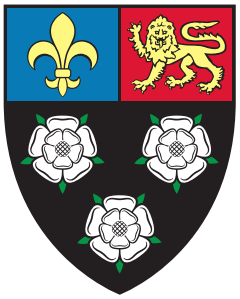
Wave Equations on Curved Spacetimes



Felicity Clare Eperon

Department of Applied Mathematics and Theoretical Physics
University of Cambridge

This dissertation is submitted for the degree of
Doctor of Philosophy

King's College

September 2019

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Preface and specified in the text.

It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text. I further state that no substantial part of my dissertation has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution except as declared in the Preface and specified in the text.

As specified at the beginning of each chapter, the results presented in this thesis have been published in the following papers:

F. C. Eperon, H. S. Reall, and J. E. Santos, "Instability of supersymmetric microstate geometries", *JHEP* **10** (2016) 031, arXiv:1607.06828.

F. C. Eperon, "Geodesics in supersymmetric microstate geometries", *Class. Quant. Grav.* **34** (2017), no. 16, 165003, arXiv:1702.03975.

O. J. C. Dias, F. C. Eperon, H. S. Reall, and J. E. Santos, "Strong cosmic censorship in de Sitter space", *Phys. Rev.* **D97** (2018), no. 10, 104060, arXiv:1801.09694.

F. C. Eperon, H. S. Reall, and J. J. Sbierski, "Predictability of subluminal and superluminal wave equations", arXiv:1802.03020.

The last of these has recently been accepted for publication in Communications in Mathematical Physics.

Felicity Clare Eperon
September 2019

Acknowledgements

First and foremost I would like to thank my supervisor, Harvey Reall, for stimulating discussions, helpful suggestions and invaluable advice. Without such excellent guidance, the last few years would have been far less interesting and productive, and this thesis would not exist.

I am very lucky to have such kind and supportive parents, who are so scientifically curious and have passed on their enthusiasm for research. They have always encouraged me in whatever I've decided to do, be it climbing, riding or indeed maths and physics. I probably knew my times tables before I could speak, thanks to Dad singing them to ever higher multiples while attempting to put me to sleep as a baby; after that there was always new physics at home to experiment with, from iron filings and magnets to light sensitive lego cars. They have continued to be interested in my work, resulting in entertaining discussions after my attempts to explain it.

Pavilion B has been an inspirational place to study general relativity. Most importantly, there have been many people to discuss maths and physics (and other non-scientific topics) with, in particular Guiseppe, Dejan, Toby, Jorge, Joe, and especially Bogdan, who constantly reminds me how exciting physics is.

I'd like to thank Ben, the best housemate imaginable, Conor and Lee for helping me appreciate the excellent local pubs and being such good friends for the eight years I've spent in Cambridge. My climbing friends, Murray, Stephen, Izzy, Luke, James K, Holly, Cameron, Alex, James L and Omar, have allowed me to regularly escape Cambridge to wilder places with mountains and sea cliffs; often I come back from trips with more ideas and motivation for maths.

Finally, thanks to Ed, for everything.

Financial support

This research has been supported by an STFC studentship.

Abstract

There are still many important unsolved problems in general relativity, two of which are the stability problem and the strong cosmic censorship conjecture. Both of these are extremely difficult, but we can gain some insight by studying linear versions of them. These simplifications give rise to wave equations on curved spacetimes.

We investigate the classical stability of supersymmetric, asymptotically flat microstate geometries with 5 non-compact dimensions. These geometries possess an evanescent ergosurface, where there are stably trapped null geodesics that have zero energy relative to an observer at infinity. We give a heuristic argument as to why this may lead to a non-linear instability, which can be seen at the linear level by studying the wave equation. We calculate the quasinormal mode frequencies and find that, due to the stable trapping, the rate of decay is extremely slow. This suggests that stability is very unlikely at the non-linear level. The behaviour of geodesics is crucial for this, so we also investigate the geodesics in these microstate geometries in some detail.

There has recently been evidence to suggest that Christodoulou's formulation of the strong cosmic censorship conjecture is actually false for Reissner-Nordstron-de Sitter black holes sufficiently close to extremality. We investigate this problem for the more physical rotating Kerr-de Sitter black holes. We look at the linear problem, and find that solutions of the wave equation decay sufficiently slowly to suggest that strong cosmic censorship is respected.

The two problems mentioned so far are both related to predictability in general relativity. We investigate predictability more generally for subluminal and superluminal Lorentz-invariant scalar wave equations. We study the Born-Infeld scalar in two dimensions, which has both a superluminal and subluminal formulation. Contrary to previous expectation, we find that, at least in some sense, the subluminal equation behaves worse the superluminal equation. It is possible to have multiple different maximal globally hyperbolic developments arising from the same initial data for the subluminal equation, but the solution is unique in the superluminal case.

Table of contents

Chapter 1

Introduction

General relativity (GR) is one of the most remarkable theories of physics, giving both a mathematically beautiful and physically reliable description of gravity. A spacetime (M, g) is a four dimensional Lorentzian manifold M equipped with a metric g and Levi-Civita connection. Einstein's equation [1]

$$G_{ab} + \Lambda g_{ab} = 8\pi T_{ab} \tag{1.1}$$

dictates that the curvature of the spacetime is conditioned by the energy-momentum tensor T_{ab}. This encodes information about the energy and matter present and, crucially for the Einstein equation, is conserved, $\nabla^a T_{ab} = 0$. The effect of gravity on observers can easily be determined, since in GR free particles follow timelike geodesics if they are massive, or null geodesics if they are massless.

GR predicts various phenomena that are absent in Newtonian gravity, so it is possible to perform various tests to verify the theory (see [2] for a review of these). The first three tests were carried out within 40 years of Einstein introducing the theory in 1915: the perihelion precession of Mercury had already been noticed prior to then [3]; light deflection around the sun was observed in 1919 [4] and the gravitational redshift of light was measured in 1954 [5]. None of these effects are expected in Newtonian gravity but are correctly predicted by GR. More recently there has also been significant work done to compare GR to alternative theories of gravity, in particular constraining parameters that measure the deviation of other theories from GR (the parametrised post-Newtonian formalism) [2].

Perhaps one of the most surprising and interesting aspects of GR is the existence of black holes. The first and simplest black hole spacetime, Schwarzschild, was unexpectedly found as a solution to the vacuum Einstein equation in 1916 [6], very shortly after GR itself was discovered. Since nothing, not even light, can escape from a black hole they are very hard to

observe. Until recently, the evidence for the presence of black holes in the universe came mostly from the trajectories of stars near the centres of galaxies and from the formation of accretion disks around them (see *e.g.* [7, 8]). However, a century after Einstein first introduced GR, more direct evidence for black holes came from the detection of gravitational waves produced by two black holes merging [9]. Since then many more similar observations have been made, providing more evidence for black holes.

GR is a classical theory of gravity; although there are several contenders for a quantum theory, such as string theory and loop quantum gravity, a consistent and experimentally verified theory has yet to be discovered. It is hoped that by being able to probe regions of spacetime with high curvature, where quantum effects are more likely to be noticeable, we may gain some insight into an appropriate quantum theory.

There are many interesting phenomena associated with black holes, some of which can be studied by looking at wave equations on various black hole backgrounds. In this thesis, we investigate different uses and properties of wave equations on curved spacetimes. In the remainder of the Introduction we will introduce black holes in more detail. We then discuss the stability problem in GR and the strong cosmic censorship conjecture; some insight into both of these problems can be gained by studying wave equations on curved spacetimes. In chapter 2 we investigate the stability of supersymmetric microstate geometries. This is related to behaviour of geodesics in these spacetimes, which we examine more thoroughly in chapter 3. In chapter 4 we move onto the strong cosmic censorship conjecture in Kerr-de Sitter spacetimes. These are all related to the predictability of GR; in chapter 5 we study predictability more generally (*i.e.* not just for GR) by looking at uniqueness of solutions to subluminal and superluminal wave equations.

1.1 Black Holes

A large proportion of this thesis is related to black holes in some way, so in this section we will introduce various black hole spacetimes in order to describe some of their key features.

Such a spacetime contains a black hole region bounded by an event horizon that observers can cross in an inwards direction only to enter the interior of the black hole; it is impossible for an observer to escape from this region. This black hole region is not in causal contact with future infinity: signals from within the black hole cannot reach an observer in the far distance. The white hole region is the 'opposite' of a black hole, since observers can only travel out of this region.

Definition 1.1.1. *The* black hole region $\mathfrak{B}$ *of a spacetime* (M,g) *that is asymptotically flat at null infinity, in the sense of [10], is the set of points* $\mathfrak{B} = M \setminus J^-(\mathscr{I}^+)$. *The* future event

horizon *is the boundary of* $\mathfrak{B}$ *(in* M*),* $\mathscr{H}^+ = M \cap \dot{J}^-(\mathscr{I}^+)$*. Analogously, the* white hole region *is* $\mathfrak{W} = M \setminus J^+(\mathscr{I}^-)$ *with* past event horizon $\mathscr{H}^- = M \cap \dot{J}^+(\mathscr{I}^-)$.

We now describe four well-known black hole spacetimes and remark on some of their key properties, see *e.g.* [10] for more details.

1.1.1 Schwarzschild

The first example of a black hole is the original asymptotically flat, spherically symmetric, non-rotating static Schwarzschild black hole with mass M. The metric is given by

$$ds^2 = -f dt^2 + f^{-1} dr^2 + r^2 d\Omega^2, \qquad f(r) = 1 - \frac{2M}{r} \tag{1.2}$$

and $d\Omega^2$ is the usual metric on the 2-sphere. The coordinate range $r > 2M$ defines the exterior of the black hole (region I of the Penrose diagram in fig. 1.1). To extend across the future event horizon $\mathscr{H}^+$ at $r = 2M$ we can define ingoing Eddington-Finkelstein coordinates (see e.g. [10]) by (v, r, θ, ϕ) where

$$v = t + r_* \qquad \text{where } \frac{dr_*}{dr} = f^{-1}.$$

The metric in these coordinates is regular at the horizon, so can be extended to $r < 2M$, the interior of the black hole (region II of the Penrose diagram in fig. 1.1). By introducing Kruskal coordinates [10] the spacetime can be extended to regions III and IV of the Penrose diagram; these are the white hole and a new asymptotically flat region respectively.

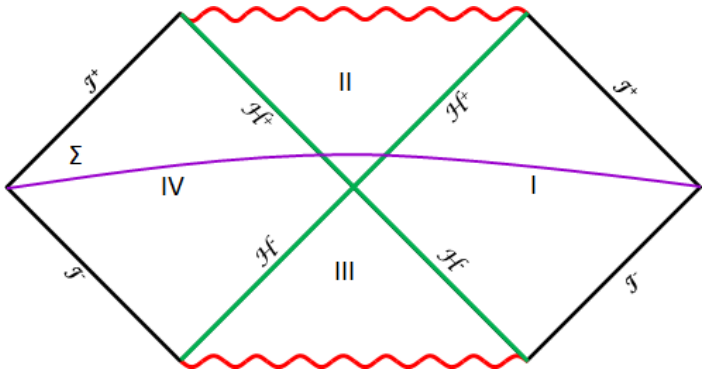


Fig. 1.1 Penrose diagram for Schwarzschild. The past/future event horizons $\mathscr{H}^{\pm}$ are the green lines, the purple line is a Cauchy surface Σ, the red squiggles represent the singularity at $r = 0$.

As for all black holes, the event horizon is a null hypersurface; in this case it is at $r = 2M$. Schwarzschild is asymptotically flat, so 'infinity' is made up of future/past null infinity (the null hypersurfaces $\mathscr{I}^{\pm}$), spacelike infinity i^0 and future/past timelike infinity $i^{\pm}$.

Σ is a Cauchy surface for the maximally extended spacetime: a spacelike hypersurface for which the domain of dependence is the whole of the Penrose diagram. Suitable initial data prescribed on Σ can be evolved using the Einstein equations to determine the entire spacetime: it is the maximal globally hyperbolic development of initial data on Σ. Note that the boundary is either infinity or the spacelike singularity at $r = 0$, across which it has been proven that one cannot extend the spacetime continuously (*i.e.* we have C^0-inextendibility) [11, 12]; it is impossible to extend the spacetime in any way.

One other noteworthy feature is that there is a photon sphere at $r = 3M$; this is made up of null geodesics that stay at constant radius. This is an example of *unstable trapping*: the geodesics stay in a bounded region of space, but if they are perturbed slightly they either fall into the black hole or escape to infinity.

1.1.2 Reissner-Nordström

The Reissner-Nordström black hole is charged: instead of being a solution to the vacuum Einstein equation it is a solution to the Einstein-Maxwell system

$$\begin{aligned} G_{ab} = 8\pi T_{ab} \quad &\text{and} \quad \nabla^b F_{ab} = 0, \quad dF = 0, \\ \text{where} \quad T_{ab} &= \frac{1}{4\pi}\left(F_a^c F_{bc} - \frac{1}{4} g_{ab} F^{cd} F_{cd}\right) \end{aligned} \tag{1.3}$$

and $F = dA$ for some 1-form A. The unique asymptotically flat and spherically symmetric solution is

$$ds^2 = -f(r)dt^2 + f(r)^{-1}dr^2 + r^2 d\Omega^2, \qquad A = -\frac{Q}{r}dt - P\cos\theta d\phi \tag{1.4}$$

$$f(r) = 1 - \frac{2M}{r} + \frac{e^2}{r^2}, \qquad e^2 = \sqrt{Q^2 + P^2} \tag{1.5}$$

where M is the mass, Q and P are the electic and magnetic charges respectively, and $M > e$ for non-extremal Reissner-Nordström black holes.

$f(r)$ has zeroes at $r_{\pm} = M \pm \sqrt{M^2 - e^2}$: the exterior of the black hole is $r > r_+$ and the event horizon is at $r = r_+$. In a similar way to Schwarzschild, by changing to regular coordinates the spacetime can be extended across both the inner and outer horizons. The Penrose diagram for the maximal analytic extension of Reissner-Nordström is shown in fig. 1.2.

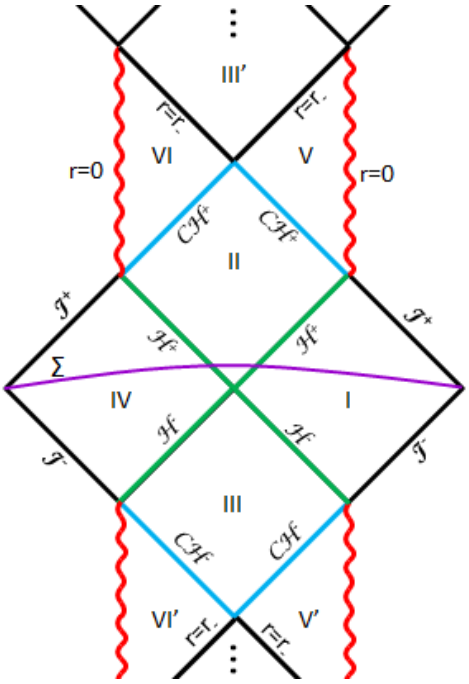


Fig. 1.2 Maximal analytic extension of Reissner-Nordström. The green lines $\mathscr{H}^{\pm}$ are the past/future event horizons, the vertical red squiggles are the (timelike) singularity at $r = 0$, the purple line Σ is a Cauchy surface for regions $I - IV$ and the blue lines $\mathscr{C}\mathscr{H}^{\pm}$ are the past/future Cauchy horizons.

Note that there are two horizons at $r = r_{\pm}$. The outer horizon at $r = r_{+}$ is the event horizon, which is also present in Schwarzschild. On the other hand, there is no equivalent of the inner horizon at $r = r_{-}$ in Schwarzschild; in that case there is only the spacelike singularity inside the black hole. This inner horizon is known as the *Cauchy horizon*: it is the boundary of the region of spacetime where the solution to the Cauchy problem, with initial data on the spacelike hypersurface Σ, is unique. We will discuss this in much more detail in section 1.3; for now we will simply remark that there are infinitely many different ways to smoothly extend the spacetime across the Cauchy horizon as a solution of the equations of motion [13]. To obtain the Penrose diagram in fig. 1.2 we have chosen to extend across the Cauchy horizon analytically.

1.1.3 Kerr

The Kerr family of spacetimes describes rotating black holes that are solutions of the vacuum Einstein equations (expected to be the unique such stationary solutions [14]). They are of particular interest because we expect these to be astrophysical black holes, *i.e.* the ones that are actually present in our universe and therefore might observe.

They are described by two parameters, (M,a), where M is the mass and $a = J/M$ describes the rotation; J is the angular momentum and for non-extremal black holes $a^2 < M^2$. In Boyer-Lindquist coordinates the metric is

$$\begin{aligned} ds^2 = &-\frac{\Delta - a^2\sin^2\theta}{\Sigma}dt^2 - 2a\sin^2\theta\frac{r^2+a^2-\Delta}{\Sigma}dtd\phi \\ &+\left(\frac{(r^2+a^2)^2 - \Delta a^2\sin^2\theta}{\Sigma}\right)\sin^2\theta d\phi^2 + \frac{\Sigma}{\Delta}dr^2 + \Sigma d\theta^2, \\ \text{where}\quad &\Sigma = r^2 + a^2\cos^2\theta, \quad \Delta = r^2 - 2Mr + a^2. \end{aligned} \tag{1.6}$$

Kerr exhibits some very similar features to Reissner-Nordström, although it is not spherically symmetric. For instance, there is an event horizon at $r = r_+$ and a Cauchy horizon at $r = r_-$, where $r_- < r_+$ are defined by $\Delta(r_\pm) = 0$. By changing to Kerr coordinates it is possible to see that the spacetime is regular at $r = r_\pm$, and can be smoothly extended across the horizons. As with Reissner-Nordström, we have chosen to analytically extend across the Cauchy horizon to obtain the Penrose diagram in 1.3.

The lack of spherical symmetry prevents us from drawing a Penrose diagram. However, it is possible to draw a Penrose diagram for the 2-dimensional submanifold corresponding to the axis of symmetry ($\theta = 0$ or $\theta = \pi$). Note that if a geodesic is initially tangent to this submanifold it remains tangent to it. The maximally analytically extended Penrose diagram of this submaifold is given in figure 1.3

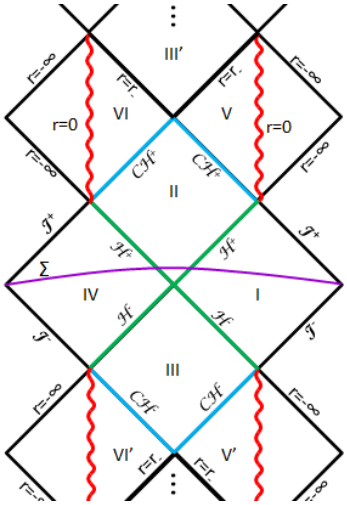


Fig. 1.3 Penrose diagram for maximal analytic extension of Kerr. The green lines are the future/past event horizon $\mathscr{H}^\pm$, the blue lines are the Cauchy horizons $\mathscr{C}\mathscr{H}^\pm$, the red squiggles are the singularity at $r = 0$, $\theta = \pi/2$; this has the structure of a ring so the solution can be extended through this to a new asymptotically flat region that goes to $r = -\infty$.

There are some crucial differences between Kerr and Reissner-Nordström. For example, the Kerr spacetimes possess an *ergoregion*: a region where the Killing vector field $\partial/\partial t$ that is timelike near infinity (generating time translation symmetries) becomes *spacelike* outside of the black hole. This can have some important implications, such as superradiance which is mentioned in section 1.2.

1.1.4 Reissner-Nordström-de Sitter

Our final example of a black hole is the Reissner-Nordström-de Sitter spacetime. This has a different structure at infinity, being asymptotically de Sitter as opposed to asymptotically flat.

It is a solution to the Einstein-Maxwell system with a positive cosmological constant $\Lambda > 0$:

$$G_{ab} + \Lambda g_{ab} = 8\pi T_{ab}, \tag{1.7}$$

and T_{ab} is the energy-momentum tensor for an electromagnetic field, the same as for Reissner-Nordström. The metric is

$$ds^2 = -f dt^2 + f^{-1} dr^2 + r^2 d\Omega^2, \quad \text{where} \quad f = 1 - \frac{2M}{r} + \frac{e^2}{r^2} - \frac{\Lambda}{3} r^2. \tag{1.8}$$

$f(r)$ has three positive roots $0 < r_- \leq r_+ \leq r_c$; here we describe non-extremal black holes for which the inequalites are strict. The maximal analytic extension is given in fig. 1.4.

Similarly to Reissner-Nordström, the null hypersurfaces $r = r_-$ and $r = r_+$ are the Cauchy horizon and event horizon respectively. Region I on the diagram is the exterior of the black hole. In contrast to the previous examples, the outer boundary of this region is not null infinity $\mathscr{I}^+$ but the null hypersurfaces at $r = r_c$, which are the future and past *cosmological horizons*. Observers can travel in an outwards direction only across the future cosmological horizon; nothing can enter region I by crossing it. The region $r > r_c$ goes all the way out to infinity, which is spacelike for asymptotically de Sitter spacetimes.

