_{\rm max} = 199$. In general, they are in excellent agreement for the first part of the evolution, but eventually they start to differ. Since ρ is the most robust parameter in the fitting (since it appears in the exponential), we define $\tau_{\rm max} \sim 0.485$ as the maximum time such that $|\rho_{199} - \rho_{99}| \leq 1/n_{\rm max}$; beyond this point the two ρ start to differ substantially and we cannot trust our numerical evolution. We see that up to $\tau_{\rm max} \sim 0.485$, ρ decays almost exponentially, so we fit it as

$$\rho(\tau) \sim \rho_0 e^{-\alpha\tau} + \rho_\infty \,. \tag{5.11}$$

Interestingly, we find that $\rho_\infty \sim 5 \times 10^{-5} \ll 1/n_{\rm max}$ strongly supporting the idea that $\rho \to 0$ at infinite τ! The value of γ varies very little before $\tau_{\rm max}$ —see Fig. (5.4)— but fluctuates quite a lot after this time. This makes it impossible to extrapolate its value to $\tau \to \infty$. We also plot the fitting function (5.10) at $\tau = \tau_{\rm max}$, finding an excellent agreement even outside of the range of n that we considered for the fits. This suggests that the formula (5.10) is actually quite robust, at least for the times at which we can trust our numerical results, $\tau \leq \tau_{\rm max}$.

It is also interesting to put to test the validity of the phase coherent conjecture [54], which states that the asymptotic spectrum of phases align coherently as in (4.17). In order to see if this is actually true, we plot in Fig. (5.5) B_n as a function of n for various values of τ.[7] Interestingly, we see an almost perfect line for $\tau \leq \tau_{\rm max}$, with deviations from linearity being indistinguishable to the naked eye. To quantify the

[7]Since $\mathrm{Arg}[\alpha_n] = B_n \in (-\pi, \pi)$ we first *unwrap* the phases before plotting the actual values. This amounts to shift $B_n \to B_n + 2\pi k$, $k \in \mathbb{Z}$ every time that $\Delta B_n = B_{n+1} - B_n$ changes sign.

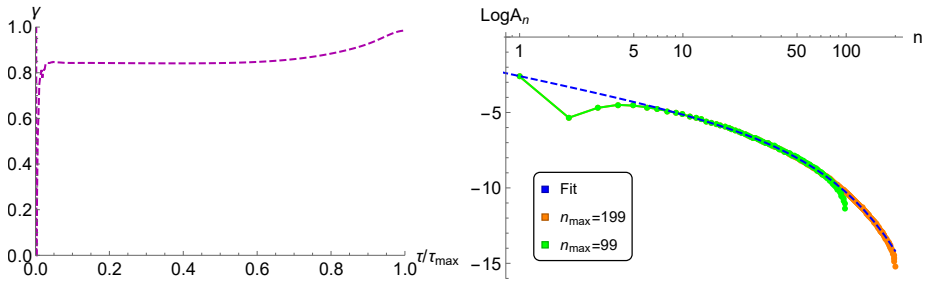


Figure 5.4: **Left:** *Evolution of $\gamma(\tau)$ up to $\tau_{max} \sim 0.485$. Beyond this point $\gamma(\tau)$ behaves quite erratically, also indicating the breakdown of our numerical solution. It seems difficult to extrapolate γ for $\tau \to \infty$.* **Right:** *Fitting of the amplitude spectrum at $\tau = \tau_{max} \sim 0.485$ according to the formula (5.10).*

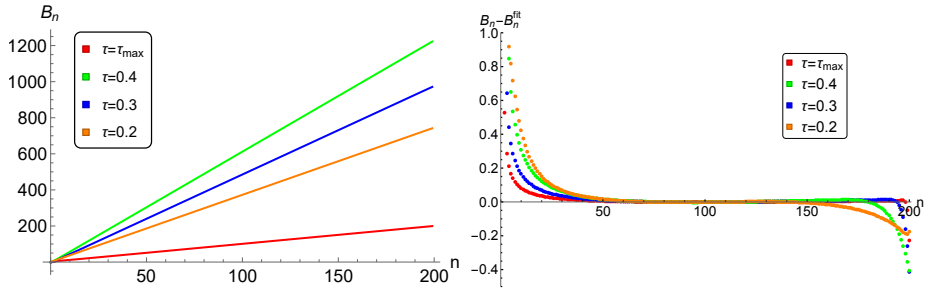


Figure 5.5: **Left:** *Unwrapped phases B_n as a function of n for various values of τ. In all cases we see that the phases are perfectly aligned to form a straight line as in (4.17).* **Right:** *The difference between the fitted and the actual values as a function of n.*

linearity of the spectrum we subtract the fitted values from the actual data. The results are presented in Fig. (5.5) as well. We observe a very good agreement with the linear, especially in the range of values used for the fit, $60 \leq n < 140$.

Non-equal energy:

For other values of κ, (5.9) implies that the initial energy of the two modes is different. We evolved the system for a wide range of κ and repeated the analysis that we presented for the case $\kappa = 3/5$. In extreme cases where most of the energy is deposited in one of the modes, the system can be though of a perturbation of a single mode solution, which is actually the center of one of the stability islands for AdS perturbations [3,13,49]. In such cases we do not expect black hole formation. On the other hand, values of κ close to $\kappa = 3/5$ should behave similarly to the equal energy data and are expected to collapse.

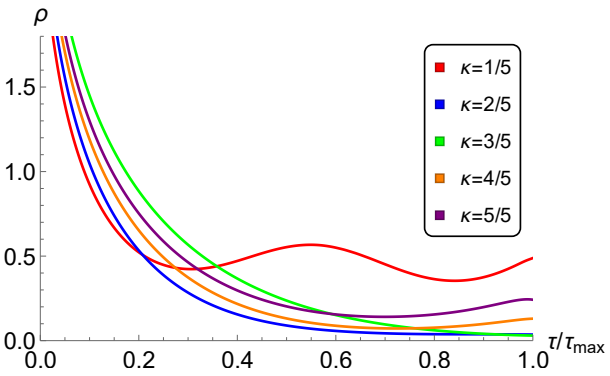


Figure 5.6: *The evolution of the analyticity strip width $\rho(\tau)$, for different ratios of the energy in the first two modes.*

In Fig. (5.6) we plot the analyticity strip width $\rho(\tau)$ as a function of τ for various values of κ, showing different kind of behaviors. For $\kappa = \{1/5, 4/5, 5/5\}$ $\rho(\tau)$ develops oscillations and never reach zero. These initial conditions are not exactly periodic as the single–mode solutions but exibith revivals at time scales of order $1/\epsilon^2$ [35, 59]. For $\kappa = 2/5$ we do not see oscillations before $\tau_{\rm max}$ but by doing the fit we find that ρ_∞ is marginally above $1/n_{\rm max}$. We conjecture that this case is close to the borderline of a subspace of solutions that collapse at infinite τ, which for the two-mode initial data (5.9) is given by an open set with $\kappa \in (\kappa_{\rm min}, \kappa_{\rm max})$. We further studied the phase spectrum of all these initial data, and found always good agreement with the coherent phase ansatz.

5.3.2 A speed limit for energy transfer

It is interesting and surprising that some initial data appears to cascade to arbitrarily high frequencies in finite time, while other initial conditions take an infinite time. To really convince ourselves that the energy cascade takes infinite slow time for some initial conditions, we need an analytic argument. In this section, we give a simple argument showing that for solutions where two low modes dominate the spectrum, the remaining modes obey a simple speed limit in how fast energy can propagate to high frequencies. Our result rests on a strong assumptions that the solution is dominated by two low modes, and it would be very interesting to relax this assumption.