The Penrose diagram in fig. 1.4 again shows the maximal analytically extended spacetime, and the difference at infinity is evident. Note that the importance of the asymptotics comes not only from the structure of infinity itself but from having a cosmological horizon instead of null infinity as part of the boundary of region I.

1.2 Stability problem in GR

Since GR is a physical theory it is important that there are solutions to the Einstein equation which are of astrophysical relevance. If a particular spacetime requires fine-tuned initial

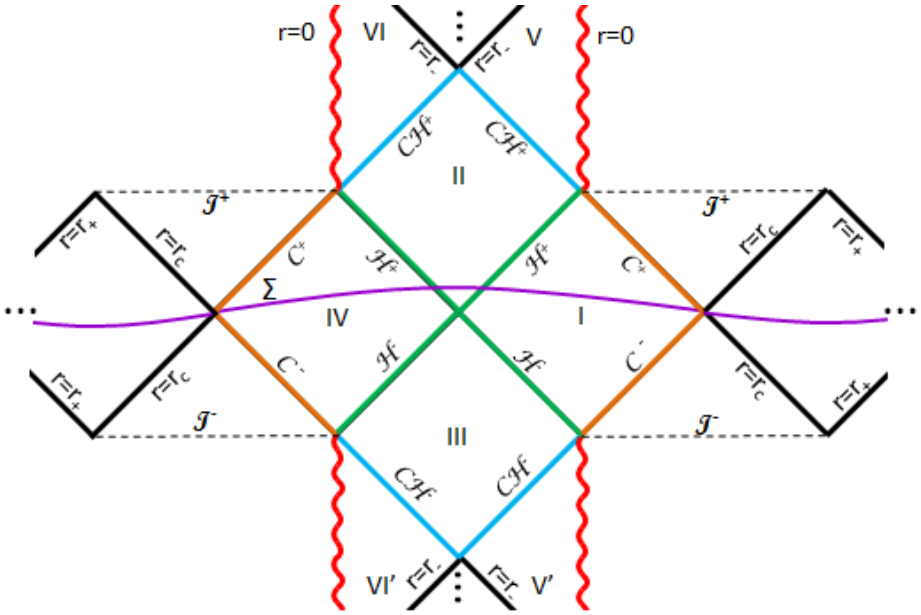


Fig. 1.4 Maximal analytic extension of Reissner-Nordström-de Sitter. The green lines $\mathscr{H}^{\pm}$ are the past/future event horizons, the vertical red squiggles are the singularity at $r = 0$, the purple line Σ is a Cauchy surface for regions $I - IV$ and the blue lines $CH^{\pm}$ are the past/future Cauchy horizons at $r = r_{-}$. The orange lines $C^{\pm}$ are the past/future cosmological horizons at $r = r_c$, and infinity $\mathscr{I}^{\pm}$ is now the spacelike dashed lines.

conditions and is unstable to perturbations it is not physical because the precise configuration of initial conditions required is unlikely to occur. In studying the stability problem we address this issue by attempting to determine whether or not a spacetime is non-linearly stable to small perturbations of the initial data. Note that by non-linear we mean that this is in the context of the full Einstein equations, including backreaction and interactions of perturbations.

The full non-linear stability of a spacetime is very hard to establish. Indeed, the original proof of the global non-linear stability of 4d Minkowski space [15] takes over 500 pages, so it is unsurprising that the non-linear stability of more complicated asymptotically flat 4d spacetimes, such as Kerr, are still open problems (although see [16–21] for recent progress in Kerr).

Before looking at non-linear stability it is important to study the rate of decay of linear perturbations (ignoring the backreaction of perturbations on the metric). If the solutions to the linear problem decay too slowly it is unlikely that the full problem will be stable when non-linear effects such as backreaction are included. For example, linear perturbations in Anti-de Sitter do not decay, leading to the expectation that AdS is unstable [22]; this instability was found in subsequent numerical investigation [23] and has been proven for the Einstein-massless Vlasov system in spherical symmetry [24].

The equations for the linearized gravitational perturbations are still difficult to solve, so it is common to first study an even simpler version, the scalar wave equation on a fixed background:

$$\Box_g \Psi = 0. \tag{1.9}$$

The first step in the stability problem is usually to study this proxy problem and find the rate of decay of solutions to the wave equation; this can either be viewed as a source for the perturbations or as the equation satisfied by a certain type of perturbation. In some spacetimes, such as the supersymmetric microstate geometries of chapters 2 and 3, there are decoupled linear perturbations that themselves satisfy the scalar wave equation.

It is not known in general what the slowest rate of decay is that allows for non-linear stability. In 4d Minkowski space the decay rate for the linear wave equation is t^{-1}. Although 4d Minkowski is non-linearly stable, this is not just because the rate is fast enough. For generic equations t^{-1} decay leads to solutions that blow up in finite time [25], but there are certain conditions on the non-linearities present in the Einstein equations that prevent this. A particular example is the “null condition" [26]: physically this prevents wave-packets that are travelling in the same null direction from interacting, so although they may be close to each other for a long time they do not form a singularity via interactions. This condition may be weakened to the "weak null condition" [27] for which global solutions still exist [28]; this has been used to prove the global non-linear stability of 4d Minkowski space using different methods to the original in [27] and more recently [29].

Since even t^{-1} decay can be problematic, non-linear stability is even less likely for a slower decay rate. There are various different factors that influence this rate. For the purposes of this thesis the most important of these is *trapping*: null geodesics that remain in a finite region of space for all time (*i.e.* are trapped). This is a common phenomenon; for example, in Schwarzschild there are trapped null geodesics at the photon sphere, and indeed for *any* spacetime with a photon sphere. It is possible for the energy of solutions to the wave equation to become concentrated near the trapped null geodesics for an arbitrarily long time, providing an obstruction to local energy decay estimates without loss of derivative [30, 31]. This has to be included somehow in the decay estimates. The usual local energy of these solutions involving up to first derivatives is constant for an arbitrarily long time; this obviously does not decay, but it is possible to obtain a decay rate for the energy when it is instead compared to an energy integral that includes up to second order derivatives.

In Schwarzschild the trapping at the photon sphere is *unstable*: if the trapped geodesic is perturbed it will either fall into the black hole or escape to infinity. In some cases, such as Kerr-AdS [32], ultracompact neutron stars [33] and black rings and black strings in higher dimensions [34], there are *stably* trapped null geodesics. This is significantly worse for

decay estimates, resulting in late time decay that goes as inverse powers of $\log t$. This is much slower than the power-law decay in asymptotically flat spacetimes with only unstable trapping, such as Schwarzschild.

The asymptotics of a spacetime also influence the rate of decay. It has already been mentioned that in AdS spacetime linear perturbations do not decay, in contrast to flat Minkowski spacetime where there is t^{-1} decay. Moreover, it has been proven that the rate of decay of solutions to the Klein-Gordon equation in Kerr-AdS is logarithmic [32] while solutions to the wave equation in asymptotically flat black holes exhibit power-law decay [18]. On the other hand, in de-Sitter spacetimes solutions decay exponentially to a constant [35–40]. Physically this is due to the behaviour at infinity: in Minkowski spacetime solutions do just disperse while in AdS spacetime there is a timelike boundary at infinity which means that waves are reflected back in to the spacetime. In de Sitter there is a cosmological horizon (similar to an event horizon but observers can only cross it going outwards) as opposed to the usual infinity. Solutions fall out across this and disperse even faster than in Minkowski. Decay in asmyptotically AdS spacetimes is expected to be much slower than in asymptotically flat spacetimes, which in turn is slower than in asymptotically de Sitter spacetimes.

There are various other factors influencing the rate of decay that are not covered in this work but are mentioned here for completeness. The first of these is superradiance. This is the idea that there are solutions of the wave equation which are amplified when they scatter off a black hole, extracting energy from it (see for example [41–44]). This happens when the Killing vector field that is timelike at infinity, used to define a conserved energy, becomes spacelike in a region known as the ergoregion. When it becomes spacelike the energy of the solution is no longer positive definite, meaning that the fact that it is conserved cannot be used to prove boundedness. In terms of particles, this is the Penrose process [45], a mechanism for a particle to extract energy from the black hole. In chapter 2 we see the implications for the rate of decay of an *evanescent* ergosurface, a submanifold on which the Killing vector field is null but timelike either side (as opposed to timelike on one side and spacelike on the other).

It is also worth mentioning that the number of dimensions has an effect on the rate of decay. In higher dimensions solutions usually decay more quickly; it is only in 4d (3+1) Minkowski space that the decay of linear perturbations is t^{-1}, which is the borderline case that makes it particularly hard to get to non-linear stability. In higher dimensions the analogous proof is much shorter [46].

The above discussion is about *linear* perturbations, ignoring non-linear interactions and backreaction on the geometry. Due to the difficulty of the problem, there are few instances of proofs of global non-linear stability, but these include 4d Minkowski space [15] and, more recently, slowly rotating Kerr-de Sitter [39]. The non-linear stability of the Kerr spacetime,

describing the black holes in our universe, remains an open problem, although there has been progress on the linear version [16–21].

1.2.1 Quasinormal modes

For several chapters of this thesis we will be concerned with a particular type of solution to the wave equation: *quasinormal modes*. These are mode solutions of the form

$$\Psi(t,x^i) = e^{-i\omega t}\tilde{\psi}(x^i) \tag{1.10}$$

that are ingoing at the event horizon and outgoing at infinity. These boundary conditions can be generalised to non-black hole spacetimes where there is no event horizon simply by demanding that the solution is outgoing at infinity and smooth everywhere. They have a definite complex frequency ω that is the *quasinormal frequency*; there is a discrete set of possible values for ω in the complex plane (see e.g. [47]). The imaginary part of the frequency governs the rate of decay (when it is negative) over time of these solutions. Mathematically they can be defined as the poles of the Green's function for the equation obtained after Fourier transforming the wave equation.

In a black hole spacetime, an external perturbation can excite the quasinormal modes which are the damped oscillations of the black hole's response [48]. The quasinormal frequencies can be used to identify the black hole, since they depend on the black hole parameters such as the mass, charge and angular momentum (see, for example, [49] for a review).

Spacetimes can have several families of quasinormal modes associated with different features of the geometry. For example, there may be modes associated with the near-horizon geometry or the asymptotics of the spacetime; these will be mentioned in the context of asymptotically de Sitter black holes in chapter 4. In this thesis we will predominantly be interested in *photon sphere* modes: these are quasinormal modes that are associated with null geodesics that stay at constant radius to form the photon sphere. This link can be explained by the geometric optics approximation.

1.2.2 Geometric optics approximation

Although it is significantly easier to study the wave equation on a fixed background than the full problem of non-linear stability, it is still difficult to determine the behaviour of solutions to this second order partial differential equation. The *geometric optics* approximation is an

effective way to find approximate high frequency solutions to the wave equation by looking at geodesics instead.

The idea is that it is possible to find solutions to the wave equation that are localised near null geodesics [30]. Using Gaussian beams, which are approximate solutions that decay exponentially away from the null geodesic, it can be proved that a solution can be localised near a null geodesic for an arbitrarily long time, and that the energy of the solution is close to the conserved energy of the null geodesic [31].

The link between geodesics and quasinormal mode solutions specifically was first mentioned in [48] for Schwarzschild and Kerr black holes, which have an unstable circular photon orbit. If we perturb this circular orbit slightly the resulting null geodesics either fall in to the black hole or escape to infinity; these are exactly the same boundary conditions as for quasinormal modes. This perturbation corresponds to a superposition of high frequency (eikonal limit) eigenmodes via the geometric optics approximation. Properties of these modes can be determined by the geodesics: the real part of the frequency is given by the orbital frequency of the circular geodesic, while the imaginary part of the frequency, corresponding to the decay of the mode, is related to the density of the geodesics (as they spread out there is less energy concentrated at the circular null geodesic).

To make use of the link between quasinormal modes and null geodesics, we first need to be able to find them; we can do this using the Hamilton-Jacobi formulation.

1.2.2.1 Hamilton-Jacobi formulation

The Hamilton-Jacobi equation for a null geodesic [50] is

$$g^{\mu\nu}\frac{\partial S}{\partial x^{\mu}}\frac{\partial S}{\partial x^{\nu}} = 0, \tag{1.11}$$

where S is the principal function. To recover the motion of massless particles following null geodesics, according to Hamilton-Jacobi theory,

$$\frac{\partial S}{\partial x^{\mu}} \equiv p_{\mu} \quad \text{and} \quad p^{\mu} = \frac{dx^{\mu}}{d\tau} \tag{1.12}$$

where τ is an affine parameter along the null geodesic.

1.2.2.2 Eikonal limit

As a brief explanation of the geometric optics approximation (for more details see [51]), consider a solution to the wave equation of the form $\Psi = Ae^{i\mu\zeta}$ in the eikonal limit: $\mu \gg 1$ is much larger than any other length scales. A is the amplitude and ζ is the phase of the

solution. If we substitute this into the wave equation and expand in powers of μ, at leading order we have

$$g^{\mu\nu}\partial_\mu\zeta\partial_\nu\zeta = 0,$$

while next to leading order terms give a transport equation for A. Now we can compare this to the Hamilton-Jacobi equation for null geodesics.

At leading order the Hamilton-Jacobi equation (1.11) for S is identical to the equation for ζ, so we expect S and ζ to be proportional at leading order. At next to leading order we can find an expression for A, which can be shown to be related to how fast neighbouring geodesics spread out.

To see more explicitly how this works in practice, consider a Kerr black hole. There is sufficient symmetry that the wave equation separates and we can look for mode solutions of the form

$$\psi = e^{-i\omega t+im\phi}\Phi_r(r)\Phi_\theta(\theta). \tag{1.13}$$

The angular equation gives spheroidal harmonics labelled by an integer ℓ with $|m| \leq \ell$ (similar to the spherical harmonics). In the eikonal limit, $\ell \gg 1$, we can compare the quasinormal frequencies obtained via the geometric optics approximation with those calculated using the WKB approximation. The WKB method can be used [52] to construct families of quasinormal modes with frequency

$$\omega = \omega_R + i\omega_I.$$

The geometric interpretation of these frequencies was also studied in [52], where it was found that, as expected from the geometric optics approximation, they are determined by the unstably trapped null geodesics on the photon sphere in the following way.

First of all, note that if E and L are respectively the conserved energy and momenta of a trapped null geodesic on the photon sphere with momentum p, then from (1.12), $E = -(\partial/\partial t)\cdot p = -p_t$ and similarly $L = p_\phi$. Using eq. (1.12) and integrating the t and ϕ components, the principal function for such geodesics can be written $S = -Et + L\phi + R(r) + \Theta(\theta)$; the ratio L/E is determined by the requirement that these are circular null geodesics,

Then the relation between the null geodesics and the quasinormal modes found in [52] is that

$$\omega_R/m \approx E/L;$$

this agrees with the expectation from the geometric optics approximation that $(-i\omega t + im\phi)$ is proportional to S at leading order. It was also found that ω_I is indeed determined by the Lyapunov exponent, which describes how quickly the null geodesics near the photon sphere

spread out, and is subleading since it is $O(1)$ while ω_R is $O(\ell)$. More precisely, for photon sphere quasinormal modes, which are associated with circular null geodesics in the equatorial plane and have $\ell = |m| \gg 1$, the quasinormal frequencies are given by [53, 48, 54–59, 52]

$$\omega \approx m\Omega_c - i\left(n+\frac{1}{2}\right)\lambda\,, \qquad n = 0,1,2,\cdots, \tag{1.14}$$

where n is the radial overtone, $\Omega_c = E/L$ is the Keplerian frequency of the orbit and λ is the Lyapunov exponent.

We will study the stability problem using these approaches in chapter 2 by looking at linear perturbations of supersymmetric microstate geometries (these are introduced in the relevant chapter). Although the stability problem has been studied extensively for other spacetimes, nothing was known about whether or not these particular geometries were stable until the work presented in this thesis (see [60]).

In that chapter we give a heuristic argument as to why we expect certain supersymmetric microstate geometries that have stably trapped zero-energy null geodesics to be non-linearly unstable. We provide evidence in support of this by studying the linear problem, the wave equation on these backgrounds. Motivated by the geometric optics approximation, we find quasinormal mode solutions that decay particularly slowly; the rate of decay we find is actually slower than is known for any other spacetime. Indeed, it was proved in [61] that in a general class of asymptotically flat spacetimes of dimension $d+1$, $d \geq 3$, the rate of decay of solutions to the wave equation is at least as fast as logarithmic. We find that the decay rate in the supersymmetric microstate geometries, which do not belong to this class of spacetime[1], is slower than logarithmic; such slow decay is likely to lead to a non-linear instability for these geometries.

Since the slow decay is entirely due to the zero-energy stably trapped null geodesics, we expect that solutions of the wave equation decay slowly in any spacetime that has such geodesics. Thus our results have implications for stability more generally; we give examples of such spacetimes in chapter 2. Note that we only study the linear problem (apart from the heuristic arguments) and the full non-linear problem has yet to be addressed, but since we gain significant insight we see how useful it can be to study the much more tractable linear problem.

Geodesics are very important in this analysis, so we study them in detail in chapter 3. Stable trapping is a local phenomenon, but it has important consequences for the decay of

[1] In 5d, where the microstate geometries are asymptotically flat, they do not possess a globally timelike Killing vector field, nor do they have an event horizon; at least one of these is required for the assumptions of [61] to be satisfied.

solutions to the wave equation. It would therefore be interesting to try to obtain a general result, independent of any other features in a spacetime, that set a lower bound on the rate of decay when there is stable trapping present.

In this section we have discussed one use of wave equations on curved spacetimes in GR; in the next we go on to discuss another unresolved problem in GR where such equations are useful.

1.3 Strong Cosmic Censorship

Predictability is an important part of any physical theory. For example, in classical theories such as Newtonian physics or Maxwell theory, if we know the initial data (particle positions and velocities or electromagnetic field) then, in principle, the equations of motion tell us exactly where the particles are or what the electromagnetic field configuration is for all time. Even in quantum mechanics, if we know the wave function initially then the Schrödinger equation determines the wave function in the future. It is only in GR that we seem to lose predictability when a spacetime has a Cauchy horizon.

As mentioned in section 1.1.2, the Cauchy horizon is the boundary of the region of spacetime where the solution to the Cauchy problem is unique; such a region, known as the maximal Cauchy development, always exists in GR [62]. The Cauchy problem is the initial value problem, for which we specify initial data on some spacelike hypersurface Σ and evolve it according to the equations of motion. For example, Reissner-Nordström (or Kerr) is the spacetime obtained by specifying the appropriate initial data on Σ (see figs. 1.2 and 1.3) and evolving it up to the Cauchy horizon, which is the boundary of the domain of dependence of Σ; this gives regions $I - IV$ of fig. 1.2 (or 1.3).

There are infinitely many ways to extend the spacetime beyond the Cauchy horizon that are all smooth and satisfy the equations of motion. One particular way is to analytically extend the spacetime across the Cauchy horizon, which is the extension we have chosen to show in the Penrose diagrams in figs. 1.2, 1.3 and 1.4. However, analyticity is not a good physical assumption, since it means that if we know the solution at a particular point then we know it everywhere. The loss of predictability comes about because we do not not know which extension to choose.

1.3.1 Blueshift effect

Fortunately a possible way out of this dilemma was first noticed by Penrose [63]: there is a mechanism known as the *blueshift effect* which suggests that the Cauchy horizon is unstable.

If it is unstable then perturbations to the initial data would result in a spacetime that does not have a Cauchy horizon, possibly having a singularity instead, and is therefore inextendible as desired. This is of physical relevance because it is unlikely that initial conditions will be so fine-tuned as to give the exact initial data required for each particular spacetime with a Cauchy horizon, such as Reissner-Nordström or Kerr (-de Sitter).

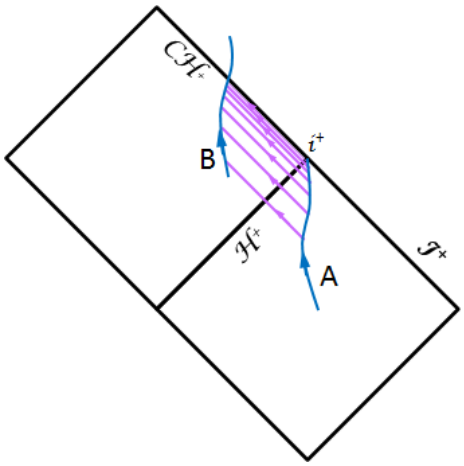


Fig. 1.5 Blueshift effect: observer A outside the black hole sends signals at regular intervals to B, who receives infinitely many signals in finite proper time.

The blueshift effect is shown in fig. 1.5. An observer A in the exterior sends signals to an observer B inside the black hole at regular intervals of, say, 1 second of A's proper time. If, hypothetically, A lives forever, she can send infinitely many signals to B. But B receives all of these signals before he crosses the Cauchy horizon in finite proper time, so he is receiving infinitely many signals in a finite time. The signals from Region I thus undergo an infinite blueshift at the Cauchy horizon in Region II. This implies that a small perturbation in Region I will have a huge energy (measured by B) at the Cauchy horizon, so the gravitational backreaction will be very large near the Cauchy horizon. This suggests that the Cauchy horizon is unstable.

The blueshift effect lead to the strong cosmic censorship conjecture of Penrose, which attempts to restore determinism to GR.

1.3.2 Strong cosmic censorship conjecture

The strong cosmic censorship (SCC) conjecture, originally suggested by Penrose in [64], essentially says that all physical spacetimes should be inextendible, *i.e.* any Cauchy horizon should be unstable [65–68]. The mathematical formulation that we shall use, first given in [69] is:

Conjecture 1.3.1 (Strong cosmic censorship)**.** *Let* (Σ, h_{ab}, K_{ab}) *be a geodesically complete, asymptotically flat initial data set for the vacuum Einstein equations. Then, generically, the maximal Cauchy development of this initial data is inextendible.*

If it is correct, we would recover determinism for GR. This conjecture has yet to be proven to be either true or false, but there has been much interest in the question; see [70] for a detailed discussion. In chapter 4 we will give an account of the recent work related to it in the context of asymptotically de Sitter black hole spacetimes and investigate whether or not it holds for Kerr-de Sitter black holes. In this section we will discuss the conjecture more generally, particularly the progress that has been made for asymptotically flat black holes.

One of the key words in the SCC conjecture is *generic*. This is the requirement that the phenomenon (in this case (in)extendibility) is physical. For example, if a particular spacetime has a Cauchy horizon then it is extendible. However, if the Cauchy horizon is unstable to perturbations of the initial data then it is not generic; it requires fine-tuned initial conditions to exist and as such is unlikely to form physically.

Inextendibility

It is important to explain more precisely what is meant by *inextendible*, as there are several possibilities. The simplest formulation of SCC is with C^0-inextendibility; the first statement of SCC in the form of Conjecture 1.3.1 in [69] used this version. The C^0-formulation states that it should be impossible to extend the spacetime across the Cauchy horizon with a continuous metric. However, there are various indications (for charged black holes) that it is possible to extend either the spacetime metric or other fields continuously across the Cauchy horizon [71–74]. Moreover, it has been proven rigorously that this formulation is false for the vacuum Einstein equations in a neighbourhood of Kerr [13]; since we would expect SCC to apply to Kerr as these are the black holes present in our universe, this would not be a good version of the conjecture.

The C^2-formulation is that the spacetime should be inextendible with a C^2-metric (*i.e.* the metric and its first and second derivatives are continuous). The motivation for this formulation is the mass inflation scenario of Poisson and Israel [68], which says that curvature invariants generally diverge at the Cauchy horizon so the extended spacetime cannot be C^2. The divergence of curvature has been proven for spherically-symmetric solutions of Einstein-Maxwell theory with a massless scalar field and zero cosmological constant [75]. However, this does *not* necessarily imply that an observer cannot cross the Cauchy horizon; indeed Ori [72] showed that the total tidal deformation an observer experiences remains finite despite the blow-up of curvature invariants.

The most physical interpretation of 'inextendible' is that an observer should not be able to cross the Cauchy horizon. Whether or not an observer can cross the Cauchy horizon depends on what the observer is made of, the equations of motion for that matter and the Einstein equations; we want to say it is impossible to extend the spacetime as a solution of the equations of motion. This depends on the minimal regularity allowed for solutions. The equations involve up to second derivatives of the metric, so one might expect that C^2 is the minimal regularity. However, it is also possible to have *weak solutions*. For a second-order quasilinear partial differential equation, if it is multiplied by a smooth, compactly supported test function and integrated by parts to get rid of the second order derivatives, a weak solution satisfies the resulting equation for any arbitrary test function. These are physical solutions to the equations of motion; for example, shocks in incompressible fluids are certainly physical since they can be observed, but they only exist as weak solutions since they are not in C^2. For the Einstein equations, a weak solution is one with locally square integrable Christoffel symbols in some coordinate chart (see beginning of [76] for a discussion).

This led to Christodoulou's formulation of SCC: the maximal development should be inextendible as a spacetime with locally square integrable Christoffel symbols [76].

Linear problem

To find out whether any version of SCC is respected or not in a certain spacetime with a Cauchy horizon it is necessary to determine whether the Cauchy horizon is stable (in the appropriate sense) to perturbations. Similar to the stability problem in GR, this is a non-linear problem that is very complicated; to simplify things the linear problem is usually studied first. Again, this reduces to the proxy problem of looking at the behaviour of solutions to the wave equation on a fixed background

$$\Box_g \psi = 0 \tag{1.15}$$

near the Cauchy horizon.

Indeed, prior to [13], the evidence against the C^0-formulation came from the study of the wave equation on sub-extremal Reissner-Nordström and Kerr backgrounds, and is roughly as follows. If ψ is a solution of eq. (1.15) arising from suitably regular Cauchy data on an initial spacelike hypersurface Σ in either Kerr or Kerr-Newman (this includes Reissner-Nordström), then ψ decays inverse polynomially to 0 on the event horizon [18, 77]. This inverse polynomial decay propagates to a spacelike hypersurface in the interior of the black hole, and it turns out that in fact ψ is bounded in the entire spacetime and can be continuously

extended across the Cauchy horizon [74, 78]. This shows that the linear version of the C^0-formulation of SCC does not hold.

For the linear problem, Christodoulou's formulation of SCC reduces to the requirement that $\psi \notin H^1_{loc}$, where the Sobolev space H^1_{loc} contains functions that are locally square integrable and for which the gradient is also locally square integrable. $\psi \notin H^1_{loc}$ is the same as the statement that the local energy of ψ diverges at the Cauchy horizon.

In asymptotically flat spacetimes, the linear version of Christodoulou's formulation holds in both Reissner-Nordström [75] and Kerr [79]. In both cases, ψ is bounded at the Cauchy horizon but the transverse derivatives of ψ blow up, resulting in infinite energy of the solution at the Cauchy horizon [80]. This is because the behaviour of ψ at the Cauchy horizon depends on what happens at the event horizon. ψ decays inverse polynomially at the event horizon, giving the well-known power-law tails. At the Cauchy horizon this decay is countered by growth coming from an exponential in time factor due to the blueshift effect; since this factor is an exponential it is much more important than the inverse polynomial decay and this leads to blow-up of some derivatives at the Cauchy horizon.

Non-linear results

The non-linear version of Christodoulou's formulation has yet to be proven, although it is generally expected to be true for both the Einstein-Maxwell equations in a neighbourhood of Reissner-Nordström and the vacuum Einstein equations in a neighborhood of Kerr. However, the non-linear version of the C^0 formulation has been proven to be false in both of these cases: for the Einstein-Maxwell equations in spherical symmetry the metric extends continuously [68, 81], whilst for the vacuum equations in a neighbourhood of Kerr, the Cauchy horizon is still a null boundary (after perturbing the initial data) across which the metric extends continuously [13].

Asymptotically de Sitter spacetimes

As long ago as 1990, Mellor and Moss [82] found evidence that the C^2 formulation of SCC was violated in a certain Reissner-Nordström-de Sitter spacetime that was close to extremality. However, several years later Brady, Moss and Myers [83] wrote a letter to say that the previous paper was in fact incorrect and that SCC was not in fact violated in this case. The state of SCC in asymptotically de Sitter spacetimes was thus left in this state of confusion until as recently as 2017.

In chapter 4 we study SCC in asymptotically de Sitter spacetimes in detail. We will emphasise the key differences between such spacetimes and the asymptotically flat black

holes mentioned above; in particular we will see why a cosmological horizon affects whether or not SCC holds. We will predominantly be concerned with the linear version of SCC for $\Lambda > 0$. As we will discuss, late-time behaviour of solutions to the wave equation in asymptotically de-Sitter spacetimes is determined by the rate of decay of quasinormal modes. Therefore much of chapter 4 will be about quasinormal modes in Reissner-Nordström- and Kerr-de Sitter.

We will start off by reviewing recent work done in relation to SCC in Reissner-Nordström-de Sitter spacetimes, which provides evidence that SCC does *not* always hold (at least when starting from smooth initial data), which is worrying.

The bulk of that chapter is about SCC in the more physical rotating Kerr-de Sitter spacetimes. We will study the linear version of SCC in this case by using both the geometric optics approximation and numerics to find the photon sphere quasinormal modes for the scalar wave equation; it turns out that these decay relatively slowly. In addition, we investigate the linearised gravitational perturbations numerically. Reassuringly, it turns out that the rate of decay for both the scalar and gravitational perturbations is sufficiently slow to provide evidence that Chistodoulou's formulation of SCC is actually respected in Kerr-de Sitter spacetimes.

It is interesting to see the effect the asymptotics can have on the linear version of SCC. Since we are interested in the behaviour of fields at the inner horizon, it might seem surprising that the asymptotics play such an important role. The difference in asymptotically flat and asymptotically de Sitter black holes comes about because the behaviour of the fields at the Cauchy horizon depends on decay in the exterior; this is heavily influenced by the asymptotics of the spacetime, being exponential in the de Sitter case but power-law in the flat case. Moreover, in asymptotically de Sitter black holes, we can simplify things by using quasinormal modes to find the decay rate because it is exponential; this is not possible in the asymptotically flat black holes. In addition, there seems to be a qualitative difference between Einstein-Maxwell theory and pure gravity, since SCC appears to be violated in the former but respected in the latter.

The results we obtain are all from the geometric optics approximation or numerics. It would be interesting to see if the behaviour of scalar waves at the Cauchy horizon could be reproduced more rigorously. This could perhaps be done by obtaining a better mathematical description of the quasinormal modes in the exterior using the fact that the wave equation separates and the effective potentials in the one-dimensional equations of motion have certain properties. These solutions could then be continued in the interior up to the Cauchy horizon, perhaps by using the WKB approximation and quantifying the error involved.

1.4 Uniqueness of solutions

The Cauchy problem, *i.e.* the correspondence between solutions of partial differential equations and their initial data, for GR was first investigated by Choquet-Bruhat in [84]. In this work, she showed that in the harmonic gauge (the freedom to choose a gauge comes from the diffeomorphism invariance of the Einstein equations), in which coordinates satisfy the wave equation:

$$\Box_g x^\mu = 0,$$

the Einstein equations reduce to a system of quasilinear wave equations of the form

$$g^{\alpha\beta}\partial_\alpha\partial_\beta g_{\mu\nu} = f(g,\partial g)_{\mu\nu} \tag{1.16}$$

together with some constraints. The initial data (Σ, h_{ab}, K_{ab}) for the Einstein equations consists of a 3-dimensional Riemannian manifold Σ that corresponds to a spacelike hypersurface in the spacetime, the pull-back of the spacetime metric to Σ, h_{ab}, and a symmetric tensor K_{ab}, which is the extrinsic curvature of Σ. It turns out that the gauge condition propagates, and then *local* well-posedness follows from general results on systems of second order quasilinear partial differential equations. Note that an initial value problem is well-posed if there exists a unique solution that depends continuously on the initial data; by local we mean that this holds in some neighbourhood of the initial surface.

This local result was extended to the problem of *global* existence and uniqueness for the vacuum Einstein equations by Choquet-Bruhat and Geroch in [62] (also much later by Sbierski [85], without appealing to Zorn's Lemma) to obtain the following:

Theorem 1.4.1. *Let (Σ, h_{ab}, K_{ab}) be initial data satisfying the Hamiltonian and momentum constraints:*

$$R' - K^{ab}K_{ab} + K^2 = 0 \quad \textit{and} \quad D_b K^b_a - D_a K = 0,$$

where R' is the Ricci scalar of h_{ab}, $K = h^{ab}K_{ab}$ and D_a is the Levi-Civita connection associated to h_{ab}. Then there exists a unique *spacetime, (M, g_{ab}), (up to diffeomorphism invariance) called the* maximal Cauchy development *of (Σ, h_{ab}, K_{ab}) such that: (i) (M, g_{ab}) satisfies the vacuum Einstein equation; (ii) (M, g_{ab}) is globally hyperbolic with Cauchy surface Σ; (iii) the induced metric and extrinsic curvature of Σ are h_{ab} and K_{ab} respectively, and finally, (iv) any other spacetime satisfying $(i)-(iii)$ is isometric to a proper subset of (M, g_{ab}).*

Note that by 'globally hyperbolic with Cauchy surface Σ', we mean that $M = D(\Sigma)$, where $D(\Sigma)$ is the domain of dependence of Σ. In other words, the spacetime only depends on the data on Σ.

Theorem 1.4.1 is of great significance. It means that in GR it makes sense to investigate the physics of a spacetime because, given initial data, the spacetime itself is entirely determined. Of course, problems with predictability may arise if the maximal Cauchy development has a boundary over which we can smoothly extend (see the discussion of SCC in section 1.3), but we do at least have uniqueness before the Cauchy horizon; if we find one maximal Cauchy development it has to be the right one.

Moving away from GR to more general quasilinear wave equations on a fixed background, suppose that we have initial data on some hypersurface Σ that we can evolve, as for GR, to obtain a globally hyperbolic development that is maximal. Now suppose that, in contrast to GR, this development is not unique: there is another maximal globally hyperbolic development that disagrees with the first in some region. How do we know which one to choose? Since they are both globally hyperbolic developments, determined entirely from the same initial data, there is no way to know if either is preferred. It is therefore impossible to attempt to discuss the physics of the theory since we can't predict it from the initial data.

We will study this particular issue with predictability in chapter 5. We do this in the context of subluminal and superluminal wave equations, which allow propagation of signals slower and faster than light respectively. The reason to study such equations is that it is often suggested that superluminal wave equations are much worse for predictability, possibly admitting 'time machine' solutions. However, despite various heuristic arguments, a rigorous example of this has yet to be found. We investigate the Born-Infeld scalar in chapter 5, which has both a subluminal and superluminal formulation. We find that the subluminal wave equation can have infinitely many different maximal globally hyperbolic developments, violating predictability badly in the way we have just discussed; this also applies to more general quasilinear wave equations. On the other hand, it turns out that superluminal wave equations have a unique maximal globally hyperbolic development, surprisingly implying that from this point of view they are actually more predictable than subluminal wave equations.

Chapter 2

Instability of supersymmetric microstate geometries

This chapter is work done in collaboration with Harvey Reall and Jorge Santos, which has been published in [60].

2.1 Introduction

Type IIB supergravity admits supersymmetric "microstate geometry" solutions [86–95]. These are asymptotically flat, geodesically complete, stationary solutions without horizons. Near infinity, they approach the product of 5 dimensional flat spacetime with 5 compact dimensions. Some of these solutions can be dimensionally reduced to give smooth solutions of 5d supergravity. In 5d, the stationary Killing vector field V is timelike everywhere except on a certain timelike hypersurface, where is it is null. This surface has infinite redshift relative to infinity, and has been called an "evanescent ergosurface" [96].

A natural question is whether these spacetimes are classically stable. This has been investigated for non-supersymmetric microstate geometries, which can have a genuine ergoregion, where V becomes spacelike [97]. Such geometries have been shown to be unstable [98]: linear perturbations localized in the ergoregion can have negative energy and there exist modes which grow exponentially in time. In the supersymmetric case, linear perturbations have non-negative energy, which excludes exponential growth so one might expect stability.

A simple argument suggests that supersymmetric microstate geometries actually have a nonlinear instability. The argument is based on the existence of the evanescent ergosurface. As we shall explain, on an evanescent ergosurface, V is tangent to affinely parameterized

null geodesics with zero energy. These geodesics are at rest relative to infinity so they are resisting the frame-dragging effect caused by the rotation of the geometry. Hence they can be regarded as carrying angular momentum opposed to that of the background spacetime. These geodesics are "trapped" in the sense that they remain in a finite region of space, i.e., they do not disperse. As mentioned in section 1.2 of the Introduction, trapping occurs in other situations, e.g., at the photon sphere of a Schwarzschild black hole. However, in the Schwarzschild case, the trapping is unstable: if one perturbs the geodesic then it will escape to infinity or fall into the black hole. At an evanescent ergosurface the trapping is *stable* because the geodesics sit at the bottom of a gravitational potential well.

Now consider perturbing the spacetime by adding an uncharged massive particle (or a tiny black hole) near to the evanescent ergosurface. If we neglect backreaction then the particle moves on a geodesic. However, if we couple it to supergravity fields then it will gradually radiate energy and angular momentum through its coupling to gravitational radiation (and other massless fields). Hence it will gradually lose energy and its trajectory will approach a geodesic which minimizes the energy. But these trajectories are precisely the zero-energy null geodesics tangent to V on the evanescent ergosurface. Hence the trajectory of our particle will approach one of these trapped null geodesics. It will have very small energy as measured at infinity. However, since the massive particle is now following an almost null trajectory, the energy measured by a *local* observer will be enormous. Hence its backreaction on the geometry will be large. This strongly suggests an instability.

What would be the endpoint of such an instability? The instability involves removing angular momentum from the microstate geometry via radiation. This will tend to shrink the evanescent ergosurface. An obvious candidate endpoint is an almost supersymmetric black hole with the same conserved charges as the microstate geometry, but different angular momenta. This could be a near-extremal BMPV black hole [99].

This heuristic argument for instability involves a massive particle. Is there also an instability involving only massless supergravity fields? Our argument relied on the fact that the particle can radiate, i.e., interactions are important. This suggests that a corresponding instability in supergravity will be a nonlinear effect, which makes demonstrating its existence difficult. But it is easy to see why the presence of an evanescent ergosurface makes nonlinear stability unlikely, as we will now explain.

As discussed in 1.2, proofs of nonlinear stability, e.g., the stability of Minkowski spacetime [15], involve first establishing that solutions of the *linearized* problem decay sufficiently rapidly. This decay occurs via dispersion to infinity (or across a black hole horizon).

Supersymmetric microstate geometries are asymptotically flat, so it is possible for linear perturbations to disperse to infinity. However, the presence of the evanescent ergosurface

implies that generic linear perturbations decay very slowly because of the stable trapping. To discuss this in more detail, we note first that there exist decoupled linear perturbations that behave like a massless scalar in these geometries [100]. Therefore we will consider the behaviour of a massless scalar field, i.e., the wave equation (this is better than just a simplification of the equations for the linearized gravitational perturbations in this case). Using geometric optics, one can construct low energy, spatially localized, solutions of the wave equation describing wavepackets propagating along the zero energy null geodesics [31]. These can decay by dispersion to infinity but, because of the stable trapping, this involves tunnelling through a potential barrier and so the decay will be very slow. This has been studied in detail for other examples of spacetimes with stable trapping, namely anti-de Sitter black holes [32] and "ultracompact" neutron stars (stars with a photon sphere) [33]. In both cases, it has been shown that the stable trapping implies that the late time decay is generically as an inverse power of $\log t$ where t labels a foliation by spacelike surfaces such that $\partial/\partial t$ is Killing. (This can be contrasted with the power-law decay of waves in asymptotically flat black hole spacetimes.)

This slow decay presents a serious problem for attempts to prove stability for a *nonlinear* equation. Given the difficulties already encountered when linear waves decay at a rate t^{-1}, slower rates appear particularly troubling. In the case where these rates are related to the phenomenon of stable trapping, the physical mechanism underlying the null condition also appears to be absent: waves can be localised along *different* null directions, but still interact for a long time. This appears particularly dangerous in the case where the stable trapping is "local", i.e., confined to a finite region of space, as in microstate geometries and ultracompact stars.[1]

For supersymmetric microstate geometries, the stable trapping appears worse than the other two examples just discussed (anti-de Sitter black holes [32] and "ultracompact" neutron stars [33]) because the associated null geodesics have zero energy. For the wave equation, the corresponding statement is that the energy degenerates on the evanescent ergosurface, so that smallness of the energy does not imply smallness of the gradient of the field there. This means that standard methods for establishing boundedness of solutions of the wave equation do not work. So even proving *linear* stability of the wave equation in these geometries is non-trivial. Even if linear stability can be established, we expect the decay of linear perturbations will be at least as slow as the examples of stable trapping just mentioned, which is far too slow for establishing nonlinear stability.

[1] It is conceivable that the stable trapping may be less of a problem for the example of AdS black holes because there the trapping occurs at infinity. Ref. [101] argues that such spacetimes will be nonlinearly stable.

In the discussion so far we have concentrated on microstate geometries from the 5d perspective. However, such geometries are often best viewed as solutions in 6 dimensions, with a compact Kaluza-Klein circle (indeed some geometries are smooth in 6d but not in 5d). We explain below how to define the evanescent ergosurface from the 6d perspective. We will also investigate the trapping in 6d. Surprisingly, we find that for any supersymmetric microstate geometry, there is a stably trapped null geodesic passing through *every* point of the 6d spacetime, i.e., not just points on the evanescent ergosurface. Away from the ergosurface, these correspond to BPS charged particle trajectories in 5d. In this paper, we will focus mainly on the stable trapping on the evanescent ergosurface.