For this purpose it is convenient to write the equations of motion for the truncated resonant system in terms of complex amplitudes $\alpha_n \equiv A_n \exp(iB_n)$. The equations

of motion become

$$2i\omega_n \frac{d\alpha_n}{d\tau} = \sum_{i+j=k+n} C_{ijkn}\alpha_i\alpha_j\bar{\alpha}_k \tag{5.12}$$

Now we will assume that the solution is dominated by two low modes; for definiteness, take modes $n = 0, 1$, but the basic structure of our results will remain unchanged. In this case, we can keep only terms in the sum where two of the indices take the values $0, 1$. Since we have to obey the resonance condition $i + j = k + n$, for $n > 2$ this means that $k = 0, 1$, leaving the equation

$$2i\omega_n \frac{d\alpha_n}{d\tau} = 2C_{1(n-1)0n}\,\alpha_1\alpha_{n-1}\bar{\alpha}_0 + 2C_{0n0n}\alpha_0\alpha_n\bar{\alpha}_0 + 2C_{1n1n}\alpha_1\alpha_n\bar{\alpha}_1 + 2C_{0(n+1)1n}\alpha_0\alpha_{n+1}\bar{\alpha}_1 \tag{5.13}$$

Collecting terms and using the symmetries of the C coefficients, we get

$$i\omega_n \frac{d\alpha_n}{d\tau} - (C_{0n0n}\alpha_0\bar{\alpha}_0 + C_{1n1n}\alpha_1\bar{\alpha}_1)\,\alpha_n = C_{1(n-1)0n}\,\alpha_1\bar{\alpha}_0\alpha_{n-1} + C_{1n0(n+1)}\alpha_0\bar{\alpha}_1\alpha_{n+1} \tag{5.14}$$

Despite the complicated coefficients, this equation has two simplifying features:

- It is linear in the amplitudes α_n.
- It is *local*: a given mode is only influenced by its nearest neighbors (in addition to modes 0 and 1, which we think of as a background.)

We are interested in large mode numbers n. Since the C coefficients are smooth functions, and n is large, we take $C_{1n0(n+1)} \approx C_{1(n-1)0n} \equiv C_n$. We treat the dominant modes $\alpha_{0,1}$ as constants (it would be interesting to allow for time dependence).

As discussed above, the overall phase and the relative phase between any *two* modes is unphysical, so we are free to choose convenient phases for $\alpha_{0,1}$. We use this freedom to take the product $\alpha_0\bar{\alpha}_1$ to be pure imaginary for convenience, leading to the equation

$$i\omega_n \frac{d\alpha_n}{d\tau} - \left(C_{0n0n}A_0^2 + C_{1n1n}A_1^2\right)\alpha_n = iC_{1(n-1)0n}\;A_1A_0(\alpha_{n-1} - \alpha_{n+1}) \tag{5.15}$$

This equation can be written in the simpler notation

$$\frac{d\alpha_n}{d\tau} + id_n\alpha_n + \frac{1}{2}c_n(\alpha_{n+1} - \alpha_{n-1}) = 0 \tag{5.16}$$

with the definitions

$$d_n \equiv \frac{C_{0n0n}A_0^2 + C_{1n1n}A_1^2}{\omega_n} \qquad c_n \equiv \frac{2C_{1(n-1)0n}\;A_1A_0}{\omega_n} \tag{5.17}$$

Since a given mode couples to nearest neighbors, we would like to identify the speed at which signals can propagate in frequency space. To estimate this, take a continuum limit of the above equation, treating n as a continuous variable, to get

$$\frac{\partial \alpha}{\partial \tau} + i d_n \alpha + c_n \frac{\partial \alpha}{\partial n} = 0 \tag{5.18}$$

We have not fully analyzed this equation, but the speed of propagation can be read off by comparing the coefficient of the time derivative to the space derivative, yielding a speed of propagation that depends on the mode number

$$\text{speed} = \frac{dn}{d\tau} = c_n \tag{5.19}$$

It is now interesting to ask whether energy can propagate to arbitrarily large mode numbers in finite time. The time to reach infinitely large mode numbers is

$$\Delta\tau = \int^{\infty} \frac{dn}{c_n} \tag{5.20}$$

Whether this is infinite depends on the function c_n, which depends on the dimension. In all dimensions, $\omega_n \sim n$. In 3+1 dimensions, the interaction coefficient $C_{1(n-1)0n} \sim n^2$ [52, 55]. Therefore, in 3+1 dimensions, $c_n \sim n$, and the integral is logarithmically divergent, and the cascade cannot reach infinity in finite time. The maximum excited mode number as a function of time can increase at most as $n_{\max} \sim \exp(a\tau)$. This agrees well with the observation in our numerical evolution that the spectrum has an exponential form $\exp(-\rho n)$ with $\rho \sim \exp(-a\tau)$.

In higher than 3+1 bulk dimensions, we expect the interaction coefficients to scale with larger powers of n, leading to a scaling $c_n \sim n^p$ with $p > 1$. This renders the integral convergent, meaning that energy can reach infinite mode number in finite time. This is in good agreement with the results obtained in [48] in 4+1 dimensions, where the spectrum was observed to approach a power law in finite time.

As a reminder, we were able to reach this strong conclusion by making a strong assumption that modes 0 and 1 dominate throughout the evolution. We expect a similar result whenever we assume that the evolution is dominated by a finite number of low-frequency modes: the equation for the high modes will still become linear and quasi-local, with the degree of nonlocality in frequency space determined by the frequency of the low modes that are excited.

On the other hand, the assumption that the solution is dominated by a few low modes can clearly break down as energy is transferred to higher modes. For this reason, it is not at all clear that the 'speed limit' found under this assumption is a robust result, although it is intriguing that it appears to agree with numerical results. It would be very interesting to derive a more generally valid speed limit on the rate of energy transfer.

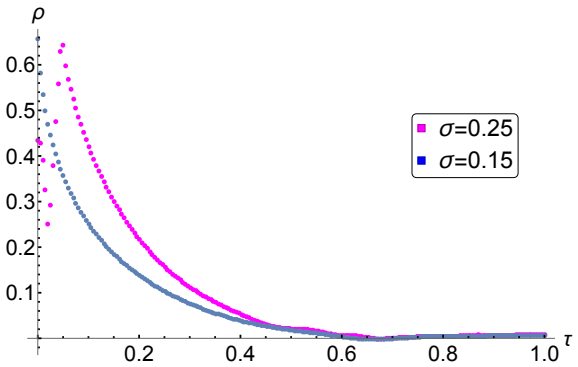


Figure 5.7: *The evolution of the analyticity strip width $\rho(\tau)$ for Gaussian data with $\sigma = 0.25$ and $\sigma = 0.15$. In both cases ρ becomes zero in finite slow time $\tau_\star \sim 0.625$.*

5.4 Gaussian Data

In this section we study scalar fields with Gaussian profiles of the form $\phi(0, x) = 2\exp\left(-\frac{\tan^2 x}{\sigma^2}\right)$ in TTF where previous simulations, in the full GR system, reported that the collapse depends on the width of profile [10,11]. We concentrate our study on profiles with widths $\sigma < 0.3$. We present results for two cases, with $\sigma = 0.15$ and $\sigma = 0.25$. We follow the same analysis as the one of Sec.(5.3), so we will not repeat the details here. One thing we would like to mention is that here the case is more clear since this initial data approach a power-law spectrum in finite time and therefore we only use the truncation of $n_{max} = 200$. One other thing is that now we have $\tau_{max} = \tau_\star$, namely, we trust the evolution up to the time of the collapse $\tau_\star$.

In Fig (5.7) we plot the evolution of the analyticity strip width, for the two above-mentioned data to make precise the contrast with the 2–mode data. Here $\rho(\tau)$ goes to zero in **finite slow time** $\tau_\star \sim 0.625$. Our results agree with the observation of [10], that collapse happens for narrow profiles ($\sigma < 0.3$), but disagree with [11] on the resulting power-law of the energy cascade. We find that by the time of the collapse, or equivalently when the analyticity strip width $\rho(\tau)$ goes to zero, the power $\gamma(\tau)$ approaches a value very close to $3/2$ and not $8/5$.

In Fig. (5.8) and Fig. (5.9) we present the evolution of the amplitude spectra for Gaussian initial data with $\sigma = 0.15$ and $\sigma = 0.25$ respectively, towards a power law and we contrast the two above-mentioned values. We see that indeed the value $\gamma(\tau_\star) = 3/2$ is a better fit. We would like to mention at this point that the numerics are very subtle in AdS_4 and although our results strongly suggest a

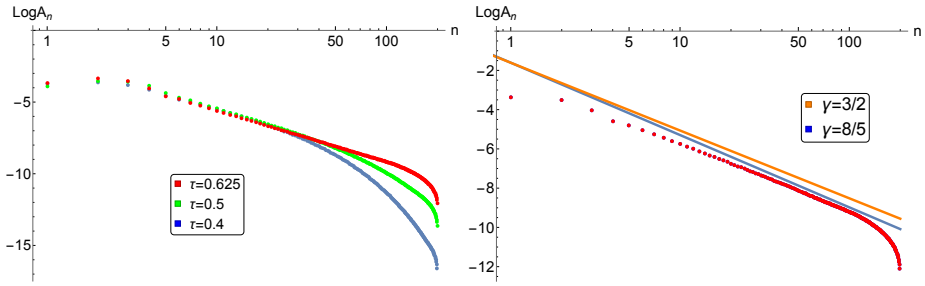


Figure 5.8: **Left**: *The evolution of the amplitude spectra for Gaussian data with* $\sigma = 0.15$ *towards a power law* $\gamma(0.625) \sim 3/2$ *in the TTF system.* **Right**: *Comparison of the two power-laws with the actual data at the time of* collapse, $\tau_\star = 0.625$.

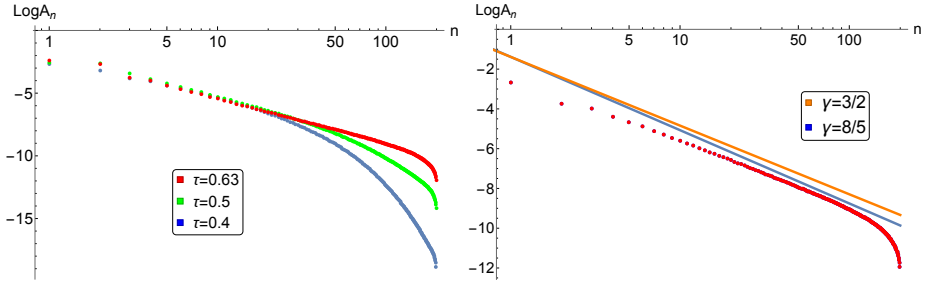


Figure 5.9: **Left**: *The evolution of the amplitude spectra for Gaussian data with* $\sigma = 0.25$ *towards a power law* $\gamma(0.63) \sim 3/2$ *in the TTF system. The evolution is almost identical to the case of* $\sigma = 0.15$. **Right**: *Comparison of the two power-laws with the actual data at the time of* collapse, $\tau_\star \sim 0.63$.

power law of 3/2, another value for γ very close to this one is still a possibility.

To illustrate the subtleties of the numerical methods we now study into more detail the fitting methods for the case of the $\sigma = 0.15$ Gaussian data close to the *collapse* point $\tau_\star \sim 0.625$. We fit the data in the range $30 \leq n \leq 110$ in order to to take into account only the asymptotic form of the spectrum and avoid cut-off effects. In Table 5.1 we present the fitting functions for three different values of $\tau = \{0.62, 0.625, 0.63\}$ and we fit both a power law spectrum and a power law spectrum with an exponential decay. We observe that the analyticity strip width turns from positive, at $\tau = 0.625$ to negative, at $\tau = 0.63$ which means that it hits zero in finite slow time. Also, at $\tau = 0.625$ it has the smallest value, and we believe that this is the most relevant time. However we don't have exact data for which $\rho = 0$. Two things that we would like to notice here is that although the value of the exponent is very small, it plays a significant role in estimating the exact value of the power γ and that the exact range of data that we use for the fitting affect

	$\tau = 0.62$		$\tau = 0.625$		$\tau = 0.63$	
	Parameter	*Error*	*Parameter*	*Error*	*Parameter*	*Error*
$\mathbf{A_n} \sim \alpha \mathbf{n}^{-\gamma} \mathbf{e}^{-\rho \mathbf{n}}$						
$\mathrm{Log}(\alpha)$	-2.17944	0.02399	-2.18595	0.0301	-2.20568	0.03769
γ	1.51033	0.0077	1.51579	0.00966	1.517	0.012
ρ	0.000754	0.00012	0.000205	0.000151	-0.000272	0.000189
$\mathbf{A_n} \sim \alpha \mathbf{n}^{-\gamma}$						
$\mathrm{Log}(\alpha)$	-2.03288	0.00649	-2.14604	0.00672	-2.25847	0.00843
γ	1.55796	0.00154	1.52876	0.0016	1.49985	0.002

Table 5.1: *The fitting values and the corresponding errors for three different times very close to the collapse time $\tau_\star \sim 0.625$, in the case of a power law with an exponential (***up***) and a power law alone (***down***).*

the result as well[8]. The fits, in both cases suggest a power law very close to 3/2, however as we mentioned earlier, values close to this one are also possible.

In Fig. (5.10) we present again the phase-spectrum as a function of the mode number n for different slow times. As we did earlier, we *unwrap* the phases, and we fit the data in the range[9] $60 \leq n < 140$ to a linear function and we quantify the deviation from linearity by subtracting the fitting value from the actual data. We see once more that the linear fit is a very good approximation, however interesting patterns appear, especially for late times, that might suggest towards a slight improvement to the perfectly coherent spectrum. We hope we will come back to this issue in a future work.

5.5 Conclusion

In this work we presented strong evidence that the turbulent cascade of energy towards modes of higher frequency in the problem of the instability of global AdS, happens in a *phase–coherent* way, as it was initially conjectured in [54]. Our results however leave room for possible small improvements on this idea and we believe it is something worth looking at in the future. On top of that, the perturbations that we studied here are also initially phase coherent and one might think that this could play a role in the later-time coherence of the phases. A natural generalization would be to study the development of phase coherence in perturbations that are

[8] We chose the above-mentioned range, $n \in [30, 110]$, since it represents the most linear part of the plot and neglects the cut-off effect.

[9] As usual, we do so in order to address the *large n limit* and to avoid cut-off effects.

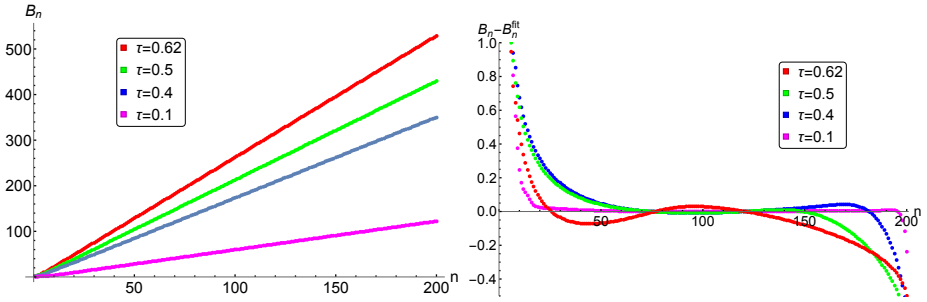


Figure 5.10: **Left:** *The phases of the modes as a function of the mode number n when* unwraped *for different slow times τ.* **Right:** *The difference between the fitted and actual values as a function of n.*

not initially phase coherent[10].