To gain some understanding of the behaviour of geodesics and linear perturbations of microstate geometries, we will study in detail two classes of solutions. In section 2.3.1 we study the 3-charge microstate geometries of Ref. [91] and the maximally rotating 2-charge microstate geometries of Ref. [87]. These solutions are special because they have extra symmetries which enable the geodesic equation or wave equation to be separated and reduced to ODEs. We will show that there are families of quasinormal modes which are localized around the stably trapped zero energy null geodesics on the evanescent ergosurface $\mathscr{S}$, and which decay very slowly. We construct these modes using a matched asymptotic expansion valid for large "total angular momentum" quantum number $\ell \gg 1$, with the result that these modes have frequency

$$\omega \approx \omega_R - i\beta e^{-2\ell \log \ell} \tag{2.1}$$

where ω_R and $\beta > 0$ are constants that are independent of ℓ to leading order. There are also quasinormal modes localized around the stably trapped null geodesics away from $\mathscr{S}$, with $\mathrm{Im}\omega \sim -\exp(-\ell \log \ell)$. These results are for $\ell \gg 1$ but we have also constructed such quasinormal modes numerically, and find that they decay very slowly even at small ℓ.

We can compare this result with the behaviour of quasinormal modes for AdS black holes [102, 103] or ultracompact stars [104]. There are two important differences. First, in these examples ω_R is proportional to ℓ at large ℓ whereas in our case, ω_R does not scale with ℓ. This is closely related to the fact that the associated null geodesics have zero energy. Second, for AdS black holes or ultracompact stars, the imaginary part of the frequency of the most slowly decaying quasinormal modes is of the form $e^{-\gamma \ell}$ (for some $\gamma > 0$) whereas we have $e^{-2\ell \log \ell}$. Hence, in our case, the decay of quasinormal modes is slower than in these other examples of stable trapping. We will explain below why this behaviour of the quasinormal modes implies that generic perturbations decay slower than for AdS black holes or ultracompact stars, and therefore cannot exhibit power law decay. A rigorous result proving this slow decay is given in [105].

Our construction of the quasinormal modes exploits the special properties of these particular microstate geometries. However, since these modes are localized around the zero energy null geodesics, we expect that the slow decay of these quasinormal modes is a generic feature of spacetimes with an evanescent ergosurface, and hence our conclusion on the slow decay of generic perturbations should apply to any such spacetime.

Note that the slowest decaying modes are those with the largest angular frequency. This suggests that the nonlinear instability of such geometries will be a short-distance effect, perhaps involving the formation of tiny (uncharged) black holes, as in the AdS instability. Such black holes would then behave as massive particles, accelerate to the speed of light and cause a large backreaction, perhaps triggering collapse of the evanescent ergosurface, with the solution finally settling down to an almost BPS black hole solution with the same conserved charges as the microstate geometry, but different angular momenta.

The "fuzzball proposal" conjectures that supersymmetric microstate geometries provide a geometrical description of certain quantum microstates of supersymmetric black holes [106]. It is therefore interesting to compare whether the decay of linear waves in a microstate geometry resembles the decay for a supersymmetric black hole. For a supersymmetric black hole, waves are expected to decay as an inverse power law of time at late time outside the horizon. This has been proved for the extremal Reissner-Nordström spacetime [107, 108]. The slowest decaying modes are those with the lowest angular frequency. However, for a microstate geometry, the stable trapping implies that the decay is slower than any inverse power law, and the slowest decaying modes are those with the highest angular frequency. Hence there is a qualitative differences between the behaviour of linear waves in microstate geometries and in supersymmetric black hole geometries.

Another family of spacetimes with an evanescent ergosurface are supersymmetric "black lens" solutions [109, 110]. A black lens is a black hole with an event horizon of lens space topology. These solutions have an evanescent ergosurface outside the event horizon. Other examples of solutions with this property are obtained by superposing black holes with microstate geometries [111]. Our heuristic particle argument for instability may not apply to these solutions because the particle can fall across the horizon. However, the presence of the evanescent ergosurface implies that it is likely that all of these solutions will exhibit slow decay of linear perturbations and a corresponding nonlinear instability.

To define the evanescent ergosurface we need a Kaluza-Klein symmetry in 6d. It has been argued that there exist microstate geometries without such a symmetry [112]. (See also Ref. [113] and references therein.) In such geometries one cannot define an evanescent ergosurface. Nevertheless, as we will explain, we expect such geometries to admit stably trapped null geodesics through every point of the spacetime. Hence we expect that such

geometries will suffer from slow decay of linear perturbations and a corresponding nonlinear instability.

This chapter is organized as follows. In section 2.2 we review the notion of an evanescent ergosurface in 5d and 6d and show that such a surface is ruled by zero energy null geodesics. For supersymmetric microstate geometries we prove that these geodesics exhibit stable trapping. We also show that a 6d microstate geometry has a stably trapped null geodesic through every point of the spacetime. We elaborate on our heuristic argument for why these geometries are unstable. We then explain why the evanescent ergosurface presents a problem for proving linear stability of these geometries. Even if this problem can be overcome, we argue that the methods required will not extend to the nonlinear problem. In section 2.3.1 we discuss in detail the 3-charge microstate geometries of Ref. [91] and briefly the 2-charge geometries of Ref. [87]. In section 2.4 we determine quasinormal modes of these geometries in two ways: first using a matched asymptotic expansion (valid for large ℓ), and then using numerical methods (for general ℓ). We then explain why the properties of these quasinormal modes imply that generic linear perturbations must decay very slowly, in particular they cannot exhibit power-law decay.

2.2 Geodesics and stable trapping

2.2.1 Zero energy null geodesics

Supersymmetric solutions of 5d supergravity admit a non-spacelike Killing vector field V which approaches a standard time translation at infinity. In a 5d microstate geometry spacetime, V is timelike everywhere except on the evanescent ergosurface: a timelike hypersurface $\mathscr{S}$, on which V is null. In fact supersymmetry implies that there exists a scalar f such that [114]

$$V^2 = -f^2 \tag{2.2}$$

and $\mathscr{S}$ is given by $f = 0$. Since V is Killing, it preserves $\mathscr{S}$, i.e., V is tangent to $\mathscr{S}$. It is easy to see that V is tangent to affinely parameterized null geodesics on $\mathscr{S}$ [115]:

$$V^b \nabla_b V_a = -V^b \nabla_a V_b = -(1/2)\nabla_a (V^2) \tag{2.3}$$

and the RHS vanishes on $\mathscr{S}$ because V^2 has a second order zero on $\mathscr{S}$. Hence V is tangent to affinely parameterized null geodesics on $\mathscr{S}$.[2] The conserved energy of a timelike or null

[2]Note that this is *not* true for a general ergosurface (e.g. in the Kerr spacetime), when V^2 has only a first order zero and so the RHS is non-zero and orthogonal to $\mathscr{S}$ hence V is non-geodesic in that case.

geodesic with momentum P^a is

$$E = -V \cdot P \geq 0 \tag{2.4}$$

where the inequality follows because V is non-spacelike and V, P are both future-directed.

Since V is null on $\mathscr{S}$, it follows that V is tangent to *zero energy* null geodesics on $\mathscr{S}$. Furthermore, these are the only causal curves with zero energy: away from $\mathscr{S}$, V is timelike so $E = 0$ would imply that P is spacelike whereas on $\mathscr{S}$, $E = 0$ implies that P is tangent to V.

Microstate geometries carry non-zero angular momentum. Since V approaches a standard time translation at infinity, a particle following an orbit of V does not rotate w.r.t. to infinity, i.e., it has zero angular velocity. This means that the particle is resisting the frame-dragging effect arising from the rotation of the spacetime geometry. In this sense, the zero energy null geodesics can be regarded as having angular momentum opposite in sign to the angular momentum of the background geometry. If the microstate geometry has appropriate rotational symmetries then one can use these to define conserved angular momenta for geodesics; we will see below that at least one of the angular momenta of the zero energy null geodesics has opposite sign to that of the background.

2.2.2 The 6d perspective

Sometimes it is more convenient to discuss microstate geometries in 6d rather than 5d. In particular, this is the case for 2-charge microstate geometries, and the 3-charge geometries of Ref. [91], which are regular in 6d but not in 5d. Therefore we will need to discuss how $\mathscr{S}$ is defined in 6d.

The 5d Killing field V is the Kaluza-Klein reduction of a 6d Killing field, which we will also call V. Supersymmetry implies that V is globally null w.r.t. the 6d metric [116]. It can be written as $V = T + Z$ where T and Z are commuting Killing vector fields, Z is the spacelike "Kaluza-Klein" Killing vector field (i.e. the 5d metric is obtained from the 6d metric by projecting orthogonally to Z and rescaling) and, near infinity, T is timelike and canonically normalized.

V is null in 5d if, and only if, it is orthogonal to Z in 6d. Hence, in 6d, $\mathscr{S}$ can be defined as the locus where V is orthogonal to Z. On $\mathscr{S}$ we therefore have (using the fact that V is null)

$$T^2 = Z^2 = -T \cdot Z \tag{2.5}$$

For 2-charge microstate geometries, which do not correspond to regular 5d solutions, $\mathscr{S}$ is a 2d timelike submanifold on which Z vanishes (and hence T is null). For 3-charge microstate geometries, $\mathscr{S}$ is a timelike hypersurface in 6d (i.e. codimension 1). In the 3-charge case,

Z is non-vanishing on $\mathscr{S}$ so the above equations imply that T is spacelike on $\mathscr{S}$. Since T generates time translations in 6d, it follows that there is a genuine ergoregion present in 6d (this has been noticed before [97]).

In 6d, since V is globally null, it is everywhere tangent to affinely parameterized null geodesics. We use T to define the energy of geodesics in 6d: $E_6 = -T \cdot P$ where P is the momentum of the geodesic. We define the Kaluza-Klein momentum as $p = Z \cdot P$. We can use $-V \cdot P \geq 0$ to obtain[3] $E_6 \geq p$. Hence the 6d energy is bounded below despite the presence of the ergoregion. Since $V \cdot Z = 0$ on $\mathscr{S}$ it follows that the null geodesics on $\mathscr{S}$ with tangent V have zero Kaluza-Klein momentum $p = 0$ as well as zero 6d energy $E_6 = 0$.

2.2.3 Stable trapping

A geodesic is trapped if it "remains within a bounded region of space". Clearly this is true for the zero energy null geodesics on $\mathscr{S}$ discussed above. We will now show that the null geodesics on $\mathscr{S}$ tangent to V are *stably* trapped in the sense that initially nearby null geodesics remain nearby. This is intuitively obvious since these geodesics minimize the energy; we will now see it explicitly using the geodesic deviation equation (i.e. Jacobi fields). We will also show that *all* null geodesics with tangent V are stably trapped in a 6d supersymmetric microstate geometry, hence there exists a stably trapped null geodesic through every point of the spacetime.

We will first consider a more general situation of d-dimensional spacetime admitting a Killing vector field V. We define $\mathscr{T}$ to be the locus where V^2 is extremized, i.e., where $\nabla_a(V^2) = 0$. Using Killing's equation as in (2.3) we then have $V^b \nabla_b V^a = 0$ on $\mathscr{T}$. Since V must be tangent to $\mathscr{T}$, we have a family of affinely parameterized geodesics on $\mathscr{T}$ with tangent V.[4]

Let γ denote one of the geodesics on $\mathscr{T}$ with tangent V. Consider a 1-parameter family of affinely parameterized geodesics which contains γ [10]. Let X^a denote the tangent vector to these geodesics, and Y^a a deviation vector within this family, i.e., $\mathscr{L}_X Y = 0$. On γ we have $X^a = V^a$. We want to determine how Y^a behaves along γ. The geodesic deviation equation gives

$$(\nabla_V \nabla_V Y^a)\,|_\gamma = (\nabla_X \nabla_X)\, Y^a|_\gamma = R^a{}_{bcd} X^b X^c Y^d|_\gamma = R^a{}_{bcd} V^b V^c Y^d|_\gamma \tag{2.6}$$

[3] In the 2-charge microstate geometries, $V' = T - Z$ is also a globally null Killing vector field, which implies $E_6 \geq |p|$.

[4] The supersymmetric microstate geometries belong to this class of spacetime, as do extremal black holes. However, note that later in this section we will use that $\mathscr{T}$ is timelike, which is not the case for extremal black holes.

To evaluate the RHS we used the Killing vector identity

$$\nabla_c \nabla_a V_b = R_{bacd} V^d \tag{2.7}$$

This implies

$$R_{bacd} V^a V^d = \nabla_c \left(V^a \nabla_a V_b \right) - \left(\nabla_c V^a \nabla_a V_b \right) = H_{bc} + \omega^a{}_c \omega_{ab} \tag{2.8}$$

where

$$H_{ab} = H_{ba} = \nabla_a \nabla_b (-V^2/2) \tag{2.9}$$

and

$$\omega_{ab} = -\omega_{ba} = \nabla_b V_a \tag{2.10}$$

The geodesic deviation equation is therefore

$$\left[\nabla_V \nabla_V Y^a + \left(H^a{}_b + \omega^{ca} \omega_{cb} \right) Y^b \right]_\gamma = 0 \tag{2.11}$$

It will be convenient to rewrite this in terms of the Lie derivative w.r.t. V as follows:

$$\mathscr{L}_V \mathscr{L}_V Y^a = \nabla_V \nabla_V Y^a - (\nabla_V Y^b) \nabla_b V^a - Y^b V^c \nabla_c \nabla_b V^a - (\mathscr{L}_V Y^b) \nabla_b V^a \tag{2.12}$$

The identity (2.7) implies that the 3rd term on the RHS of (2.12) is zero. The first term is given by (2.11). Using this, (2.12) becomes

$$\left(\mathscr{L}_V \mathscr{L}_V Y^a + 2 \omega^a{}_b \mathscr{L}_V Y^b + H_{ab} Y^b \right)_\gamma = 0 \tag{2.13}$$

This is a second order ODE governing the evolution of Y^a along γ. Note that

$$\mathscr{L}_V \omega_{ab} = \mathscr{L}_V H_{ab} = 0 \tag{2.14}$$

which implies that (2.12) admits the first integral

$$\left(\mathscr{L}_V Y_a \right) \left(\mathscr{L}_V Y^a \right) + H_{ab} Y^a Y^b = C \tag{2.15}$$

where C is constant along the geodesic.

Now we assume that γ is a *null* geodesic and that Y^a is a deviation vector pointing to a nearby causal geodesic. To do this we consider a 1-parameter family of causal geodesics, so $X^2 \leq 0$. Since $X^2 = 0$ on γ, we see that X^2 is maximized on γ within our 1-parameter family.

Hence on γ we have

$$0 = \nabla_Y(X^2) = 2X^b Y^a \nabla_a X_b = 2X^b X^a \nabla_a Y_b = 2X^a \nabla_a (X \cdot Y) \tag{2.16}$$

where we used $\mathscr{L}_X Y = 0$ and the geodesic equation for X. It follows that $X \cdot Y$ is constant along γ, therefore $V \cdot Y$ is constant along γ so $V_a \mathscr{L}_V Y^a = 0$. Hence $\mathscr{L}_V Y^a$ must be spacelike or null so the first term in (2.15) is non-negative.

Note that H_{ab} is the Hessian of $-V^2/2$, which is extremized on $\mathscr{T}$. Therefore H_{ab} has components only in directions normal to $\mathscr{T}$. If we assume that $\mathscr{T}$ is a timelike submanifold then these normal directions are all spacelike. If $-V^2/2$ is *minimized* on $\mathscr{T}$ (as for a microstate geometry) then H_{ab} will be positive semi-definite, so we deduce that $C \geq 0$. Generically, H_{ab} will be positive definite when restricted to the space of vectors normal to $\mathscr{T}$. In this case, H_{ab} is a Riemannian metric on the space of vectors normal to $\mathscr{T}$. But we know that $H_{ab} Y^a Y^b \leq C$ hence the components of Y^a normal to $\mathscr{T}$ remain bounded. In other words, at the (infinitesimal) level of geodesic deviation, causal geodesics near to γ cannot move away from $\mathscr{T}$.

For a 5d supersymmetric microstate geometry, $\mathscr{T}$ coincides with the evanescent ergosurface $\mathscr{S}$, which is a hypersurface (i.e. a 4d submanifold). Furthermore, V^2 has a *second order* zero on $\mathscr{S}$. This implies that the Hessian can be written $H_{ab} = \alpha n_a n_b$ where $\alpha > 0$ is constant along γ and n_a is a unit spacelike normal to $\mathscr{S}$. The argument of the previous paragraph then gives $(n \cdot Y)^2 \leq C/\alpha$ hence the component of Y normal to $\mathscr{S}$ remains bounded so we have stable trapping in the direction normal to $\mathscr{S}$. Hence causal geodesics that are initially close to γ will remain close to $\mathscr{S}$.

Now consider the case in which V is globally null, e.g. a supersymmetric microstate geometry in 6d. In this case $\mathscr{T}$ is the entire spacetime and H_{ab} vanishes. However, we can see stable trapping as follows. From (2.7) we see that $\nabla_V \omega_{ab} = 0$ so the geodesic deviation equation (2.11) admits a first integral[5]

$$(\nabla_V Y_a)(\nabla_V Y^a) + \omega_{ac} \omega_b{}^c Y^a Y^b = C' \tag{2.17}$$

where C' is constant along the geodesic. As above, $V \cdot Y$ is constant along a geodesic γ with tangent V so $V_a \nabla_V Y^a = 0$. Hence $\nabla_V Y^a$ is spacelike or null so the first term above is non-negative. Hence we have

$$\omega_{ac} \omega_b{}^c Y^a Y^b \leq C'' \tag{2.18}$$

[5]Note that we cannot do this when $H_{ab} \neq 0$ because $\nabla_V H_{ab} \neq 0$ in general. The constants C and C' differ by a multiple of $\omega_{ab} Y^a \nabla_V Y^b$ which can be shown to be constant along γ using (2.11).

for some new constant C''. Note that the LHS is non-negative because ω_{ab} is orthogonal to V hence $\omega_{ab}Y^b$ is non-timelike.

Note that ω_{ab} is the rotation of the null geodesic congruence with tangent V.[6] As is usual when dealing with such a congruence, we can pick a null basis $\{e^a_\mu\}$ where $e_0 = V$ and e_1 is null with $e_0 \cdot e_1 = -1$ and e_i $(i = 2,3,\ldots,d-1)$ are orthonormal spacelike vectors orthogonal to e_0 and e_1. Furthermore, we can choose our basis to be parallelly transported along the geodesics of the congruence. In such a basis, the components $\omega_{\mu\nu}$ are constants along γ and $\omega_{0\mu} = 0$. Equation (2.18) becomes

$$\left(\omega_{i1}Y^1 + \omega_{ij}Y^j\right)\left(\omega_{i1}Y^1 + \omega_{ij}Y^j\right) \leq C'' \tag{2.19}$$

Next note that $Y^1 = -e_0 \cdot Y = -V \cdot Y$, which we showed above is constant along γ. Hence $\omega_{i1}Y^1$ is constant along γ so it follows from this equation that $\omega_{ij}Y^j$ is bounded (w.r.t. the norm δ_{ij}).

Now assume that our spacetime contains an evanescent ergosurface $\mathscr{S}$, i.e., a timelike surface with equation $Z \cdot V = 0$. Any covector normal to $\mathscr{S}$ is parallel to

$$n_a = \nabla_a(Z \cdot V) = Z^b\nabla_a V_b + V^b\nabla_a Z_b = -Z^b\nabla_b V_a - V^b\nabla_b Z_a = -2Z^b\nabla_b V_a = -2\omega_{ab}Z^b \tag{2.20}$$

with n_a spacelike (because $\mathscr{S}$ is timelike). Note that

$$n \cdot Y = 2\omega_{ab}Z^aY^b = 2\omega_{ij}Z^iY^j \tag{2.21}$$

where we used $Z^1 = Z \cdot V = 0$. We have just shown that $\omega_{ij}Y^j$ is bounded along γ, hence $n \cdot Y$ is also bounded. It follows that $\mathscr{S}$ exhibits stable trapping: deviation vectors cannot become large in the direction orthogonal to an evanescent ergosurface $\mathscr{S}$.

We can deduce a little more from the above analysis. We no longer assume that γ is on $\mathscr{S}$. We showed above that, along γ, Y^1 is constant and $\omega_{ij}Y^j$ is bounded. Now *assume* that ω_{ij} is non-degenerate. It follows that Y^j must be bounded along γ. In fact, it is easy to solve explicitly the geodesic deviation equation (2.11) to see that Y^i oscillates along γ, such that the mean value of $\omega_{ij}Y^j$ is $-\omega_{i1}Y^1$. One can then solve for Y^0, finding an oscillating term plus a term that grows linearly. The latter is "pure gauge": it can be eliminated by a change of affine parameter along the geodesics of the 1-parameter family. Having done this, all components of Y^a are bounded along γ. This is stable trapping. Hence if the congruence of null geodesics with tangent V has non-degenerate rotation matrix ω_{ij} then *any* geodesic in

[6] We emphasize that our 1-parameter family is not assumed to belong to this congruence, i.e., Y^a is a general deviation vector, not necessarily one associated with this congruence.

this congruence exhibits stable trapping. The constant Y^1 represents a shift from a geodesic γ in this congruence to a nearby geodesic γ' also within this congruence and the deviation vector describes oscillations about γ'.