We have also studied the evolution of the amplitude spectra for the controversial two mode data and our work, combining numerical and analytical results, suggests that the collapse will happen in infinite slow time $\tau_\star \to \infty$. For the Gaussian initial data with $\sigma < 0.3$ we observed that the spectrum of the amplitudes approaches a power-law at finite slow time $\tau_\star \sim 0.625$, and the results suggest the power at the time of the collapse being very close to $\gamma(\tau_\star) \sim 3/2$. It would be very interesting to determine the power-law in the case of the two mode data and see if it agrees with the above-mentioned value, as well as to obtain a definite answer for the precise value in the case of the Gaussian data, perhaps through analytic techniques.

[10]We will come back to this issue in future work, but some preliminary results that we have suggest that even in data for which the phases are initially randomly distributed, the higher modes are excited coherently

6 Outlook

The end of a melody is not its goal: but nonetheless, had the melody not reached its end it would not have reached its goal either.

— F. Nietzsche

In this final chapter, the main results of this work are summarized and considerations for possible future directions are presented.

A position space analysis

In chapter 2 we studied the problem of the stability of *AdS* under small perturbations directly in position space. Previous work on the subject had focused on the Fourier analysis of the problem, namely the exchange of energy between the different normal modes in which the scalar field can be decomposed. Here instead, we took a different route by studying directly an infalling spherical shell of a massless scalar field. We resorted to perturbation theory, where we used the so called *near Minkowski expansion* according to which all of the interesting dynamics takes place near the center of the spacetime, which is approximately of Minkowski type. The role of *AdS* is simply to confine the shell due to a reflecting boundary at spatial infinity. Two important results of our perturbative analysis, which we truncate at the first nonlinear order, are that the resulting approximate equations possess a scaling symmetry similar to the one obtained from Fourier analysis and that the first order correction to the scalar field profile has a *reflection anti-symmetry*. The latter means that if the initial profile is asymmetric and the first order correction leads towards a focusing (defocusing) of the energy, then if we would start with a profile that is a reversed version of the initial profile, then its first order correction would have the opposite effect. Namely it would defocus (respectively focus) the energy. We conclude this chapter by clarifying the interpretation of the black hole formation in the bulk in terms of thermalization of the boundary field theory. We argue that black hole formation in *AdS* doesn't always correspond to an efficient thermalization of the field theory. Whether this is the case or not depends on the way in which one takes the limits of classical and weak gravity.

Extended validity of perturbation theory

Since all analytical results we have obtained so far rely on first order perturbation theory, it is natural to worry up to what time scales are our approximate schemes reliable. Organising our perturbation theory in terms of the amplitude of the perturbation ϵ, common lore suggests that naive *first order* perturbation theory will be valid for time scales of the order $t < \epsilon^{-2}$, whereas from numerical results of the full nonlinear system we know that black hole is expected to form at $t \sim \epsilon^{-2}$. It seems therefore that naive perturbation theory is not capable of capture the actual dynamics at the relevant time scales and one might need to consider more elaborate perturbative expansions, like TTF. In chapter 3 we prove however that this is not always the case. In particular, when a first order perturbation theory is applied recursively, like the one of [4], it can faithfully describe solutions of the full nonlinear theory, up to the first nonlinear time scale, as long as those solutions have *amplitudes* that remain sufficiently small during the evolution. In the context of scalar field collapse in AdS, this means that recursive perturbation theory faithfully describes the *non-collapsing* solutions up to $t \sim \epsilon^{-2}$. One can then employ the scaling symmetry of the approximate equations to establish the existence of these solutions at the vanishing amplitude limit ($\epsilon \to 0$), establishing this way an open set of stable solutions.

Gauge dependence of the problem

If one wants to establish, in a similar manner, an open set of collapsing solutions, at the vanishing amplitude limit, he would have to find an open set of singular solutions of the approximate/perturbative equations of motion. However, in that case one should be very careful about how to interpret the observed blow up. It could represent either a genuine singular solution to the full theory or it could merely signal the breakdown of the perturbative expansion by invalidating one of the approximations used during the derivation. A first attempt towards this direction appeared in the work of [48] in which they studied numerically the TTF system and reported a solution with an oscillating singularity. Although it was conjectured that this represents a genuine singular solution of the full theory, a physical understanding was missing. Later on, it was argued [52] and then numerically shown [53] that this oscillating singularity is a gauge artefact, raising doubts about the initial conjecture. In chapter 4 we show however that actually the combination of those results clarifies the situation and the observed singularity actually represents a large backreaction to the full nonlinear theory since it corresponds to a diverging redshift between the boundary and the center of the spacetime.

Fast and slow coherent cascades

In momentum space the formation of the black hole is realised as a cascade of energy towards modes of higher frequency. This behaviour bears resemblance to the Kolmogorov's theory of turbulence with one crucial difference however. In Kolmogorov's theory with a *random phase* ansatz, there is no transfer of energy at the $t \sim \epsilon^{-2}$ time scale and any interesting dynamics appears at the next order $t \sim \epsilon^{-4}$. An alternative was therefore proposed in [54] where a *coherent phase* ansatz was assumed and the corresponding power laws of the cascade were derived. In chapter 5 we studied numerically this phase ansatz by looking both at the resulting power laws or at the phase spectrum directly. Our results are in good agreement with the predictions of [54] but they leave open room for small corrections. In this chapter we also studied the disputed *two mode data* in AdS_4 and we conjectured, based on both numerical and analytical results, that they collapse in infinite *slow time* $\tau = \epsilon^2 t$.

Future directions

Until now, almost all of the studies have been concentrated to the spherically symmetric case and to first order in perturbation theory. The natural subsequent steps would be to move beyond spherical symmetry as well as to higher orders in perturbation theory. One might hope for example to unequivocally answer for or against the stability of AdS at the next time scale $t \sim \epsilon^{-4}$. A perhaps more ambitious expectation would be to have a resolution of the problem in terms of a theorem, either along the lines of the coherent cascade conjecture of [54] or some form of ergodic theorem for AdS.

Other possible directions would be to understand the nature of the problem outside the context of AdS. Namely whether the resulting phenomenological picture is unique to AdS or can be found in other systems as well? It is also very interesting to understand the case from the AdS/CFT perspective. As it was noted earlier in this work, the formation of the black hole in the bulk theory can be interpreted as the evolution towards thermal equilibrium for the boundary field theory. The fact that some excitations would thermalize, whereas some others would not, seems very puzzling. From this point of view the apparent relation of oscillating solutions to quantum revivals of the initial state in the boundary field theory [59] is also very interesting

Bibliography

[1] F. V. Dimitrakopoulos, B. Freivogel, and J. F. Pedraza, "Fast and Slow Coherent Cascades in Anti-de Sitter Spacetime," arXiv:1612.04758 [hep-th].

[2] F. V. Dimitrakopoulos, B. Freivogel, J. F. Pedraza, and I.-S. Yang, "Gauge dependence of the AdS instability problem," *Phys. Rev.* **D94** no. 12, (2016) 124008, arXiv:1607.08094 [hep-th].

[3] F. Dimitrakopoulos and I.-S. Yang, "Occasionally Extended Validity of Perturbation Theory: Persistence of AdS Stability Islands," arXiv:1507.02684 [hep-th].

[4] F. V. Dimitrakopoulos, B. Freivogel, M. Lippert, and I.-S. Yang, "Instability corners in AdS space," arXiv:1410.1880 [hep-th].

[5] F. V. Dimitrakopoulos, L. Kabir, B. Mosk, M. Parikh, and J. P. van der Schaar, "Vacua and correlators in hyperbolic de Sitter space," *JHEP* **06** (2015) 095, arXiv:1502.00113 [hep-th].