We can apply this argument to supersymmetric microstate geometries in 6d.[7] We will show later that ω_{ij} is indeed everywhere non-degenerate for the 3-charge microstate geometries of [91–93], and also the 2-charge geometries of [87, 88]. It seems very unlikely that more complicated microstate geometries would have degenerate ω_{ij} so we expect ω_{ij} to be non-degenerate for general supersymmetric microstate geometries (including those lacking the Kaluza-Klein Killing vector field Z as in [112]). So we expect that the null geodesics with tangent V are all stably trapped in any supersymmetric microstate geometry. Hence there is a stably trapped null geodesic through every point of the 6d spacetime. Of course, these include the zero energy null geodesics on $\mathscr{S}$, which are singled out by the additional condition of having zero Kaluza-Klein momentum.

Away from $\mathscr{S}$ the stably trapped null geodesics have non-zero Kaluza-Klein charge p. From the 5d perspective, these null geodesics look like "BPS" charged particles, i.e., with mass equal to charge, which are at rest relative to infinity. It is familiar that such particles can remain at rest because they experience a cancellation of forces. But often this corresponds to neutral equilibrium (degenerate ω_{ij}, which allows linear growth of deviation vectors), whereas we have stable equilibrium. It would be interesting to investigate how this stability arises from the interaction of the particle with the various 5d fields.

In arguing for instability, we will focus on the consequences of the stable trapping on $\mathscr{S}$ because in this case we have stable trapping of null geodesics in 5d as well as is 6d. The consequences of the stable trapping away from $\mathscr{S}$ in 6d would be interesting to explore further.

2.2.4 Heuristic argument for instability

In the Introduction, we presented a heuristic argument that supersymmetric microstate geometries experience an instability because a massive uncharged 5d particle will accelerate to the speed of light on $\mathscr{S}$, and cause a large backreaction. We will now discuss this in more detail.

Let Σ_0 be a spacelike Cauchy surface for a 5d microstate geometry. Choose coordinates x^i on Σ_0 and let t be the parameter distance from Σ_0 along the integral curves of V. Carry the coordinates x^i along these integral curves to define coordinates (t, x^i). The metric can then be

[7] In 10d, ω_{ij} is degnerate in directions associated with the internal T^4. However, the compactness of this space prevents the geodesics from dispersing in these directions.

written in ADM form

$$ds^2 = -N^2 dt^2 + h_{ij}\left(dx^i - \Omega^i dt\right)\left(dx^j - \Omega^j dt\right) \tag{2.22}$$

where

$$N^2 = f^2 + h_{ij}\Omega^i\Omega^j, \tag{2.23}$$

$V = \partial/\partial t$ is the stationary Killing vector field, and $f = 0$ on $\mathscr{S}$. In general there is freedom to shift t by a function of the other coordinates.

For the 3-charge microstate geometries that we will study later,[8] we can split the coordinates as $x^i = (x^I, x^\alpha)$ such that $\partial/\partial x^I$ $(I = 1,2)$ are Killing vectors associated to rotational symmetries, and $\Omega^\alpha = 0$, and it is natural to chose Σ_0 so that $\partial/\partial x^I$ are tangent to it, which eliminates the freedom to shift t.

We will consider a family of local observers whose velocity is othogonal to surfaces of constant t. The velocity of such an observer is

$$u^a = -N(dt)^a = \frac{1}{N}\left(\frac{\partial}{\partial t} + \Omega^i \frac{\partial}{\partial x^i}\right) \tag{2.24}$$

For a microstate geometry with rotational symmetries, the velocity of these observers is orthogonal to $\partial/\partial x^I$ and so they have zero angular momentum. Hence they are referred to as "zero angular momentum observers" (ZAMOs). Note that they rotate with angular velocities Ω^I w.r.t. a stationary observer at infinity. This is because of the frame-dragging caused by the rotation of the spacetime. For a general microstate geometry we don't expected any rotational symmetries but we will still refer to these observers as ZAMOs. In general there is the freedom to shift t by a function of x^i so there are many different families of ZAMOs.

Now consider a particle with mass μ. Its momentum P_a obeys

$$-\mu^2 = g^{ab}P_a P_b \tag{2.25}$$

which can be rearranged to give

$$E^2 - 2EJ - \frac{f^2}{h_{jk}\Omega^j\Omega^k}J^2 = N^2\left(\mu^2 + H^{kl}P_k P_l\right) \equiv \Delta^2 \tag{2.26}$$

Here $E = -P_t \geq 0$ is the energy of the particle (conserved if it follows a geodesic) and

$$J = \Omega^i P_i \tag{2.27}$$

[8] These have a pair of orbifold singularities when reduced to 5d but that is not relevant to this argument.

We have decomposed P^i so that the component of P_i along Ω^i appears on the LHS of (2.26) and the orthogonal component appears on the RHS where we have defined H^{ij} to be the projection of h^{ij} orthogonal to Ω^i:

$$H^{ij} = h^{ij} - \frac{\Omega^i \Omega^j}{h_{kl}\Omega^k \Omega^l} \tag{2.28}$$

For a microstate geometry with rotational symmetries, we have $J = \Omega^I P_I$ and P_I are the angular momenta of the particle, which are conserved if the particle follows a geodesic.

Note that the energy of the particle according to a ZAMO is

$$E_{\text{ZAMO}} = -u \cdot P = \frac{1}{N}(E - J) \tag{2.29}$$

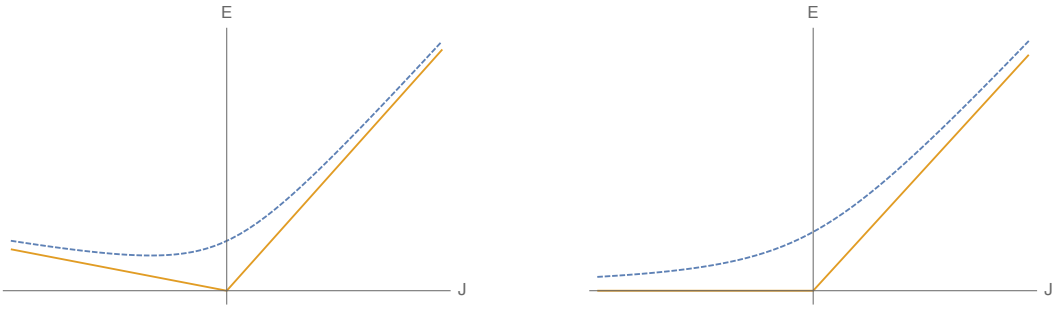


Fig. 2.1 Plots of E against J. Dashed blue curves: $\Delta > 0$, solid orange lines: $\Delta = 0$. Left: a generic point of spacetime. Right: on an evanescent ergosurface.

To formulate our argument for instability, it is useful to consider equation (2.26). At a generic point of a microstate geometry spacetime we have $f \neq 0$ and Figure 2.1 (left) shows E as a function of J for fixed Δ. The minimum value of E is positive and occurs at some finite value of J. However, at an evanescent ergosurface, we have $f = 0$ and the corresponding figure is shown on the right of Fig. 2.1. If $\Delta > 0$ then E is minimized at $J = -\infty$.

First consider a massive particle $\mu > 0$. If the particle is free then it will move on a geodesic, so E is conserved. However, when interactions are included, the particle couples to gravitational radiation (and other massless field), and therefore gradually loses energy through radiation. If $E < \mu$ then the particle cannot escape to infinity. Its energy E will decrease over time and approach its minimum value. From the plots, it is clear that the energy is minimized on the evanescent ergosurface, and this minimum occurs at $J = -\infty$ for a massive particle (as $\Delta > 0$). Hence the particle must "roll down the hill" to $J = -\infty$. This

implies that E_{ZAMO} will diverge, i.e., the local observer will measure infinite energy. This strongly suggests that the spacetime will be unstable.[9]

Now consider a massless particle, $\mu = 0$. If the particle starts on a stably trapped geodesic then it cannot escape to infinity. As for the massive particle, E will gradually decrease so we can apply the above argument when $\Delta > 0$. However, it is possible that the particle will radiate in such a way that it approaches a final state with $\Delta = 0$, in which case it can eventually reach $E = 0$ at finite negative J. This corresponds to to one of the null geodesics tangent to V on $\mathscr{S}$. However, there is nothing preventing this endpoint from having arbitrarily large J, so one might expect generically that this will be the case simply because there is more phase space available at large J. This again suggests instability.

2.2.5 The energy functional

We will now discuss the consequences of the existence of an evanescent ergosurface for linear perturbations of microstate geometries. We will explain how establishing even *linear* stability in such backgrounds is problematic, and then discuss the consequences for nonlinear stability.

Known microstate geometry solutions can be obtained as solutions of 6d supergravity. For these solutions, Ref. [100] showed that one can identify certain decoupled sectors of linear perturbations for which the 6d equation of motion is simply that of a massless, uncharged, scalar field, i.e., the wave equation. If this field does not vary around the Kaluza-Klein circle then it will also satisfy the wave equation in 5d.

The usual method for establishing that solutions of the wave equation remain bounded in time is based on the existence of a conserved energy functional. Consider a globally hyperbolic spacetime with a causal Killing vector field V. A field Φ satisfying the wave equation has a conserved energy momentum tensor

$$T_{ab} = \partial_a \Phi \partial_b \Phi - \frac{1}{2} g_{ab} \left(\partial \Phi\right)^2 \tag{2.30}$$

We can define a conserved energy-momentum current for Φ:

$$j^a = -T^a{}_b V^b \tag{2.31}$$

[9]Note that one could not apply this argument in a supersymmetric black hole spacetime because the particle would fall across the horizon with non-zero E.

Let Σ_0 be a spacelike Cauchy surface and let Σ_t be the image of Σ_0 by moving parameter distance t along the integral curves of V. The energy of Φ on Σ_t is then

$$E_t[\Phi] = -\int_{\Sigma_t} \sqrt{h}\, n \cdot j \tag{2.32}$$

where h is the determinant of the induced metric on Σ_t and n is the future-directed unit normal to Σ_t.

Since T_{ab} satisfies the dominant energy condition, j^a must be causal and future-directed, or zero. This implies that $E_t \geq 0$. Since j is conserved, it follows that if $t' > t$ then we have $E_{t'} \leq E_t$. (Here we allow for the possibility of the surfaces extending to future null infinity, in which case energy can be lost by radiation through null infinity.) Hence if E_0 is small then E_t remains small for all $t > 0$.

Consider the integrand of E_t. The dominant energy condition implies that $-n \cdot j \geq 0$ with equality if, and only if, $j = 0$. But $j = 0$ implies (by contracting with $d\Phi$) that $V \cdot \partial\Phi = 0$ and $(\partial\Phi)^2 = 0$. If V is timelike then this implies $d\Phi = 0$. However, if V is null then it implies only that $d\Phi$ parallel to V.

If V is timelike everywhere then E is a positive-definite functional of $d\Phi$, i.e., E defines a norm for $d\Phi$. If there exist additional Killing vector fields K^I that span the tangent space of Σ_t then one can commute the wave equation several times with these vector fields to obtain bounds on $E[K^{I_1} \ldots K^{I_N}\Phi]$ and hence control the norm of higher derivatives of Φ. The Sobolev embedding theorem can then be used to bound Φ. This process may be adapted in several ways: the commuting vector fields need not be exactly Killing, they may only span a submanifold of Σ_t (e.g. [117]), or the commutation may be with higher order, tensorial operators rather than vector fields (e.g. [118]).

Now consider a 5d supersymmetric microstate geometry. In this case, V is null on $\mathscr{S}$. Hence on $\mathscr{S}$, E fails to control the component of $d\Phi$ in the direction of V so E is not positive definite and the above argument for demonstrating boundedness of Φ does not work. Conservation of energy does not prevent $d\Phi$ from becoming large on $\mathscr{S}$.[10]

This problem arises also for stationary black hole geometries, where V becomes null at the horizon. For a non-extremal black hole, this problem is overcome by exploiting the "horizon redshift effect". This arises from the fact that affinely parameterized horizon generators have tangent $e^{-\kappa t}V$ where κ is the surface gravity and t is a parameter along the integral curves of

[10] From the 6d perspective, the functional E gives the difference $E_6 - p$ where E_6 is the 6d energy (defined using the Killing field T) and p the Kaluza-Klein momentum (defined using the Killing field Z). If we restrict attention to fields Φ invariant around the KK circle, i.e., $Z \cdot \partial\Phi = 0$, then we have $p = 0$ so $E_6 = E \geq 0$. Since V is globally null, E fails everywhere to control the component of $d\Phi$ along V. But we have imposed the additional condition $Z \cdot \partial\Phi = 0$, so $d\Phi$ can be proportional to V only when V is orthogonal to Z, i.e., on $\mathscr{S}$.

V. Hence a photon travelling along a horizon generator suffers a redshift $e^{-\kappa t}$. The wave analogue of this effect enables one to control the behaviour of the problematic component of $d\Phi$ at the horizon [81, 119]. However, this effect is absent for an extremal black hole. In the extremal case, it turns out that the problematic component of $d\Phi$ remains bounded but higher derivatives blow up along the horizon, i.e., there is an instability [107, 108, 120, 121].

For a supersymmetric microstate geometry, V is tangent to affinely parameterized geodesics on $\mathscr{S}$ so there is no analogue of the horizon redshift effect that can be used to control the behaviour of $d\Phi$ on $\mathscr{S}$. To control the problematic component of $d\Phi$ on $\mathscr{S}$ one might attempt to proceed as follows. First introduce an everywhere timelike vector field W which agrees with V everywhere except near $\mathscr{S}$. Now use W to define an energy functional. This new energy functional will be non-degenerate (i.e. it defines a norm on $d\Phi$) but non-conserved. The idea is that we can control the problematic component of $d\Phi$ by commuting the wave equation with Killing vector fields or higher order operators. In particular, if the microstate geometry admits angular momentum operators which commute with the wave operator, then we can first commute with these operators, in order to obtain a bound on the associated higher order energy. We could then integrate this bound in time to show that the non-degenerate energy can grow at most linearly in time. But of course this does not exclude an instability. Alternatively, if a version of Hardy's inequality (see e.g. [17]) can be proved on these backgrounds, then a similar argument could be employed in order to show that the nondegenerate energy is bounded for all time.

These arguments will only work when the background has appropriate symmetries, which will not be the case for a general microstate geometry. Furthermore, even when the background has such symmetries, these arguments are unlikely to extend to the nonlinear problem. In the nonlinear problem we would no longer have an exactly conserved energy so if we were to try to bound the energy of a perturbation by its initial value then we would encounter various error terms. In order to prove stability, we need to bound these error terms in a suitable way in terms of the initial data. This is often done in the context of a bootstrap argument: the error terms are assumed to satisfy certain bounds, which allows the energy to be bounded, and this in turn allows the initial assumptions on the error terms to be verified and improved. However, if we take the approach suggested above for the linear problem, and first commute the equation with (approximate) angular momentum operators, then the error terms will involve higher derivatives of the field, so we will need to assume bounds on higher-order energies in order to be able to bound lower-order energies. However, this scheme can never "close" – in order to bound these higher-order energies, we would need to assume bounds on even higher order energies, and so on.

In summary, the existence of an evanescent ergosurface implies that standard methods for establishing boundedness of solutions of the linear wave equation do not work in supersymmetric microstate geometries. It is conceivable that this problem could be overcome for microstate geometries admitting suitable rotational symmetries. But such geometries are not typical and furthermore, the methods required are not robust enough to extend to the nonlinear problem.

2.3 Supersymmetric microstate geometries

2.3.1 3-charge metric and charges

In this section we will study in detail the 3-charge microstate geometries of Refs. [91–93]. These are supersymmetric solutions of type IIB supergravity compactified on T^4. The resulting 6d geometry asymptotically approaches the product of 5d Minkowski spacetime with a Kaluza-Klein circle of radius R_z. We will focus on the case for which the 6d geometries are smooth with no conical or orbifold singularities. These geometries can be reduced to 5d however the 5d metric has a pair of orbifold singularities so it is more convenient to work in 6d.

These solutions admit 4 Killing vector fields and a "hidden" symmetry (associated to a Killing tensor field) which enables one to separate the wave equation (and Hamilton-Jacobi equation for geodesics) into ODEs.

The 3 charges of these solutions arise from n_1 D1-branes wrapped around the Kaluza-Klein S^1, n_2 D5-branes wrapped around $S^1 \times T^4$, and n_p units of momentum around the S^1 where

$$n_p = n(n+1)n_1 n_2 \qquad n \in \mathbb{Z} \tag{2.33}$$

The solution is written in terms of dimensionful charges

$$Q_1 = \frac{(2\pi)^4 g\alpha'^3}{V} n_1 \qquad Q_2 = g\alpha' n_2 \qquad Q_p = a^2 n(n+1) = \frac{4G^{(5)}}{\pi R_z} n_p \tag{2.34}$$

where g is the string coupling constant, V is the volume of the T^4, $G^{(5)}$ is the 5d Newton constant, α' is a constant related to the tension of the string and the length scale a is defined by

$$a = \frac{\sqrt{Q_1 Q_2}}{R_z} \tag{2.35}$$

The 10d string frame metric is:

$$\begin{aligned}
ds^2 =& -\frac{1}{h}(dt^2-dz^2)+\frac{Q_p}{hf}(dt-dz)^2+hf\Big(\frac{dr^2}{r^2+(\tilde\gamma_1+\tilde\gamma_2)^2\eta}+d\theta^2\Big)\\
&+h\Big(r^2+\tilde\gamma_1(\tilde\gamma_1+\tilde\gamma_2)\eta-\frac{(\tilde\gamma_1^2-\tilde\gamma_2^2)\eta Q_1Q_2\cos^2\theta}{h^2f^2}\Big)\cos^2\theta d\psi^2\\
&+h\Big(r^2+\tilde\gamma_2(\tilde\gamma_1+\tilde\gamma_2)\eta+\frac{(\tilde\gamma_1^2-\tilde\gamma_2^2)\eta Q_1Q_2\sin^2\theta}{h^2f^2}\Big)\sin^2\theta d\phi^2\\
&+\frac{Q_p(\tilde\gamma_1+\tilde\gamma_2)^2\eta^2}{hf}(\cos^2\theta d\psi+\sin^2\theta d\phi)^2\\
&-2\frac{\sqrt{Q_1Q_2}}{hf}\Big(\tilde\gamma_1\cos^2\theta d\psi+\tilde\gamma_2\sin^2\theta d\phi\Big)(dt-dz)\\
&-2\frac{(\tilde\gamma_1+\tilde\gamma_2)\eta\sqrt{Q_1Q_2}}{hf}\Big(\cos^2\theta d\psi+\sin^2\theta d\phi\Big)dz+\sqrt{\frac{H_1}{H_2}}\Sigma_{i=1}^4dx_i^2\\
=&ds_6^2+\sqrt{\frac{H_1}{H_2}}\Sigma_{i=1}^4dx_i^2
\end{aligned}\tag{2.36}$$

where

$$\eta=\frac{Q_1Q_2}{Q_1Q_2+Q_1Q_p+Q_2Q_p},\tag{2.37}$$

$$\tilde\gamma_1=-an,\ \ \tilde\gamma_2=a(n+1),\tag{2.38}$$

$$\begin{aligned}
f&=r^2+(\tilde\gamma_1+\tilde\gamma_2)\eta(\tilde\gamma_1\sin^2\theta+\tilde\gamma_2\cos^2\theta)\\
&=r^2+a^2\eta(-n\sin^2\theta+(n+1)\cos^2\theta),
\end{aligned}\tag{2.39}$$

$$H_1=1+\frac{Q_1}{f},\ \ H_2=1+\frac{Q_2}{f}\ \text{ and }\ h=\sqrt{H_1H_2},\tag{2.40}$$

where $\theta\in[0,\pi/2]$, $r>0$ and $0\le\phi,\psi\le2\pi$.

It was shown in [91] that these geometries are indeed smooth. Note that as $f\to0$, hf takes a finite (non-zero) value, and f only appears explicitly in the metric when multiplied by h. There could also be a problem coming from the fact that $h\to\infty$ as $f\to0$, but it turns out that the brackets multiplying h all tend towards f in this limit, again resulting in the finite combination hf. Thus there are no singularities as $f\to0$. In the limit $r\to0$ it is necessary to change coordinates to show that there is no singularity here; this was done in [91]. The coordinate transformation is given in section 2.4.2.1 where we discuss the boundary condition at $r=0$.