[6] D. Christodoulou and S. Klainerman, "The global nonlinear stability of the minkowski space," *Séminaire Équations aux dérivées partielles (Polytechnique)* (1993) 1–29.

[7] H. Friedrich, "On the existence of n-geodesically complete or future complete solutions of einstein's field equations with smooth asymptotic structure," *Commun. Math. Phys.* **107** no. 587-609, (1986) .

[8] P. Bizon and A. Rostworowski, "On weakly turbulent instability of anti-de Sitter space," *Phys.Rev.Lett.* **107** (2011) 031102, arXiv:1104.3702 [gr-qc].

[9] M. Dafermos, "unpublished,".

[10] A. Buchel, S. L. Liebling, and L. Lehner, "Boson stars in AdS spacetime," *Phys.Rev.* **D87** no. 12, (2013) 123006, arXiv:1304.4166 [gr-qc].

[11] M. Maliborski and A. Rostworowski, "Lecture Notes on Turbulent Instability of Anti-de Sitter Spacetime," *Int.J.Mod.Phys.* **A28** (2013) 1340020, arXiv:1308.1235 [gr-qc].

[12] V. Balasubramanian, A. Buchel, S. R. Green, L. Lehner, and S. L. Liebling, "Holographic Thermalization, Stability of AdS, and the FPU Paradox," arXiv:1403.6471 [hep-th].

[13] S. R. Green, A. Maillard, L. Lehner, and S. L. Liebling, "Islands of stability and recurrence times in AdS," *Phys. Rev.* **D92** no. 8, (2015) 084001, arXiv:1507.08261 [gr-qc].

[14] O. J. Dias, G. T. Horowitz, and J. E. Santos, "Gravitational Turbulent Instability of Anti-de Sitter Space," *Class.Quant.Grav.* **29** (2012) 194002, arXiv:1109.1825 [hep-th].

[15] O. J. Dias, G. T. Horowitz, D. Marolf, and J. E. Santos, "On the Nonlinear Stability of Asymptotically Anti-de Sitter Solutions," *Class.Quant.Grav.* **29** (2012) 235019, arXiv:1208.5772 [gr-qc].

[16] G. T. Horowitz and J. E. Santos, "Geons and the Instability of Anti-de Sitter Spacetime," arXiv:1408.5906 [gr-qc].

[17] . Dias and J. E. Santos, "AdS nonlinear instability: moving beyond spherical symmetry," *Class. Quant. Grav.* **33** no. 23, (2016) 23LT01, arXiv:1602.03890 [hep-th].

[18] A. Rostworowski, "Comment on AdS onlinear instability: moving beyond spherical symmetry, arXiv:1612.00042 [hep-th].

[19] Martinon, Grgoire and Fodor, Gyula and Grandclment, Philippe and Forgcs, Peter, "Gravitational geons in asymptotically anti-de Sitter spacetimes," arXiv:1701.09100 [hep-th].

[20] M. Maliborski and A. Rostworowski, "Time-Periodic Solutions in an Einstein AdS–Massless-Scalar-Field System," *Phys.Rev.Lett.* **111** no. 5, (2013) 051102, arXiv:1303.3186 [gr-qc].

[21] B. Craps, O. Evnin, and J. Vanhoof, "Renormalization group, secular term resummation and AdS (in)stability," arXiv:1407.6273 [gr-qc].

[22] I.-S. Yang, "Missing top of the AdS resonance structure," *Phys.Rev.* **D91** no. 6, (2015) 065011, arXiv:1501.00998 [hep-th].

[23] H. de Oliveira, L. A. Pando Zayas, and E. Rodrigues, "A Kolmogorov-Zakharov Spectrum in AdS Gravitational Collapse," *Phys.Rev.Lett.* **111** no. 5, (2013) 051101, arXiv:1209.2369 [hep-th].

[24] S. L. Liebling, "Nonlinear collapse in the semilinear wave equation in AdS space," *Phys.Rev.* **D87** no. 8, (2013) 081501, arXiv:1212.6970 [gr-qc].

[25] M. Maliborski, "Instability of Flat Space Enclosed in a Cavity," *Phys.Rev.Lett.* **109** (2012) 221101, arXiv:1208.2934 [gr-qc].

[26] A. Buchel, L. Lehner, and S. L. Liebling, "Scalar Collapse in AdS," *Phys.Rev.* **D86** (2012) 123011, arXiv:1210.0890 [gr-qc].

[27] P. Bizon, "Is AdS stable?," *Gen.Rel.Grav.* **46** (2014) 1724, arXiv:1312.5544 [gr-qc].

[28] M. Maliborski and A. Rostworowski, "What drives AdS unstable?," arXiv:1403.5434 [gr-qc].

[29] P. Bizon and A. Rostworowski, "Comment on "Holographic Thermalization, stability of AdS, and the Fermi-Pasta-Ulam-Tsingou paradox" by V. Balasubramanian et al," arXiv:1410.2631 [gr-qc].

[30] S. Trotzky, Y.-A. Chen, A. Flesch, I. P. McCulloch, U. Schollwöck, J. Eisert, and I. Bloch, "Probing the relaxation towards equilibrium in an isolated strongly correlated one-dimensional Bose gas," *Nature Physics* **8** (2012) 325–330, arXiv:1101.2659 [cond-mat.quant-gas].

[31] M. Gring, M. Kuhnert, T. Langen, T. Kitagawa, B. Rauer, M. Schreitl, I. Mazets, D. A. Smith, E. Demler, and J. Schmiedmayer, "Relaxation and Prethermalization in an Isolated Quantum System," *Science* **337** (2012) 1318–, arXiv:1112.0013 [cond-mat.quant-gas].

[32] A. D. Rendall, "Convergent and divergent perturbation series and the post-minkowskian approximation scheme," *Classical and Quantum Gravity* **7** no. 5, (1990) 803.

[33] P. Bizon, T. Chmaj, and A. Rostworowski, "Late-time tails of a self-gravitating massless scalar field revisited," *Class.Quant.Grav.* **26** (2009) 175006, arXiv:0812.4333 [gr-qc].

[34] P. Basu, C. Krishnan, and A. Saurabh, "A Stochasticity Threshold in Holography and and the Instability of AdS," arXiv:1408.0624 [hep-th].

[35] J. Abajo-Arrastia, E. da Silva, E. Lopez, J. Mas, and A. Serantes, "Holographic Relaxation of Finite Size Isolated Quantum Systems," *JHEP* **1405** (2014) 126, arXiv:1403.2632 [hep-th].

[36] M. A. Amin, E. A. Lim, and I. Yang, "A scattering theory of ultrarelativistic solitons," *Phys.Rev.* **D88** (2013) 105024, arXiv:1308.0606 [hep-th].

[37] M. A. Amin, E. A. Lim, and I.-S. Yang, "A Clash of Kinks: Phase shifts in colliding non-integrable solitons," *Phys.Rev.Lett.* **111** (2013) 224101, arXiv:1308.0605 [hep-th].

[38] P. Bizon and J. Jalmuzna, "Globally regular instability of AdS_3," *Phys.Rev.Lett.* **111** no. 4, (2013) 041102, arXiv:1306.0317 [gr-qc].

[39] J. Jalmuzna, "Three-dimensional gravity and instability of AdS_3," arXiv:1311.7409 [gr-qc].

[40] S. Lin and E. Shuryak, "Toward the AdS/CFT Gravity Dual for High Energy Collisions. 3. Gravitationally Collapsing Shell and Quasiequilibrium," *Phys.Rev.* **D78** (2008) 125018, arXiv:0808.0910 [hep-th].