The angular momenta of these geometries are

$$J_\psi = -nn_1n_5 \qquad J_\phi = (n+1)n_1n_5, \tag{2.41}$$

It is worth noting that we will need to work in the Einstein frame in 6d but that when we reduce from 10 to 6 dimensions and then go to the Einstein frame, the factors involved cancel so the 6d Einstein metric is exactly the same as ds_6^2, the 6d part of the 10d string frame metric in (2.36).

In 6d, the spacetime is asymptotically Kaluza-Klein, *i.e.* the direct product of 5d Minkowski space with a compact direction. In the 'near region' the geometry forms a throat with topology $AdS_3 \times S^3$, which caps off smoothly at $r = 0$ [91–93].

2.3.2 Evanescent ergosurface and zero energy null geodesics

The above solution is supersymmetric and therefore admits a globally defined null Killing vector field:

$$V = T + Z \tag{2.42}$$

where

$$T = \frac{\partial}{\partial t} \qquad Z = \frac{\partial}{\partial z}. \tag{2.43}$$

As discussed in section 2.2.2, the evanescent ergosurface $\mathscr{S}$ is defined as the surface where the Kaluza-Klein Killing vector field Z is orthogonal to V. We have $V \cdot Z = 1/h$ and hence $\mathscr{S}$ is the surface where h diverges, i.e., where $f = 0$. Solving the equation $f = 0$ for $0 < r < \infty$ gives the following ranges of θ on $\mathscr{S}$ [92]:

- $n > 0$: $\theta \in I_{n>0} = [\tilde{\theta}, \pi/2]$ where $\tan\tilde{\theta} = \sqrt{\frac{n+1}{n}}$;
- $n < 0$: $\theta \in I_{n<0} = [0, \tilde{\theta}]$.

It was shown in [93] that the 6d metric is regular on $\mathscr{S}$ and that $\mathscr{S}$ has topology $S^1 \times S^3$.

Due to the symmetries of the spacetime, if U is the tangent vector to an affinely parameterized geodesic then the quantities $p_I = (\partial/\partial x^I) \cdot U$ are conserved along the geodesic, where $x^I \in \{t, z, \phi, \psi\}$. As discussed in section 2.2.2, V is everywhere tangent to null geodesics. The conserved quantities associated to these geodesics are

$$p_t = -h^{-1} \qquad p_z = h^{-1} \qquad p_\psi = -\frac{\sqrt{Q_1Q_2}}{hf} a\eta\cos^2\theta \qquad p_\phi = -\frac{\sqrt{Q_1Q_2}}{hf} a\eta\sin^2\theta. \tag{2.44}$$

On $\mathscr{S}$, these become

$$p_t = 0,\ \ p_z = 0,\ \ p_\psi = -a\eta\cos^2\theta,\ \ p_\phi = -a\eta\sin^2\theta. \tag{2.45}$$

so the energy ($-p_t$) and Kaluza-Klein charge (p_z) both vanish on $\mathscr{S}$, as expected from section 2.2.2. Note that $p_\phi + p_\psi$ has opposite sign to $J_\phi + J_\psi$; in this sense, the geodesics have angular momenta opposed to those of the background geometry. If we define $J_L = J_\phi - J_\psi$ and $J_R = J_\phi + J_\psi$ then the background geometry has $J_L = (2n+1)n_1n_5$, $J_R = n_1n_5$ so if $n, n_1, n_5 \gg 1$ then $J_L \gg J_R \gg 1$. The backreaction of particles following geodesics on $\mathscr{S}$ will tend to reduce J_R so it is plausible that the final state of the instability will be a near-extremal BMPV black hole [99], which has $J_R \approx 0$.[11] A black ring always has $J_R \neq 0$ [125], so although it is possible for J_R to become arbitrarily small as the radius of the ring tends to zero, it seems more likely that the end result wil be a near-extremal BMPV black hole.

The energy of these geodesics as measured by a *local* observer is not small. For example, consider a zero angular momentum observer (ZAMO) (as in section 2.2.4) with velocity u^a given by

$$u^a = -\frac{(dt)^a}{\sqrt{-g^{tt}}} \tag{2.46}$$

On $\mathscr{S}$, a ZAMO measures the energy of a null geodesic with momentum V to be

$$E_{ZAMO} = -u\cdot V = \sqrt{Q_1Q_2}\left(Q_1+Q_2+Q_p+\frac{Q_1Q_2+Q_1Q_p+Q_2Q_p}{a^2\eta\left((n+1)\sin^2\theta - n\cos^2\theta\right)}\right)^{-\frac{1}{2}}. \tag{2.47}$$

As discussed in section 2.2.3, the condition for the null geodesics with tangent V to be stably trapped everywhere is for the rotation matrix ω_{ij} of the null geodesic congruence with tangent V to be non-degenerate. One can define the rotation as follows [10]. At any point, consider the space of vectors orthogonal to V quotiented by the subspace of vectors proportional to V. This defines a 4d vector space $\mathscr{V}$, and $\omega = -(1/2)dV$ can be regarded as a 2-form acting on vectors in this space. We want to ask whether this 2-form is non-degenerate. So we need to calculate dV. We start from

$$V = -h^{-1}(dt-dz) + C(hf)^{-1}\left(\cos^2\theta\, d\psi + \sin^2\theta\, d\phi\right) \tag{2.48}$$

[11] The BMPV bound in this notation is $J_L^2/4 \leq n_1n_5n_p$ [122–124].

where C is a constant and hence

$$\begin{aligned} dV = &\frac{1}{2}\big((Q_1+Q_2)f+2Q_1Q_2\big)(hf)^{-3}\Big[r(dt-dz)\wedge dr - a^2\eta(2n+1)\sin\theta\cos\theta(dt-dz)\wedge d\theta\Big] \\ &+\frac{C}{2}(2f+Q_1+Q_2)(hf)^{-3}r\Big[\cos^2\theta\, d\psi\wedge dr+\sin^2\theta\, d\phi\wedge dr\Big] \\ &+\frac{C}{2}\sin\theta\cos\theta(hf)^{-3}\Big[2(hf)^2(d\psi\wedge d\theta - d\phi\wedge d\theta) \\ &\qquad -a^2\eta(2n+1)(2f+Q_1+Q_2)\left(\cos^2\theta\, d\psi\wedge d\theta+\sin^2\theta\, d\phi\wedge d\theta\right)\Big]. \end{aligned} \tag{2.49}$$

Now we want to show that this is non-degenerate by acting on an arbitrary vector $X\in\mathscr{V}$. Since $X\sim X+\alpha V$ we can choose X so that $X^t=0$. The condition $X\cdot V=0$ then fixes X^z. We now consider $(dV)_{ab}X^b$ as a covector acting on $\mathscr{V}$ so we neglect terms proportional to V_a in $(dV)_{ab}X^b$. The result is that this covector vanishes if, and only if, $X^r=X^\theta=X^\phi=X^\psi=0$ and hence $X^z=0$. Therefore dV is non-degenerate, viewed as a quadratic form on $\mathscr{V}$. Hence the rotation matrix is non-degenerate. By setting $Q_p=0$ one sees that this result applies also to the 2-charge microstate geometries discussed in section 2.3.3.

2.3.3 2-charge microstate geometries

2.3.3.1 The metric

We consider the 2-charge supersymmetric microstate geometries constructed in [87]. These are obtained by setting $Q_p=0$ in the solution described in section 2.3.1. Ref. [89] describes a whole family of such solutions, but we will only consider the maximally rotating solution with a circular profile. The 6d metric for this 2-charge $D1-D5$ microstate geometry (in the form given in [126]) is

$$\begin{aligned} ds^2 = &-\frac{1}{h}(dt^2-dz^2)+hf\left(d\theta^2+\frac{dr^2}{r^2+a^2}\right)-\frac{2a\sqrt{Q_1Q_2}}{hf}\left(\cos^2\theta dz d\psi+\sin^2\theta dt d\phi\right) \\ &+h\Big[\left(r^2+\frac{a^2Q_1Q_2\cos^2\theta}{h^2f^2}\right)\cos^2\theta d\psi^2+\left(r^2+a^2-\frac{a^2Q_1Q_2\sin^2\theta}{h^2f^2}\right)\sin^2\theta d\phi^2\Big] \end{aligned} \tag{2.50}$$

where

$$f=r^2+a^2\cos^2\theta,\ \ h=\Big[\left(1+\frac{Q_1}{f}\right)\left(1+\frac{Q_2}{f}\right)\Big]^{1/2}, \tag{2.51}$$

a is defined in (2.35), $\theta\in[0,\pi/2]$, $r>0$ and $0\le\phi,\psi\le 2\pi$..

The spacetime is non-singular [87]; it is asymptotically $R^5 \times S^1$, while the 'near region' is the direct product $AdS_3 \times S^3$. The S^3 asymptotes to non-zero size as r tends to zero; the'throat' is the region where the size of the S^3 is close to its asymptotic value [126].

2.3.3.2 Evanescent ergosurface

As in the 3-charge microstate geometries, the globally null Killing vector field is

$$V = \frac{\partial}{\partial t} + \frac{\partial}{\partial z} \tag{2.52}$$

and the evanescent ergosurface $\mathscr{S}_2$ defined by $V \cdot Z = 0$ is at $f = 0$; this is defined by $r = 0$ and $\theta = \pi/2$.

In the 2-charge geometry the Kaluza-Klein circle pinches off smoothly at $f = 0$ [89]. The $\psi-$direction also shrinks to zero size at $f = 0$ (in the same way as at the origin of polar coordinates) so that at constant t, $\mathscr{S}_2$ has topology S^1 where the coordinate around this circle is ϕ. There are several differences between the evanescent ergosurface in the 2- and 3-charge geometries. First of all they have different dimensions: the 3-charge $\mathscr{S}$ is 5 dimensional whilst the 2-charge $\mathscr{S}_2$ is only 2 dimensional. In the 2-charge case the Killing vector field $T = \partial/\partial t$ is timelike everywhere except on $\mathscr{S}_2$ where it is null (V is null everywhere and $Z = \partial/\partial z$ vanishes on $\mathscr{S}_2$) and so in this case there is no ergoregion, in contrast with the 3-charge case where T is spacelike on $\mathscr{S}$.

There are zero energy null geodesics with tangent vector V which are stably trapped on $\mathscr{S}_2$ and thus stay at constant $r = 0$, $\theta = \pi/2$. In the same way as for the 3-charge geometry this follows from the discussion in section 2.2.

2.4 Quasinormal modes

2.4.1 Relation to null geodesics

We will now consider the wave equation

$$\Box\Phi = 0 \tag{2.53}$$

in the geometry (2.36). The geometric optics approximation tells us that we can expect to find rapidly varying solutions of this equation which are localized around null geodesics for

an arbitrarily long time.[12] Therefore we expect there to exist solutions of the wave equation that are localized around a null geodesic with tangent V. Of course, such solutions will eventually decay by dispersion to infinity.

In this section, we will show that such solutions can be constructed as quasinormal modes, i.e., modes with definite frequency ω. For black hole solutions, it is known that quasinormal mode frequencies can be related to properties of trapped null geodesics in the geometric optics limit [127, 52].

Recall the relation between quasinormal mode solutions and unstably trapped null geodesics on the photon sphere for a Kerr black hole, discussed in section 1.2.2. We will do something similar for the wave equation in the spacetime (3.54). It has been shown that the wave equation separates in this geometry [92] so we will look for solutions of the form

$$\Phi(t,z,r,\theta,\phi,\psi) = e^{-i\omega t+i\lambda z+im_\psi\psi+im_\phi\phi}\Phi_r(r)\Phi_\theta(\theta). \tag{2.54}$$

where the angular harmonics Φ_θ are labelled by an integer ℓ.

By analogy with the Kerr case in section 1.2.2, for large ℓ we expect there to exist quasinormal modes which are closely related to the trapped null geodesics. There are several important differences to the Kerr case. First, in the geometry (2.36), the trapping is *stable* so we expect ω_I to be much smaller than in the Kerr case. Second, on $\mathscr{S}$, the trapped null geodesics have zero energy and KK momentum. Hence we expect to find quasinormal modes with $\lambda = 0$ such that $\omega_R/\ell \approx 0$, i.e., ω_R does not scale with ℓ.

We can also consider a null geodesic with tangent V that does not lie on $\mathscr{S}$. Such geodesics have $-p_t = p_z$ so we would expect there to exist corresponding quasinormal modes with $\lambda \neq 0$ and $\omega \approx \lambda$. We will look for these modes by taking $\lambda = \mathscr{O}(\ell)$ and $\omega - \lambda = \mathscr{O}(1)$.

We will determine quasinormal modes in two ways. For large ℓ we will use a matched asymptotic expansion inspired by a similar calculation in [128]. For general ℓ we will determine quasinormal modes numerically. For both methods we will need to use the ODEs resulting from separation of variables, which are [92]

$$\frac{1}{\sin 2\theta}\frac{\mathrm{d}}{\mathrm{d}\theta}\left(\sin 2\theta\frac{\mathrm{d}\Phi_\theta(\theta)}{\mathrm{d}\theta}\right) + \left[A - \frac{m_\psi^2}{\cos^2\theta} - \frac{m_\phi^2}{\sin^2\theta} + (\tilde{\omega}^2 - \tilde{\lambda}^2)\frac{a^2\eta}{R_z^2}(\cos^2\theta + n\cos 2\theta)\right]\Phi_\theta(\theta) = 0 \tag{2.55a}$$

$$\frac{1}{r}\frac{\mathrm{d}}{\mathrm{d}r}\left[r(r^2+\alpha^2)\frac{\mathrm{d}\Phi_r(r)}{\mathrm{d}r}\right] + \left(\tilde{\kappa}^2 r^2 + 1 - \tilde{\nu}^2 + \frac{\xi^2 s^2}{r^2+\alpha^2} - \frac{\zeta^2 s^2}{r^2}\right)\Phi_r(r) = 0\,, \tag{2.55b}$$

[12] Furthermore, the results of Ref. [31] prove that the energy of the solution is close to the energy of the corresponding null geodesic.

where A is a constant arising from the separation of variables and

$$\tilde{\omega} = \omega R_z, \qquad \tilde{\lambda} = \lambda R_z, \qquad s = \frac{\sqrt{Q_1 Q_2}}{R_z^2}, \qquad \alpha = s\sqrt{\eta}, \qquad \tilde{\kappa} = \sqrt{\tilde{\omega}^2 - \tilde{\lambda}^2} \tag{2.56}$$

$$\tilde{\nu} = \sqrt{1 + A - \tilde{\kappa}^2 \frac{Q_1 + Q_2}{R_z^2} - (\tilde{\omega} - \tilde{\lambda})^2 \frac{Q_p}{R_z^2}}, \tag{2.57}$$

$$\xi = \sqrt{\eta}\left[\frac{\tilde{\omega}}{\eta} - \tilde{\lambda}\frac{Q_p(Q_1+Q_2)}{Q_1 Q_2} + n m_\psi - m_\phi\,(n+1)\right], \tag{2.58}$$

$$\zeta = \sqrt{\eta}\left[\tilde{\lambda} + m_\psi\,(n+1) - n m_\phi\right]. \tag{2.59}$$

2.4.2 Matched asymptotic expansion

We will look first for quasinormal modes corresponding to the null geodesics with tangent V that are on, or near to, $\mathscr{S}$. On $\mathscr{S}$ these have $p_t = p_z = 0$ and non-zero p_ϕ, p_ψ in general. Therefore we look for quasinormal modes with $|m_\psi|, |m_\phi| \gg 1$ while keeping $\{\tilde{\omega}, \tilde{\lambda}\} = O(1)$ in (2.54). Our aim is to solve the coupled system of equations (2.55) for the eigenvalue pair $\{A, \tilde{\omega}\}$. It turns out that if either $|m_\phi|$ or $|m_\psi|$ are large, the two eigenvalues essentially decouple. That is to say, one can first determine A and a *posteriori* determine $\tilde{\omega}$.

To see how this works in more detail, we start by looking at the angular equation (2.55a). In the $|m_\psi|, |m_\phi| \to \infty$ limit, while keeping $\{\tilde{\omega}, \tilde{\lambda}\}$ fixed, we can introduce the effect of $\tilde{\omega}$ and $\tilde{\lambda}$ perturbatively. At leading order, we can ignore the term proportional to $\tilde{\omega}^2 - \tilde{\lambda}^2$ in (2.55a), so that it becomes the equation for spherical harmonics on S^3 with known eigenvalues $A = \ell(\ell+2) \equiv \mu_\ell^2$ where

$$\ell \geq |m_\psi| + |m_\phi|, \quad \ell \in \mathbb{Z}. \tag{2.60}$$

From (2.60), $|m_\psi|, |m_\phi| \to \infty$ is equivalent to taking $\ell \to \infty$ and $|m_\psi|, |m_\phi| = O(\ell)$; we will work in this limit for simplicity in keeping track of the orders of various terms. The next order term in the large ℓ expansion will only affect the ℓ independent piece of A, that is to say, at large[13] ℓ

$$A \approx \mu_\ell^2 + O(1)\,.$$

It turns out that we only need to know A up to this order in ℓ to know the leading behaviour of the imaginary part of the quasinormal modes in this sector of perturbations.

[13] This correction can be easily computed, but will not be needed in what follows. For the interested reader, when $\ell = |m_\phi| + |m_\psi|$

$$A \approx \mu_\ell^2 + \left(n\frac{m_\phi}{\ell} - (n+1)\frac{m_\psi}{\ell}\right)(\tilde{\omega}^2 - \tilde{\lambda}^2)\frac{a^2\eta}{R_z^2} + O(\ell^{-1})\,.$$

We now turn our attention to the radial equation. Unlike the angular equation, we cannot use standard perturbation theory to determine $\tilde{\omega}$. Instead, we have to resort to a matched asymptotic expansion.

The radial equation (2.55b) can be written as

$$-y(y^2+s^2\eta)\frac{\mathrm{d}}{\mathrm{d}y}\left[y(y^2+s^2\eta)\frac{\mathrm{d}\Phi_r}{\mathrm{d}y}\right]+V(y)\Phi_r(y)=0 \tag{2.61}$$

where we introduce the dimensionless variable $y=r/R_z$ and define

$$V(y)=-\tilde{\kappa}^2y^6+ay^4-by^2+c \tag{2.62}$$

where $a=\ell^2a_0+\ell a_1+O(1)$, $b=\ell^2b_0+\ell b_1+O(1)$ and $c=\ell^2c_0+\ell c_1+O(1)$,

$$\begin{aligned}
&a_0=1,\ \ a_1=2\\
&b_0=-s^2\eta+\frac{m_\phi^2}{\ell^2}(2n+1)(1-j^2)s^2\eta,\\
&b_1=-2s^2\eta+2\eta s^2\frac{m_\phi}{l}\left(\frac{\tilde{\omega}}{\eta}-\tilde{\lambda}\frac{Q_p(Q_1+Q_2)}{Q_1Q_2}\right)(nj-(n+1))-2\tilde{\lambda}s^2\eta\frac{m_\phi}{l}\left(j(n+1)-n\right)\\
&c_0=s^4\eta^2\frac{m_\phi^2}{\ell^2}(j(n+1)-n)^2\ \text{ and }\ c_1=2s^4\eta^2\tilde{\lambda}\frac{m_\phi}{\ell}\left(j(n+1)-n\right).
\end{aligned} \tag{2.63}$$

For later use, we also define

$$j\equiv\frac{m_\psi}{m_\phi},\ \ m\equiv\frac{m_\phi}{\ell}\ \Rightarrow|m|\leq\frac{1}{1+|j|}. \tag{2.64}$$

The wave equation is invariant under complex conjugation and so we have an overall choice of sign in the exponent in (2.54). Geodesics with tangent vector V on $\mathscr{S}$ have $p_\phi<0$ so we will fix the sign by assuming $m<0$.

To calculate the frequencies of quasinormal modes we find solutions of (2.61) obeying the necessary boundary conditions in the limit $\ell\to\infty$. We use an asymptotic matching procedure with $\ell\to\infty$ a large parameter, similar to that used in ref. [128] for the decoupling limit of non-supersymmetric 3-charge microstate geometries.

Note that $\{a,b,c\}=O(\ell^2)$ but $\tilde{\kappa}=O(1)$ so that we can split the $y-$axis into 3 regions, approximate the potential $V(y)$ and then solve the remaining equation exactly in each region. The regions and approximations of the potential are as follows:

1) $y\ll\sqrt{\ell}$: $V(y)\approx ay^4-by^2+c$

2) $1 \ll y \ll \ell$: $V(y) \approx ay^4$

3) $y \gg \sqrt{\ell}$: $V(y) \approx -\tilde{\kappa}^2 y^6 + ay^4$.

Since region 2 overlaps with both regions 1 and 3 we can find solutions in each of the regions then match them where they overlap. We will label the solution of Φ_r in each of the regions by Φ_i, where i indexes the region in question.