[41] S. Bhattacharyya and S. Minwalla, "Weak Field Black Hole Formation in Asymptotically AdS Spacetimes," *JHEP* **0909** (2009) 034, arXiv:0904.0464 [hep-th].

[42] T. Albash and C. V. Johnson, "Evolution of Holographic Entanglement Entropy after Thermal and Electromagnetic Quenches," *New J.Phys.* **13** (2011) 045017, arXiv:1008.3027 [hep-th].

[43] V. Balasubramanian, A. Bernamonti, J. de Boer, N. Copland, B. Craps, *et al.*, "Holographic Thermalization," *Phys.Rev.* **D84** (2011) 026010, arXiv:1103.2683 [hep-th].

[44] S. Hawking and D. N. Page, "Thermodynamics of Black Holes in anti-De Sitter Space," *Commun.Math.Phys.* **87** (1983) 577.

[45] C. M. Will, "Post-Newtonian effects in N-body dynamics: conserved quantities in hierarchical triple systems," *Classical and Quantum Gravity* **31** no. 24, (2014) 244001, arXiv:1404.7724.

[46] P. Basu, C. Krishnan, and P. Bala Subramanian, "AdS (In)stability: Lessons From The Scalar Field," *Phys.Lett.* **B746** (2015) 261–265, arXiv:1501.07499 [hep-th].

[47] A. Buchel, S. R. Green, L. Lehner, and S. L. Liebling, "Universality of non-equilibrium dynamics of CFTs from holography," arXiv:1410.5381 [hep-th].

[48] P. Bizon, M. Maliborski, and A. Rostworowski, "Resonant dynamics and the instability of anti-de Sitter spacetime," arXiv:1506.03519 [gr-qc].

[49] A. Buchel, S. R. Green, L. Lehner, and S. L. Liebling, "Reply to "Comment on two-mode stability islands around AdS"," arXiv:1506.07907 [gr-qc].

[50] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, "Holographic representation of local bulk operators," *Phys.Rev.* **D74** (2006) 066009, arXiv:hep-th/0606141 [hep-th].

[51] A. Hamilton, D. N. Kabat, G. Lifschytz, and D. A. Lowe, "Local bulk operators in AdS/CFT: A Holographic description of the black hole interior," *Phys.Rev.* **D75** (2007) 106001, arXiv:hep-th/0612053 [hep-th].

[52] B. Craps, O. Evnin, and J. Vanhoof, "Ultraviolet asymptotics and singular dynamics of AdS perturbations," *JHEP* **10** (2015) 079, arXiv:1508.04943 [gr-qc].

[53] N. Deppe, "On the stability of anti-de Sitter spacetime," arXiv:1606.02712 [gr-qc].

[54] B. Freivogel and I.-S. Yang, "Coherent Cascade Conjecture for Collapsing Solutions in Global AdS," *Phys. Rev.* **D93** no. 10, (2016) 103007, arXiv:1512.04383 [hep-th].

[55] B. Craps, O. Evnin, P. Jai-akson, and J. Vanhoof, "Ultraviolet asymptotics for quasiperiodic AdS_4 perturbations," *JHEP* **10** (2015) 080, arXiv:1508.05474 [gr-qc].

[56] B. Craps, O. Evnin, and J. Vanhoof, "Renormalization, averaging, conservation laws and AdS (in)stability," arXiv:1412.3249 [gr-qc].

[57] C. Sulem, P. Sulem, and F. Helene, "Tracing complex singularities with spectral methods.," *Comput. Phys.* **50** no. 138, (1983) .

[58] N. Deppe and A. R. Frey, "Classes of Stable Initial Data for Massless and Massive Scalars in Anti-de Sitter Spacetime," *JHEP* **12** (2015) 004, arXiv:1508.02709 [hep-th].

[59] E. da Silva, E. Lopez, J. Mas, and A. Serantes, "Collapse and Revival in Holographic Quenches," *JHEP* **04** (2015) 038, arXiv:1412.6002 [hep-th].

Contributions to Publications

Below I will describe my personal contributions to the publications on which this thesis is based.

[4] F.V. Dimitrakopoulos, B. Freivogel, M. Lippert and I. S. Yang
Position space analysis of the AdS instability problem
JHEP **1508**, 077 (2015), arXiv:1410.1880 [hep-th].
I was responsible for subsection 3.1, section 4 and I contributed to all conceptual discussions to all parts and topics discussed in the paper.

[3] F. Dimitrakopoulos and I-S. Yang
Conditionally extended validity of perturbation theory: Persistence of AdS stability islands
Phys. Rev. D **92**, 083013 (2015), arXiv:1507.02684 [hep-th].
I was responsible for sections 2 and 3 as well as appendix A. I also contributed to all conceptual discussions.

[2] F.V. Dimitrakopoulos, B. Freivogel, J.F. Pedraza and I-S. Yang
Gauge dependence of the AdS instability problem
Phys. Rev. D **94**, 124008 (2016), arXiv:1607.08094 [hep-th].
I was the main author of this article and responsible for all calculations. I contributed to all conceptual discussions.

[1] F.V. Dimitrakopoulos, B. Freivogel and J.F. Pedraza
Fast and slow coherent cascades in AdS
arXiv:1612.04758 [hep-th].
I was responsible for section 4 and I conducted, together with J.F. Pedraza all the numerical computations. I contributed to all conceptual discussions.

Summary

NONLINEAR DYNAMICS AND THE (IN)STABILITY OF AdS

Context

Historically, stability considerations and perturbation theory date back to the era of celestial mechanics and the question of stability of the solar system over long time scales. Between 1609 and 1618 Johannes Kepler determined the trajectories of the planets as they revolve around the Sun. Following the work of Copernicus, Kepler placed the Sun at the centre of the universe and based on observations of the famous astronomer of the time Tycho Brache, he succeeded to show that planets move in ellipses around the Sun and at the end of the revolution the planets find themselves back to where they started.

However, this picture of a perfectly stable solar system would be soon challenged. After Isaac Newton developed his theory about gravity, he derived the Keplerian orbits by restricting to the interaction of a planet with the Sun alone. Although this is the leading contribution to the gravitational force exerted to each planet, it is not the only one. Planets attract each other as well. When these perturbations are taken into account they might lead to small effects which accumulate in the course of time destroying in that way the Keplerian orbits.

The study of the stability of the solar system has led to remarkable discoveries in Physics and Mathematics with the most prominent one being perhaps the celebrated Kolmogorov-Arnold-Moser (KAM) theory in which it was rigorously shown that both stable and unstable orbits exist depending on whether the ratio of the unperturbed frequencies is a rational number.

Newton's theory was superseded when Albert Einstein published in 1915 his theory of gravitation, known as General Relativity (GR). According to Einstein, gravity is not a force but rather the manifestation of the geometry of spacetime in which the

masses move. Massive objects curve the spacetime and spacetime back-reacts to the masses by dictating them which paths they should follow. Einstein's equations possess three vacuum solutions, namely three different empty spacetimes depending on whether the cosmological constant of the theory is positive (de Sitter), zero (Minkowski) or negative (Anti-de Sitter). The most important question regarding a vacuum state is whether it is stable under small perturbations.

Motivation of research

The stability of the vacuum solutions of GR comes second (perhaps even first) only to the stability of the solar system and has led to one of the greatest developments in mathematical relativity [6]. Of the three vacuum spacetimes the two where proven to be stable long ago [6, 7]. The stability of the third one (AdS) was not even raised, let alone answered, until very recently [8].

Anti-de Sitter (AdS) spacetime plays a prominent role in modern Theoretical Physics mainly due to its role in the only concrete example of a gauge/gravity duality, the AdS/CFT correspondence [5]. In this picture, a Quantum Field Theory (QFT) living on the boundary of AdS is equivalent to a String Theory in the AdS background. Despite the great importance of (asymptotically) AdS spacetime(s), the study of its (nonlinear) stability was initiated only very recently, albeit it was earlier conjectured that AdS would be nonlinearly unstable [9].

Results

Chapter 2 :

In this chapter we presented an alternative and complementary method of studying the problem of the stability of AdS, directly in position space. We derived an approximate/perturbative equation of motion which has a similar scaling symmetry, as the one observed in Fourier space methods (TTF equations). We also showed that the gravitational interaction near the center of the spacetime obeys an exact antisymmetry under time reversal and therefore it is equally likely that the energy be focused or defocused. Finally, we touched on the thermalization process of the boundary field theory and we argued that even if black holes form in the first nonlinear time scale (ϵ^{-2}), it doesn't always represent efficient thermalization of the boundary theory.

Chapter 3 :

Approximating nonlinear dynamics with a truncated perturbative expansion may be accurate for a while, but it in general breaks down at a long time scale that

is one over the small expansion parameter (in our case, $t \sim \epsilon^{-2}$). In this chapter we presented cases where such a break down doesn't happen and the perturbation theory is valid up to this time scale, as long as it is applied recursively. There are cases where one can try and *guess* the form of the (regular) solution and then set up a smarter perturbation theory that reproduces this solution. Such is the case of the Two Time Framework (TTF) for example. As we argue in this chapter, if one uses for example a perturbation theory similar to the one of [4], the *regular* solutions of this approximate equation are valid up to $t \sim \epsilon^{-2}$ and not only for $t < \epsilon^{-2}$, as conventional wisdom would suggest. Using these results we then establish the existence of an open set of initial conditions that do not collapse up to this *long time scale*.

Chapter 4 :

An effort to establish collapsing solutions at the vanishing amplitude limit $\epsilon \to 0$ was made in [48] where solutions of the TTF that develop an *oscillating singularity* were reported. However this singularity is merely a *gauge artefact*. One can work in a different gauge and not observe this blow up of the derivatives of the phases. In this chapter we showed that these solution are genuine singular solutions and the discrepancy of the results in the two gauges was realized as a diverging redshift between the boundary and the center of the spacetime.

Chapter 5 :

In this chapter we studied the amplitude and the phase dynamics of small perturbations in AdS_4 using the *Two Time Framework* approximation. Our intention was to test the *phase coherent cascade* conjecture of [54] for different initial data. This is done in two ways; either by directly checking the phase coherent ansatz, eq. (5.8) or by studying the resulting power-law for the spectrum of collapsing solutions. Our results suggest that this ansatz works pretty well, however small modifications/improvements might be necessary. We found that the energy spectrum of narrow Gaussian wavepackets scales as $E_n \sim n^{-1}$ and also the phases are coherently aligned ($B_n \sim n$), although some small divergences from this linear behaviour were seen.

We also studied the contentious two-mode equal energy data, and we conjectured that they belong to a new class of solutions that collapse at infinite *slow time* τ, at the vanishing amplitude limit.

Outlook

Stability considerations have led to tremendous discoveries in Theoretical Physics and Mathematics and the stability of AdS could not be an exception to this rule.

Although the question has not yet been unequivocally answered, and perhaps there is a long way to go, the studies so far have already unveiled a very rich phenomenology. However there are still many questions to be answered, like what happens for example if we abolish spherical symmetry, or what is the fate of the perturbations at longer time scales.

The problem at hand is not only interesting from the pure mathematical point of view of GR, but can shed light to understanding the thermalization process of the boundary theory via the AdS/CFT correspondence. There have already been very interesting developments in this direction [4, 12], as well as interesting results in relating quantum revivals (of the boundary QFT) with bouncing geometries in the bulk [59].

Samenvatting

Niet-lineaire dynamica and de (in)stabiliteit van anti-de Sitter

Context

Historisch gezien gaan stabiliteitsoverwegingen en storingsrekening terug naar de tijd van de klassieke mechanica en toen men vragen ging stellen over de stabiliteit van het zonnestelsel voor grote tijdschalen. Voortbordurend op het werk van Copernicus, plaatste Johannes Kepler de zon in het midden van het zonnestelsel om vervolgens de banen van onze planeten te bepalen. Met de observaties van de bekende astronoom Tycho Brache, kon hij laten zien dat de banen van onze planeten ellipsen zijn en dat de planeten weer terugkomen bij hun begin positie na elke omlooptijd.

Dit plaatje zou echter snel op de proef gesteld worden. Nadat Newton zijn wetten had geformuleerd en de ellips banen van Kepler had afgeleid door alleen de interactie van de zon met de planeten mee te nemen, begon het ook duidelijk te worden dat de aantrekkende kracht van de planeten onderling ook een belangrijk effect zou kunnen zijn. Hoewel de zon de grootste zwaartekracht levert, kunnen deze onderlinge interacties voor kleine veranderingen zorgen die groeien in de tijd en zo grote verandering kunnen worden die de Kepler-banen kunnen vernietigen.

De studie naar dit soort vraagstukken over stabiliteit onder kleine verstoringen heeft tot buitengewone ontdekkingen geleidt in zowel natuur- en wiskunde. Het belangrijkste resultaat is waarschijnlijk het bekende Kolmogorov-Arnold-Moser (KAM) theorema waarin werd aangetoond dat stabiele of instabiele planeet banen kunnen bestaan afhankelijk van of de verhouding tussen de onverstoorde frequenties een rationaal getal is.

De vraagstukken over stabiliteit kregen een nieuwe twist toen Albert Einstein in 1915 zijn theorie van gravitatie, de Algemene Relativiteitstheorie publiceerde. Vol-

gens de theorie van Einstein, is gravitatie geen kracht, maar een manifestatie van de meetkunde van de ruimtetijd waarin massas bewegen. Massieve objecten krommen ruimtetijd en de ruimtetijd reageert hierop door de massas voor te schrijven in welke paden zij moeten bewegen. De vergelijkingen van Einstein laten drie vacuum oplossingen toe, namelijk drie verschillende lege ruimtetijden die gekarakteriseerde worden door de kosmologische constante. Bij een positieve (negatieve) constante spreken we over de Sitter (anti-de Sitter) ruimtetijd en een ruimtetijd met een kosmologische constante gelijk aan nul noemen we Minkowski. Bij het bestuderen van deze vacuum oplossingen, is een van de belangrijkste vragen of ze stabiel zijn onder kleine verstoringen.

Motivatie van het onderzoek

Stabiliteit van vacuum oplossingen van Einsteins theorie van gravitatie is, na de stabiliteit van ons zonnestelsel, een van de belangrijkste vraagstukken en heeft voor een van de grootste ontwikkelingen in de wiskunde achter de algemene relativiteitstheorie gezorgd [6]. Stabiliteit van twee van de drie vacuum oplossingen was een lange tijd geleden al bewezen [6, 7]. De stabiliteit van de derde oplossing, anti-de Sitter (AdS) werd niet eens beschouwd, laat staan werd zijn stabiliteit bewezen, tot vrij recentelijk [8].

Anti-de Sitter ruimtetijd speelt een belangrijke rol in de moderne theoretische natuurkunde. Dit komt doordat het terugkomt in een concreet voorbeeld van de ijk/gravitatie dualiteit, de AdS/CFT correspondentie. In deze dualiteit is een quantum velden theorie die leeft op de rand van AdS volkomen gelijk aan een snaartheorie in de AdS ruimtetijd. Ondanks deze reden is er pas recentelijk in detail gekeken naar (niet-lineaire) stabiliteit van AdS, ook al vermoedde men eerder al dat AdS niet-lineair instabiel zou zijn [9].