2.4.2.1 Region 1: $y \ll \sqrt{\ell}$

We approximate the equation by

$$y(y^2+s^2\eta)\frac{\mathrm{d}}{\mathrm{d}y}\left[y(y^2+s^2\eta)\frac{d\Phi_1}{dy}\right] - (ay^4 - by^2 + c)\Phi_1(y) = 0. \tag{2.65}$$

To make the expressions more compact, we define

$$\alpha \equiv s\sqrt{\eta},\ \ \beta \equiv \sqrt{a + \frac{c}{\alpha^4} + \frac{b}{\alpha^2}},\ \ \nu \equiv \sqrt{1+a} = \ell + 1 + O(\ell^{-1}). \tag{2.66}$$

Eq. (2.65) can be brought to a more familiar form by a suitable change of variable. We define

$$\Phi_1(y) = y^{\frac{\sqrt{c}}{\alpha^2}}(y^2+\alpha^2)^{\frac{\beta}{2}} Q\left(-\frac{y^2}{\alpha^2}\right),$$

where we implicitly have changed to a new coordinate $\tilde{z} = -y^2/\alpha^2$. The resulting equation for $Q(\tilde{z})$ is that of a Gaussian hypergeometric function of the second kind, ${}_2F_1(\tilde{a},\tilde{b},\tilde{c},\tilde{z})$ with

$$\tilde{a} = \frac{1}{2}\left(1 - \nu + \beta + \frac{\sqrt{c}}{\alpha^2}\right), \quad \tilde{b} = \frac{1}{2}\left(1 + \nu + \beta + \frac{\sqrt{c}}{\alpha^2}\right) \quad \text{and} \quad \tilde{c} = 1 + \frac{\sqrt{c}}{\alpha^2}.$$

Our boundary conditions demand that we choose the regular Gaussian hypergeometric function at $\tilde{z} = y = 0$. Our final solution, in this region of the potential, can simply be written as

$$\Phi_1(y) = A_1 y^{\frac{\sqrt{c}}{\alpha^2}}(y^2+\alpha^2)^{\frac{\beta}{2}}\, {}_2F_1\left(\tilde{a},\tilde{b},\tilde{c},-\frac{y^2}{\alpha^2}\right). \tag{2.67}$$

where A_1 is a constant.

Boundary conditions

To see why this solution is indeed smooth at $r=0$ we need to use coordinates that are themselves regular there. Such coordinates can be found in [91]: let

$$\phi \to \tilde{\phi} + \frac{an}{\sqrt{Q_1 Q_2}} z, \quad \psi \to \tilde{\psi} - \frac{a(n+1)}{\sqrt{Q_1 Q_2}} z, \quad z \to \tilde{z} = \frac{az}{\sqrt{Q_1 Q_2}} \tag{2.68}$$

(which are all $2\pi-$periodic). In the limit $r \to 0$, in the new coordinates, the metric takes the form [91]

$$ds^2 = \frac{hf}{\eta a^2}\left(dr^2 + r^2 d\tilde{z}^2\right) + \bar{g}_{\alpha\beta} d\bar{x}^\alpha d\bar{x}^\beta + ds^2_{T^4}$$

where $hf = \eta a^2\left((n+1)\cos^2\theta - n\sin^2\theta\right)$, $\bar{x}^\alpha \in \{t, \theta, \tilde{\phi}, \tilde{\psi}\}$ and $\bar{g}_{\alpha,\beta}$ are differentiable functions of r and θ that are non-zero in the limit $r \to 0$.

In these coordinates the solution (2.67) takes the form

$$\Phi(t, z, r, \theta, \phi, \psi) \sim e^{-i\omega t + i m_\psi \tilde{\psi} + i m_\phi \tilde{\phi}} (r e^{\pm i\tilde{z}})^{|(n+1)m_\psi - n m_\phi|} F_r(r) \Phi_\theta(\theta). \tag{2.69}$$

where

$$F_r(r) = (r^2/R_z^2 + \alpha^2)^{\frac{\beta}{2}} {}_2F_1\left(\tilde{a}, \tilde{b}, \tilde{c}, -\frac{r^2}{R_z^2 \alpha^2}\right)$$

is a regular function of r^2 as $r \to 0$. We have used that $\lambda = 0$ and $\sqrt{c}/\alpha^2 = |(n+1)m_\psi - n m_\phi|$ at leading order, while the $\pm$ comes from the sign of $\left((n+1)m_\psi - n m_\phi\right)$.

If we were then to change coordinates to $\hat{x} = r\cos\tilde{z}$, $\hat{y} = r\sin\tilde{z}$ we would see that (2.69) is indeed regular at $r = 0$, and hence we should choose the regular hypergeometric function as stated above.

To match this solution in region 1 to region 2 take the limit $y \to \infty$ ($\ell \to \infty$ and the overlap region is $1 \ll y \ll \sqrt{\ell}$, so we can have for example $y \approx O(\ell^{\frac{1}{4}})$):

$$\Phi_1(y) \approx A_1 \Gamma(1 + \frac{\sqrt{c}}{\alpha^2}) \alpha^{\frac{\sqrt{c}}{\alpha^2} + \frac{1}{2} + \frac{\beta}{2}} \left[\alpha^{\frac{\nu}{2}} y^{-\nu-1} \frac{\Gamma(-\nu)}{\Gamma(\tilde{c}-\tilde{b})\Gamma(\tilde{a})} + \alpha^{-\frac{\nu}{2}} y^{\nu-1} \frac{\Gamma(\nu)}{\Gamma(\tilde{c}-\tilde{a})\Gamma(\tilde{b})}\right]. \tag{2.70}$$

2.4.2.2 Region 2: $1 \ll y \ll \ell$

In this region the equation is approximated by

$$y^3 \frac{d}{dy}\left(y^3 \frac{d\Phi_2}{dy}\right) - a y^4 \Phi_2(y) = 0 \tag{2.71}$$

since $s^2\eta \ll y^2$. This has solution

$$\Phi_2(y) = B_1 y^{-\nu-1} + B_2 y^{\nu-1} \tag{2.72}$$

where B_1, B_2 are constants.

Matching (2.70) to (2.72) in the overlapping region gives the condition:

$$\frac{B_1}{B_2} = \alpha^\nu \frac{\Gamma(-\nu)}{\Gamma(\nu)} \frac{\Gamma(\tilde{c}-\tilde{a})\Gamma(\tilde{b})}{\Gamma(\tilde{c}-\tilde{b})\Gamma(\tilde{a})}. \tag{2.73}$$

2.4.2.3 Region 3: $y \gg \sqrt{\ell}$

In this region at highest order in ℓ,

$$y^3 \frac{d}{dy}\left(y^3 \frac{d\Phi_3}{dy}\right) - \left(-\tilde{\kappa}^2 y^6 + a y^4\right)\Phi_3 = 0 \tag{2.74}$$

with solution

$$\Phi_3(y) = \frac{1}{y}\left(C_1 J_\nu(\tilde{\kappa} y) + C_2 Y_\nu(\tilde{\kappa} y)\right) \tag{2.75}$$

where C_1, C_2 are constants and $J_\nu(x), Y_\nu(x)$ are Bessel functions of the first and second kind respectively.

In the asymptotic region as $y \to \infty$,

$$\Phi_3(y) = \frac{1}{y^{\frac{3}{2}}} \frac{1}{\sqrt{\tilde{\kappa}\pi}} \left[e^{i\tilde{\kappa}y} e^{-i\frac{\nu\pi}{2}} \left(\frac{1}{2} - \frac{i}{2}\right)(C_1 - iC_2) + e^{-i\tilde{\kappa}y} e^{i\frac{\nu\pi}{2}} \left(\frac{1}{2} + \frac{i}{2}\right)(C_1 + iC_2) \right] + O(y^{-\frac{5}{2}}). \tag{2.76}$$

Imposing the boundary condition that there are only outgoing waves at infinity gives

$$C_1 + iC_2 = 0. \tag{2.77}$$

To match to Region 2 in the overlap region $\sqrt{\ell} \ll y \ll \ell$ we take $\tilde{\kappa} y \ll \ell$ while $\nu \to \infty$. Using the formulae for the asymptotic form of the Bessel functions at large orders [129]

gives:

$$\Phi_3 = \left[C_1 y^{\nu-1} \left(\frac{\tilde{\kappa}}{2}\right)^{\nu} \frac{1}{\sqrt{2\pi\nu}} \frac{e^{\nu}}{\nu^{\nu}} - C_2 y^{-\nu-1} \left(\frac{\tilde{\kappa}}{2}\right)^{-\nu} \sqrt{\frac{2}{\pi\nu}} \frac{e^{-\nu}}{\nu^{-\nu}} \right] \left[1 + O(\ell^{-1})\right] \tag{2.78}$$

and so we find

$$\frac{C_1}{C_2} = -2 \left(\frac{\tilde{\kappa}}{2}\right)^{-2\nu} e^{2\nu} \nu^{-2\nu} \frac{B_2}{B_1}. \tag{2.79}$$

2.4.2.4 Real part of the frequency

The conditions (2.73), (2.77) and (2.79) all together imply that the quasinormal mode frequencies $\tilde{\omega}$ are solutions of the equation

$$\alpha^{-\nu} \frac{\Gamma(\tilde{c}-\tilde{b})\Gamma(\tilde{a})}{\Gamma(\tilde{c}-\tilde{a})\Gamma(\tilde{b})} = 2i \left(\frac{\tilde{\kappa}}{2}\right)^{2\nu} \frac{\Gamma(-\nu)}{\Gamma(\nu)} e^{2\nu} \nu^{-2\nu}. \tag{2.80}$$

We have that $\nu = \sqrt{1+a} = O(\ell) \gg 1$ so the RHS is extremely small; the only way to solve (2.80) is to have a pole in one of the $\Gamma-$functions in the denominator of the LHS *i.e.*

$$(\tilde{c}-\tilde{a} = -N \vee \tilde{b} = -N) \Rightarrow \frac{1}{2}\left(1 + \nu \pm \beta + \frac{\sqrt{c}}{\alpha^2}\right) = -N. \tag{2.81}$$

The leading order dependence on $\tilde{\omega}$ in (2.81) comes from

$$\beta = \ell |m(jn-(n+1))| + \frac{|m(nj-(n+1))|}{m(nj-(n+1))} \left(\frac{\tilde{\omega}}{\eta} - \tilde{\lambda} \frac{Q_p(Q_1+Q_2)}{Q_1 Q_2}\right) + O(\ell^{-1}). \tag{2.82}$$

From the condition that $|\tilde{\omega}|, |\tilde{\lambda}| \ll \ell$, all the terms that are proportional to ℓ in (2.81) must cancel:

$$1 + |m(j(n+1)-n)| \pm |m(jn-(n+1))| = 0. \tag{2.83}$$

Clearly, this condition does not hold for general values of m and j, and so we will use (2.83) to find possible values for m in terms of j for which there are quasinormal modes with $|\tilde{\omega}|, |\tilde{\lambda}| \ll \ell$. By examining (2.83) we see that it can only be solved if we choose the minus sign (otherwise all terms on the left hand side are positive). The equation remains non-trivial. We will use geometric optics to help us find a solution.

In geometric optics, $j = p_\psi / p_\phi$ to leading order in ℓ. In section 2.3.2 we found that the zero energy geodesics with tangent vector V have:

$$0 \leq \frac{p_\psi}{p_\phi} \leq \frac{n}{n+1} \quad \text{for} \quad n > 0, \qquad \frac{p_\psi}{p_\phi} \geq \frac{n}{n+1} \quad \text{for} \quad n < 0 \tag{2.84}$$

This suggests that we look for a solution of (2.83) with

$$0 \leq j \leq \frac{n}{n+1} \quad \text{for} \quad n > 0, \qquad j \geq \frac{n}{n+1} \quad \text{for} \quad n < 0 \tag{2.85}$$

In both cases we have $j \geq 0$ and $(n+1)j - n \leq 0$, and these imply $nj - (n+1) < 0$. Using these, along with $m < 0$, equation (2.83) reduces to

$$m = -\frac{1}{1+j} \tag{2.86}$$

which is equivalent to

$$\ell = -m_\phi - m_\psi \tag{2.87}$$

So in summary, we have found values of ℓ, m_ϕ, m_ψ that are consistent with our assumptions by taking $m_\phi, m_\psi < 0$ and $j = m_\psi/m_\phi$ in the range (2.85), with ℓ given by (2.87). Substituting these values into (2.81), the real part of $\tilde{\omega}$ at leading order is

$$\tilde{\omega}_R = 2\eta(N+1) + \tilde{\lambda}. \tag{2.88}$$

The expression (2.88) for $\tilde{\omega}_R$ is remarkably simple. As a check on this formula we can take the decoupling limit $Q_p \ll \sqrt{Q_1 Q_2} \ll R_z^2$, which gives $\eta \to 1$, in (2.88). In this limit the geometry reduces to $\mathrm{AdS}_3 \times S^3$ and our expression for $\tilde{\omega}_R$ reduces to the formula for certain normal modes in $\mathrm{AdS}_3 \times S^3$, see e.g. Eq. (6.12) of [130].[14]

2.4.2.5 Imaginary part of the frequency

To find the imaginary part of the frequency we look at the next order terms in (2.80) by substituting $\tilde{\omega} = \tilde{\omega}_R + \delta\tilde{\omega}$. Then $\beta = \beta(\tilde{\omega}_R) + \delta\beta$ where $\delta\beta = \frac{\delta\tilde{\omega}}{\eta}$ and we substitute

$$\Gamma(-N - \frac{\delta\beta}{2}) = \frac{(-1)^{N+1}}{N!}\frac{2}{\delta\beta}(1 + O(\delta\beta)) \tag{2.89}$$

in the left hand side of (2.80), which is the only term that depends on $\delta\beta$ at highest order. We also use the well known identities

$$\Gamma(-\nu) = -\frac{\pi}{\nu \sin \pi\nu}\frac{1}{\Gamma(\nu)}, \quad \text{and} \quad \Gamma(-N-\nu) = \frac{(-1)^{N+1}\pi}{(N+\nu)\sin\pi\nu}\frac{1}{\Gamma(N+\nu)}.$$

[14] A similarly simple expression was found for the real part of the frequencies of *unstable* modes in the non-supersymmetric 3-charge geometries in the decoupling limit in [128] although in that case the real part of the frequency scales as ℓ in general.

Substituting these into (2.80) and rearranging:

$$\delta\beta = -i\left(\frac{\tilde{\kappa}}{2}\right)^{2\nu}\alpha^{\nu}\frac{4(N+\nu)\Gamma(N+1+\nu+\frac{\sqrt{c}}{\alpha^2})}{N!\nu\Gamma(N+1+\frac{\sqrt{c}}{\alpha^2})}\frac{\Gamma(N+\nu)}{\Gamma(\nu)^2}e^{-2\nu\log\nu+2\nu}. \tag{2.90}$$

The size of the corrections to the real part of the frequency $\tilde{\omega}_R$ from Eq. (2.81) are of order $O(\ell^{-1})$ and are thus much larger than the corrections to $\tilde{\omega}$ here. However, the corrections to $\tilde{\omega}$ in (2.81) will all be real (all the coefficients are real apart from dependence on $\tilde{\omega}$) and so the imaginary part of the frequency does not have any terms that are proportional to inverse powers of ℓ. We therefore use (2.90) to find the imaginary part of $\tilde{\omega}$ at leading order and we in fact have $\delta\tilde{\omega} = \delta\tilde{\omega}_R + i\tilde{\omega}_I$. Substituting this in to (2.90), we find

$$\tilde{\omega}_I = -\eta\left(\frac{\tilde{\kappa}}{2}\right)^{2\nu}\alpha^{\nu}\frac{4(N+\nu)\Gamma(N+1+\nu+\frac{\sqrt{c}}{\alpha^2})}{N!\nu\Gamma(N+1+\frac{\sqrt{c}}{\alpha^2})}\frac{\Gamma(N+\nu)}{\Gamma(\nu)^2}e^{-2\nu\log\nu+2\nu}. \tag{2.91}$$

Define $\mu = -\frac{j(n+1)-n}{1+j} > 0$, then use $\ell \gg 1$ in (2.91) gives

$$\tilde{\omega}_I = -D\eta\,\alpha\,\tilde{\kappa}_0^2 e^{-2\ell\log\ell+\ell\left[2-\mu\log\mu+(1+\mu)\log(\mu+1)+2\log\frac{\tilde{\kappa}_0\sqrt{\alpha}}{2}\right]+(N-\frac{3}{2})\log\ell+O(1)} \tag{2.92}$$

where $\tilde{\kappa}_0 = \sqrt{\tilde{\omega}_{R,0}^2 - \tilde{\lambda}^2}$, $\tilde{\omega}_{R,0}$ is the real part of $\tilde{\omega}$ calculated to leading order only (i.e. $\tilde{\omega}_R$ in (2.88)). D is a positive constant that is independent of ℓ at leading order but depends on the higher order corrections to the real part of $\tilde{\omega}$ from the term $\tilde{\kappa}^{2(\ell+1)}$ in (2.91).

Equation (2.92) is one of our main results. We see that $\tilde{\omega}_I < 0$ so the waves decay as expected. However, the rate of decay is very slow, since at leading order the term that controls it is $e^{-2\ell\log\ell}$ which is very small for large ℓ.

As discussed above, in the decoupling limit we expect our quasinormal modes to reduce to normal modes in $\mathrm{AdS}_3\times\mathrm{S}^3$ so $\tilde{\omega}_I$ should vanish in this limit. This is indeed the case because $\alpha \to 0$ in the decoupling limit.

The calculation above assumes $n \neq 0$, i.e., $Q_p \neq 0$ so it does not apply to 2-charge microstate geometries. When $n = 0$, $\mathscr{S}$ becomes the 2-dimensional submanifold $r = 0$, $\theta = \pi/2$. In section 2.4.4 we show that it is straightforward to modify the above calculation to cover this case too. The result is the same, i.e, $\tilde{\omega}_I$ is $\mathscr{O}(e^{-2\ell\log\ell})$ at large ℓ. Hence the dimension of $\mathscr{S}$ does not seem to affect the slow decay, which is to be expected since the slowly decaying quasinormal modes are associated to individual null geodesics on $\mathscr{S}$ rather than to global properties of $\mathscr{S}$.

2.4.3 Kaluza-Klein momentum scaling with ℓ

In section 2.3.2 we saw that at every point in the six-dimensional spacetime there is a stably trapped geodesic with tangent V. We have found quasinormal modes that correspond to the zero energy null geodesics that are trapped near $\mathscr{S}$ but we also expect to be able to find slowly decaying modes that are localised near geodesics that are trapped elsewhere in the spacetime. These geodesics have tangent V and conserved quantities $p_z = -p_t$. Under the geometric optics approximation we expect that the corresponding solutions of the wave equation will have $\tilde{\omega} \approx \tilde{\lambda}$. We will now consider $\tilde{\lambda} = \mathscr{O}(\ell)$ but keep the difference $|\tilde{\omega} - \tilde{\lambda}| = O(1)$ in the limit $|m_\psi|, |m_\phi| \to \infty$. In this case, $\tilde{\kappa}^2 = (\tilde{\omega} - \tilde{\lambda})(\tilde{\omega} + \tilde{\lambda}) = O(|m_\psi|, |m_\phi|)$.

Since $\tilde{\kappa}^2 \ll m_\phi^2,\ m_\psi^2$, we can ignore the $\tilde{\kappa}^2$ in the angular equation (2.55a) at leading order in m_ϕ, m_ψ. This means that we have

$$A \approx \ell^2 + A_1 \ell + O(1)$$

with ℓ defined previously in (2.60). If we set $\ell = |m_\phi| + |m_\psi|$, *i.e.* $m = -1/(1+j)$, we can find A_1 using standard perturbation theory. It turns out that

$$A_1 = 2 - 2\frac{\tilde{\lambda}\alpha^2}{\ell}(\tilde{\omega} - \tilde{\lambda})\left(\frac{n - (n+1)j}{1+j}\right) \tag{2.93}$$

We will find later that we must have $m = -1/(1+j)$ to have modes $|\tilde{\omega} - \tilde{\lambda}| = O(1)$ so this assumption is consistent.

The expressions for $a,\ b,\ c$ in (2.61) at the various orders change: we now have

$$\begin{aligned} a &= \tilde{\nu}^2 - 1 - \tilde{\kappa}^2\alpha^2 \\ b &= \alpha^2(1 - \tilde{\nu}^2) + s^2(\xi^2 - \zeta^2) \\ c &= \alpha^2 s^2 \zeta^2 \end{aligned} \tag{2.94}$$

where $\tilde{\nu},\ \xi,\ \zeta$ are defined in (2.59).

2.4.3.1 Asymptotic matching

The asymptotic matching procedure in 2.4.2 only needs to be slightly modified to find solutions with frequencies with $\tilde{\kappa}^2 = O(\ell)$. Regions 1, 2 and 3 must be changed so that the potentials can be approximated in the same way as before in each region.