Resultaten

Hoofdstuk 2:

In dit hoofdstuk presenteren we een alternatieve en complementaire methode om het probleem van de stabiliteit van AdS direct in reele ruimte te bestuderen. We leiden een benaderende/perturbatieve bewegingsvergelijkingen af welke een zelfde schalingssymmetrie heeft als die gezien wordt door de Fourier ruimte methoden (TTF vergelijkingen) te gebruiken. We laten ook zien dat de gravitationele interactie nabij het centrum van de ruimtetijd een exacte antisymmetie onder tijdsomkering heeft en daarom zal focusering of defocusering van energie met gelijke kansen optreden. Tenslotte, beschouwen we in het kort de thermalisatie van de randtheorie en beargumenteren we dat zelfs als zwarte gaten gevormd worden in

de eerste niet-lineaire tijdschalden (ϵ^{-2}), er niet altijd effectieve thermalisatie in de randtheorie is.

Hoofdstuk 3:

Het benaderen van niet-lineaire dynamica met een beperkte perturbatieve expansie is accuraat voor een tijdje, maar verliest in het algemeen precisie bij lange tijdschalen die inverse proportioneel is aan de expansie parameter (in ons geval, $t \sim \epsilon^{-2}$). In dit hoofdstuk laten we een aantal gevallen zien waarbij zon verlies van precisie niet optreedt en de perturbatieve expansie ook geldt bij deze tijdschalen, zolang het recursief wordt toegepast. Er zijn gevallen waarbij men kan proberen een vorm van de (oppassende) oplossing te gebruiken om een slimmere storingstheorie op te zetten die deze oplossing kan reproduceren. Dit is bijvoorbeeld het geval bij het *Two Time Framework* (TTF). We laten inderdaad zien in dit hoofdstuk dat als men een storingsrekening opzet lijkend op die van [4], dan zijn de oppassend oplossingen van deze benadering geldig voor tijden tot $t \sim \epsilon^{-2}$ and niet alleen voor $t < \epsilon^{-2}$ zoals men zou verwachten. Door deze resultaten te gebruiken kunnen we het bestaan van een open verzameling van begincondities bevestigen die niet ineenstorten tot deze *lange tijdschaal*.

Hoofdstuk 4:

Bij pogingen om ineenstortende oplossingen bij de verdwijnende amplitude limiet $\epsilon \to 0$ te construeren [48] werden oplossingen van de TTF gevonden die een oscillerende singulariteit hadden. Echter is deze singulariteit een ijk-artefact; in een andere ijk keuze wordt deze singulariteit van afgeleiden van de fasen niet geobserveerd. In dit hoofdstuk laten we zien dat deze oplossingen echte singuliere oplossingen zijn en dat de tegenstrijdigheid van de resultaten in de twee ijk keuzes gerealiseerd werd als een divergerende roodverschuiving tussen de rand en het centrum van de ruimtetijd.

Hoofdstuk 5:

In dit hoofdstuk onderzoeken we de amplitude en fase van de dynamica van de kleine verstoringen in AdS4 door gebruik te maken van de Two Time Framework benadering. Onze intentie was het testen van een vermoeden, het phase coherent cascade vermoeden, van [54] voor verschillende begincondities. Dit hebben we op twee manieren gedaan; ofwel door direct de fase coherente ansatz te verifieren, verg. (5.8) dan wel door het machtsverband van het spectrum van de ineenstortende oplossingen te bestuderen. De resultaten suggereren dat deze ansatz best wel goed werkt, maar dat kleine aanpassingen eventueel nodig zijn. Wij vinden dat het energie spectrum van smalle Gaussiaanse golfpakketjes schaalt als $E_n \sim n^{-1}$ en ook dat de fasen coherent uitgelijnd zijn ($B_n \sim n$), hoewel enkele kleine divergenties van het lineaire gedrag ook gezien werden.

We hebben ook de omstreden two-mode equal energy data onderzocht en wij vermoeden dat zij tot een nieuwe klasse van oplossingen behoren die ineenstorten bij een oneindige *slow time* τ , in de verdwijnende amplitude limiet.

Vooruitblik

Stabiliteitsvraagstukken hebben tot waanzinnige ontdekkingen in de theoretische natuurkunde en wiskunde geleidt en de stabiliteit van AdS is hierop geen uitzondering. Hoewel de vraag nog niet ondubbelzinnig beantwoord is, en misschien is er nog een lange weg te gaan, maar het onderzoek tot dusver heeft al een zeer rijke fenomenologie laten zien. Toch zijn er nog steeds veel vragen die beantwoord moeten worden, zoals wat er gebeurt als er bijvoorbeeld geen sferische symmetrie wordt aangenomen, of wat het lot is van kleine verstoringen op grote tijdschalen.

De stabiliteit van AdS is niet alleen interessant vanuit de puur wiskundige aspecten van de algemene relativiteitstheorie, maar kan ook helpen bij het begrijpen van thermalisatie van de randtheorie via de AdS/CFT correspondentie. Er zijn al enkele zeer interessante ontwikkelingen gaande in deze richting [4, 12], alsmede een interessante connectie tussen quantum opwekkingen (van de randtheorie) en stuiterende meetkundes in de binnenkant van AdS [59] .

Acknowledgments

This thesis is the final destination of a wonderful journey that started roughly four years ago and would not have been possible (or so pleasant) without the invaluable assistance of many *beautiful* people. I still remeber my first impression from UvA and its inhabitants during the GRAPPA PhD recruitment days. I was fascinated by the environment and thankfully I was not erred.

I would like to commence with expressing my sincere gratitude to my supervisor, Ben Freivogel, for his guidance, inspiration, support and of course his patience. For a beginning PhD student, it is very important to have a balance between the amount of workload that is assigned to him and the free space he is given to explore his own interests. I believe Ben excelled in finding the balance between the two. He didn't let me get overwhelmed by the infinite possiblilites and my low expertise by giving my structured assistance and hands-on tasks to perform but at the same time he gave me enough space to explore other research topics of my own interest and initiate new collaborations. Thank you very much for everything. I have learned a lot from you.

I would also to like to thank my collaborators, Juan, Mat, Jan Pieter, Laurens, Benjamin, Alex, Joanna, Alejandra, Eva and Nabil for all the wonderful conversations and exchange of knowledge. Especially, I would like to thank I-Sheng Yang, for his invaluable help and guidance during the first years of my PhD.

Pursuing a PhD can be a very lonely endeavor. Thankfully this was not the case for me and this is due to everyone in the insitute of Theoretical Physics who made my daily life way more bearable. I would like to start by thanking my *fellow travellers*, Laurens, Diego C., Irfan, Robert, Hamish, Michael, Sebastian, Niki, Jorrit, Sam, Gerben, Manus, Francesca, Sagar and Eva for sharing the load and the stress throughout the years. I also owe a special "thank you", together with a crate of beers, to Jorrit for taking care the Dutch summary! I would also like to thank everyone else in the ITF for all the knowledge I acquired from them. Erik, Jan, Diego H., Alejandra, Gianfranco, Shinichiro, Jan Pieter, Daniel, Miranda, you gave me more than you know. Thank you.

Special thanks is also reserved to the support staff of IoP, especially Natalie, Anne-

Marieke, Joost and Fatima who made our lives much easier, by helping us with all the practical issues, and much more pleasent by providing us with endless quantities of coffee and cookies!

I would also like to thank seperately Bert, Diego C., Laurens, Gerben and Alex for our short term career as musicians and Irfan for all our philosophical conversations. I had a lot of fun guys!

Obtaining a PhD, and especially one in Theoretical Physics, is far from an easy task. The intrinsic difficulties, the uncertainty, the anxiety and the frustration can turn anyone to a distant and introvert person. And this has been the case for me in periods of huge stress and pressure. For the understanding and support during these dark eras, I would like to thank my family, my friends and everyone in my life. I owe you a lot...

A man needs a little madness, or else... he never dares cut the rope and be free.

— Nikos Kazantzakis