We define the new regions as:

1') $y \ll \ell^{\frac{1}{4}}$: $V(y) \approx ay^4 - by^2 + c$

2') $1 \ll y \ll \sqrt{\ell}$: $V(y) \approx ay^4$;

3') $y \gg \ell^{\frac{1}{4}}$: $V(y) \approx -\tilde{\kappa}^2 y^6 + ay^4$.

Note that the regions still overlap so we can match the solutions in different regions.

Exactly the same matching procedure as in section 2.4.2 then follows through to give that the real part of the frequency is defined by the condition

$$\frac{1}{2}(1+\nu \pm \beta + \frac{\sqrt{c}}{\alpha^2}) = -N. \tag{2.95}$$

We expect $\tilde{\omega} - \tilde{\lambda}$ to be small so we must take the minus sign for the same reasons as in section 2.4.2.4. However, the leading order behaviour of β and ν differs to the previous case; we find that now

$$\begin{aligned}
\beta &= \tilde{\lambda} + m\ell[jn-(n+1)] + (\tilde{\omega}-\tilde{\lambda})\left\{\frac{1}{\eta} - \frac{\tilde{\lambda}\alpha^2}{\tilde{\lambda}+m\ell[jn-(n+1)]}\right\} + O(\ell^{-1}) \\
\frac{\sqrt{c}}{\alpha^2} &= \tilde{\lambda} + m\ell[j(n+1)-n], \\
\nu &= \ell + \frac{A_1}{2} - (\tilde{\omega}-\tilde{\lambda})\frac{\tilde{\lambda}}{\ell}\left(\frac{Q_1+Q_2}{R_z^2} + \alpha^2\right) + O(\ell^{-1}) = \ell + \nu_1 + O(\ell^{-1}).
\end{aligned} \tag{2.96}$$

We assume as before that $m<0,\ jn-(n+1)<0,\ j>0$ and $\tilde{\lambda} \geq 0$. Substituting this into equation (2.95) and imposing the condition $|\tilde{\omega}-\tilde{\lambda}| = O(1)$, we find that we must take $m = -1/(1+j)$ so that the higher order terms cancel. Then the real part of the frequency is given by (2.95):

$$\tilde{\omega}_R = \tilde{\lambda} + \frac{2\eta}{P}(N+1) + O(\ell^{-1}) \tag{2.97}$$

where we use the definitions of $a,\ b,\ c$ and $\tilde{\omega} = \tilde{\lambda} + O(1)$ to find

$$P = 1 + \frac{\tilde{\lambda}\alpha^2\eta}{\ell}\left(1 - \frac{\ell}{\tilde{\lambda}+m\ell[jn-(n+1)]}\right) + \frac{\tilde{\lambda}}{\ell}\eta\frac{Q_1+Q_2}{R_z^2} + \frac{\tilde{\lambda}\alpha^2\eta}{\ell}\left(\frac{n-j(n+1)}{1+j}\right) > 0. \tag{2.98}$$

It is interesting to see that if we take $\tilde{\lambda} \ll \ell$ in (2.97), although this limit does not apply here, we nevertheless recover the real part of the frequency for $\tilde{\lambda} = O(1)$ as given in eq. (2.88).

The calculation for the imaginary part is also very similar to that of section 2.4.2.5; we simply have to replace $\delta\beta$ with $P\delta\tilde{\omega}$ in (2.91). Then let

$$\mu' = \frac{\tilde{\lambda}}{\ell} - \frac{j(n+1)-n}{1+j} > 0.$$

In the limit $\ell \to \infty$, the imaginary part of the frequency at leading order is

$$\tilde{\omega}_I = -D' e^{-\ell \log \ell + \ell \left[2 - \mu' \log \mu' + (1+\mu') \log(\mu'+1) + 2 \log \frac{\tilde{\kappa}_0 \sqrt{\alpha}}{2\sqrt{\ell}}\right] + (N + \frac{1}{2} - \nu_1) \log \ell} + O(l^{-1}) \qquad (2.99)$$

for some positive constant D' that is independent of ℓ and where $\tilde{\kappa}_0^2 = 2\tilde{\lambda}(\tilde{\omega}_R - \tilde{\lambda})$ with $\tilde{\omega}_R$ evaluated using (2.97) and ν_1 is given in (2.96) with the terms $\tilde{\omega} - \tilde{\lambda}$ also evaluated at leading order using (2.97). D' is proportional to α^{ν_1}; in the decoupling limit $\nu_1 \to 1$ and $\alpha \to 0$ so we see that the imaginary part vanishes in this limit, as expected. The real part reduces to the expression for certain normal modes in $\mathrm{AdS}_3 \times S^3$, as given in [130].

We have constructed quasinormal modes with $\tilde{\omega}_I \sim -e^{-\ell \log \ell}$ at leading order for $\ell \gg 1$. We expect that such a mode will be localised near a stably trapped geodesics with tangent V, whose location is determined by the matching the ratios p_ψ/p_ϕ, p_z/p_ϕ to m_ψ/m_ϕ and λ/m_ϕ. Note that there is no longer a factor of 2 multiplying $-\ell \log \ell$ in the exponent so these modes decay faster than the modes localized near $\mathscr{S}$ that we found in the previous section. However, the decay is still very slow and therefore likely to be problematic for nonlinear stability.

The above calculation assumes $n \neq 0$, i.e., $Q_p \neq 0$ but in the next section 2.4.4 we show that it is straightforward to modify the calculation to cover the 2-charge case. The result is $\tilde{\omega}_I = \mathscr{O}(e^{-\ell \log \ell})$ as for the 3-charge case.

2.4.4 2-charge quasinormal modes

The wave equation separates in the 2-charge microstate geometries in the coordinates of (3.22) (see ref. [126]) in the same way as for the 3-charge geometry but with $n = 0$. In the wave equation we will therefore again use the ansatz

$$\Phi(t, z, r, \theta, \phi, \psi) = e^{-i\omega t + i\lambda z + i m_\psi \psi + i m_\phi \phi} \Phi_r(r) \Phi_\theta(\theta). \qquad (2.100)$$

However, if we are looking for modes that correspond, via the geometric optics approximation, to null geodesics with tangent vector V that are stably trapped on $\mathscr{S}_2$ we must set $m_\psi = 0$ because the corresponding geodesics are localized at $\theta = \pi/2$ so they have $p_\psi = 0$.

Ref. [126] discusses scattering solutions of the wave equation with low frequencies. Here we will find quasinormal modes with $|m_\phi| \gg 1$. As for the 3-charge case, we look specifically for solutions where $\tilde{\omega}, \tilde{\lambda} = O(1) \ll |m_\phi|$, motivated by the geometric optics approximation since the geodesics with tangent V on $\mathscr{S}_2$ have zero energy and Kaluza-Klein momentum.

2.4.4.1 2-charge matched asymptotic expansion

After separating variables, the equation for $\Phi_\theta(\theta)$ is exactly the same as (2.55a) with $m_\psi = 0$, $n = 0$ and $\eta = 1$. Note that $m_\psi = 0$ implies that $j = 0$ and that if we write (2.55a) in the form of a Schrödinger equation the potential is not strictly positive at $\theta = \pi/2$ on $\mathscr{S}_2$ so we have an 'allowed' region there.

Exactly as in section 2.4.2, from eq. (2.55a) the separation constant is $A = \ell(\ell+2) + O(1)$ where

$$\ell \geq |m_\phi|,\ \ell \in \mathbb{Z}. \tag{2.101}$$

We will construct quasinormal modes satisfying $\ell \gg 1$ and $|m_\phi| = O(\ell)$.

The differences to the calculation for the 3-charge case arise in the radial equation. We still have equation (2.61) for $\Psi_r(y)$ but there are important differences in the coefficients b and c:

$$\begin{gathered} b_0 = s^2(m^2-1),\ b_1 = -2s^2(1+\tilde{\omega}) \\ c_0 = 0 = c_1 \Rightarrow c = c_2 = \alpha^4 \tilde{\lambda}^2. \end{gathered} \tag{2.102}$$

From the calculation for the 3-charge case we expect that we will have to set $m = -1$; in this case $b_0 = 0$ and $b = O(\ell)$. When we define each region we will allow either $b_0 = 0$ or $b_0 \neq 0$ and use (assuming $\tilde{\kappa} = O(1)$):

1) $y \ll \ell^{\frac{1}{4}}$: $\tilde{\kappa}^2 y^6 \ll : V(y) \approx ay^4 - by^2 + c$

2) $1 \ll y \ll \ell$: $V(y) \approx ay^4$;

3) $y \gg \sqrt{\ell}$: $\ell^2(y^2+C) \ll \tilde{\kappa}^2 y^6$ and $V(y) \approx -\tilde{\kappa}^2 y^6 + ay^4$.

Although the regions themselves are slightly different to those used in the 3-charge case, region 2 still overlaps both regions 1 and 3 and we approximate the equation in the same way as before in each region.

Therefore the analysis of section 2.4.2 follows through in exactly the same way as before; the fact that $c = O(1)$ doesn't change anything in the method or matching and we reach the same conditions as in the 3-charge case.

First of all, substituting $j = 0$ and $n = 0$ into equation (2.83), the requirement that the frequencies do not scale with ℓ, implies that

$$m = -1 \tag{2.103}$$

as we anticipated so that we do indeed have $b = O(\ell)$.

For the real part of the frequency we substitute $\eta = 1$ into (2.88) (or substitute c_2 and the other necessary values into (2.81)) to find that at leading order

$$\tilde{\omega}_R = 2(N+1) + \tilde{\lambda}. \tag{2.104}$$

For the imaginary part of the frequency given in (2.92) we set $\mu = 0$ to find

$$\tilde{\omega}_I = -D_2 s \tilde{\kappa}_0^2 e^{-2\ell \log \ell + \ell\left(2 + 2\log \frac{\tilde{\kappa}_0\sqrt{\alpha}}{2}\right) + (N - \frac{3}{2})\log \ell + O(1)} \tag{2.105}$$

for some positive constant D_2 and $\tilde{\kappa}_0 = \sqrt{\tilde{\omega}_{R,0}^2 - \lambda^2}$ where $\tilde{\omega}_{R,0}$ is the real part of $\tilde{\omega}$ calculated to leading order only in (2.104).

In both the 2- and 3-charge geometries the imaginary part of ω is negative and $O(e^{-2\ell \log \ell})$ as $\ell \to \infty$ when $\tilde{\kappa} = O(1)$. Hence the dimension of the evanescent ergosurface does not seem to make a difference to the rate at which the modes decay.

2.4.4.2 2-charge quasinormal mode frequencies scaling with ℓ

The angular equation for the 2-charge case is exactly the same as in the 3-charge case, but we had to modify the calculation of section 2.4.2 because some of the coefficients in the potential for the radial equation were zero at leading order. However, if we now assume that $|\tilde{\omega} - \tilde{\lambda}| = O(1)$ but $\tilde{\lambda} = O(\ell)$ so that the frequency scales with ℓ, the coefficients in the potential are non-zero at leading order and the calculation for the quasinormal frequencies that scale with ℓ is exactly the same as in section 2.4.3.1.

To obtain the quasinormal modes for the 2-charge case from the 3-charge calculation we set $n = 0$. Previously we also had to set $j = 0$ because we were looking for quasinormal modes localised near null geodesics stably trapped on the evanescent ergosurface, $\mathscr{S}_2$. Now we want to find solutions of the wave equation localised near null geodesics that are stably trapped away from $\mathscr{S}_2$; these do not necessarily have $j = 0$. However, in the calculation of section 2.4.3.1 we assume that we still have $n - (n+1)j \geq 0$; for ease of calculation we will therefore still assume that $j = 0$ here so we are looking for solutions localised near $\theta = \pi/2$ but not on $\mathscr{S}_2$.

In this case we can simply substitute $n = 0$ and $j = 0$ into the results of section 2.4.3.1. We find the real and imaginary parts of the quasinormal frequencies from equations (2.97) and (2.99) respectively: at leading order

$$\tilde{\omega}_R = \tilde{\lambda} + \frac{2\eta}{P'}(N+1) + O(\ell^{-1}) \tag{2.106}$$

where

$$P' = 1 + \frac{\tilde{\lambda}\,\alpha^2}{\ell}\left(1 - \frac{\ell}{\tilde{\lambda}+\ell}\right) + \frac{\tilde{\lambda}}{\ell}\frac{Q_1+Q_2}{R_z^2}. \tag{2.107}$$

If we define

$$\mu'' = \frac{\tilde{\lambda}}{\ell} > 0$$

we find that the imaginary part in the limit $\ell \to \infty$ is

$$\tilde{\omega}_I = -D_2' e^{-\ell\log\ell+\ell\left[2-\mu''\log\mu''+(1+\mu'')\log(\mu''+1)+2\log\frac{\tilde{\kappa}_0\sqrt{\alpha}}{2\sqrt{\ell}}\right]+(N+\frac{1}{2}-\nu_1)\log\ell} + O(l^{-1}) \tag{2.108}$$

where ν_1 is independent of ℓ and defined in (2.96) with $n=0$ and $j=0$, D_2' is a constant proportional to s^{ν_1} that vanishes in the decoupling limit and $\tilde{\kappa}_0 = \sqrt{\tilde{\omega}_R^2 - \tilde{\lambda}^2}$ with $\tilde{\omega}_R$ defined in (2.106).

2.4.5 Numerical determination of quasinormal modes

2.4.5.1 Method

In the previous sections we have determined certain quasinormal modes in the limit of large quantum number ℓ, we now aim to determine the behaviour of the corresponding modes at finite ℓ numerically. In doing so, we can also understand the regime of validity of the approximation scheme detailed in our previous sections. For the sake of presentation, we will restrict ourselves to the case with $\tilde{\lambda} = 0$, *i.e.* modes that do not depend on the Kaluza-Klein momentum.

Our separation ansatz reads

$$\Phi(r,\theta) = X(\cos\theta)\,W\left(\frac{rR_z}{\sqrt{Q_1\,Q_2}}\right),$$

which yields the following pair of ordinary differential equations for $X(x)$ and $W(w)$ to be solved numerically:

$$\frac{1}{x}\frac{\mathrm{d}}{\mathrm{d}x}\left[x(1-x^2)\frac{\mathrm{d}}{\mathrm{d}x}X(x)\right] + \left\{A + \alpha_1\alpha_2\eta\,\tilde{\omega}^2\left[-n\left(1-x^2\right)+x^2(1+n)\right] - \frac{m_\psi^2}{x^2} - \frac{m_\phi^2}{1-x^2}\right\}X(x) = 0 \tag{2.109a}$$

$$\frac{1}{w}\frac{\mathrm{d}}{\mathrm{d}w}\left[w(w^2+\eta)\frac{\mathrm{d}}{\mathrm{d}w}W(w)\right]+ \left\{\tilde{\omega}^2\left[\alpha_1+\alpha_2+\alpha_1\alpha_2 n(1+n)+\alpha_1\alpha_2 w^2\right]-A-\frac{\eta\left[nm_\phi-(1+n)m_\psi\right]^2}{w^2}+ \frac{\eta\left[(\alpha_1+\alpha_2)n(1+n)\tilde{\omega}+\tilde{\omega}+nm_\psi-(1+n)m_\phi\right]^2}{w^2+\eta}\right\}W(w)=0\,, \quad (2.109b)$$

where we have changed variables to $x\equiv\cos\theta$ and $w\equiv rR_z/\sqrt{Q_1\,Q_2}$ and defined $Q_i=\alpha_i R_z^2$. Here, as in previous sections, A is a separation constant to be determined in what follows.

Before detailing our numerical method, we need to investigate the boundary conditions at the edges of our integration domain. Our equations have five real singular points (three for the angular equation, and two for the radial equation). For the angular equation (2.109a) these are $x=0$, $x=1$ and $x=\infty$. For the radial equation these are $w=0$ and $w=\infty$.

Let us start with the angular equation. Since our integration domain is $x\in(0,1)$, we only need to understand what happens at the singular points $x=0$ and $x=1$. A Frobenius expansion at $x=0$, yields the following behaviour

$$X\sim x^{\pm|m_\psi|}\left[1+\mathscr{O}(x)\right],$$

while at $x=1$ we find

$$X\sim(1-x)^{\pm\frac{|m_\phi|}{2}}\left[1+\mathscr{O}(1-x)\right].$$

In order to have a regular solution, we must choose the $+$ signs at both integration edges. To solve the problem numerically, we change to a new variable that relates to X in the following manner:

$$X=x^{|m_\psi|}(1-x^2)^{\frac{|m_\phi|}{2}}\widetilde{X},$$

and impose Robin boundary conditions for $\widetilde{X}$ at $x=0$ and $x=1$. These can be found by solving the equations for $\widetilde{X}$ in a Taylor expansion around the two singular points; at either point they give a constraint on some combination of $\widetilde{X}$ and its first derivative.

Next we address the radial equation. The singular point at $w=0$ is a regular singular point, and its behaviour can be extracted via a Frobenius expansion (similar to the angular equation):

$$W(w)\sim w^{\pm|n(m_\phi-m_\psi)-m_\psi|}\left[1+\mathscr{O}(w)\right],$$

again regularity demands keeping the $+$ sign only (see section 2.4.2.1 for why this results in smoothness at $r=0$). Finally, at $w=+\infty$, there is an essential singularity, which is to be expected since we want to impose outgoing boundary conditions there. The singular

behaviour can be easily extracted, and takes the following form

$$W(w) \sim \frac{e^{\pm i\sqrt{\alpha_1}\sqrt{\alpha_2}w\tilde{\omega}}}{w^{\frac{3}{2}}}\left[1+\mathscr{O}(w^{-1})\right].$$

Demanding outgoing boundary conditions yields demands choosing the $+$ sign. As we have done for the angular equation, we now change to a new variable that is more adequate for the numerical procedure. We chose the following:

$$W(w) = \frac{e^{i\sqrt{\alpha_1}\sqrt{\alpha_2}w\tilde{\omega}}}{w^{\frac{3}{2}}} w^{|n(m_\phi - m_\psi)-m_\psi|}\widetilde{W}(w).$$

Finally, since w is a non-compact coordinate, we do a further change of coordinates of the form

$$w = \frac{\tilde{w}\sqrt{2-\tilde{w}^2}}{1-\tilde{w}^2},$$

which maps $w=0$ to $\tilde{w}=0$ and $w=+\infty$ to $\tilde{w}=1$. Robin boundary conditions, involving $\widetilde{W}(\tilde{w})$ and its first derivative, at $\tilde{w}=0$ and $\tilde{w}=1$ can now be found by solving the respective equation for $\widetilde{W}(\tilde{w})$.

Our original system of equations (2.109) has been mapped to two equations for $\widetilde{X}(x)$ and $\widetilde{W}(\tilde{w})$, with two coupled eigenvalues $(\tilde{\omega}, A)$. In order to solve these, we use a Newton-Raphson routine which has been outlined in [131] for a similar problem. Regarding the implementation of the algorithm, the only nuance that is worth emphasising is that we had to work with arbitrary-precision arithmetic, since the magnitude of the imaginary part of our quasinormal modes can be as small as 10^{-170} (for an order of magnitude, this is more than the number of decimal places captured by octuple precision - 10^{-71}).

2.4.5.2 Results

We have varied parameters in our search, *i.e.* different values of n, α_i, but the results look qualitatively similar. We divide the types of quasinormal modes we find into two types: *i*) those for which ω_R does not scale with ℓ and *ii*) those for which ω_R does scale with ℓ. In this paper we will focus on type *i*) modes, which is the sector that is responsible for the slow decay of generic perturbations. As we have seen in section 2.4.2.4 (see discussion around Eq. (2.86)), the slow decay will only hold for modes satisfying $\ell = |m_\phi| + |m_\psi|$, which are the modes we are going to focus on.

For the sake of presentation, we will keep $\alpha_1 = \alpha_2 = 1 = n = 1$. Changing α_1 or α_2 will just change the regime at which the matched asymptotic expansion analysis settles in. The

larger α_1 or α_2, the larger the value of ℓ we need to reach in order to see matching with the matched asymptotic expansion analysis of the previous sections.

In Fig. 2.2 we show a linear plot (left panel) of the real part of $\tilde{\omega}$ as a function of $m_\phi < 0$ for $m_\psi = -1$. We see that as $|m_\phi|$ increases, $\tilde{\omega}_R$ approaches the value predicted in Eq. (2.88). The approach to this value (solid red curve) can also be determined via the explicit construction of "quasimodes", which we detail in section 2.5. On the right panel of the same figure, we show a log-log plot of the imaginary part of $\tilde{\omega}$ as a function of $|m_\phi|$: the blue dots are the numerical data, whereas the red dashed curve is a one parameter fit to (2.92), with D, the overall scale, being the fitting parameter.

The agreement of the fit with the numerical data is very reassuring. In fact, the agreement is much better than one might have expected: our analytical result (2.92) works well down to small values of ℓ whereas this result was only expected to hold for $\ell \gg 1$. Note that the imaginary part of $\tilde{\omega}$ is very small even for small values of ℓ. So there exist very slowly decaying quasinormal modes even at small ℓ. The decay becomes even slower at high ℓ, in agreement with our analytical result.

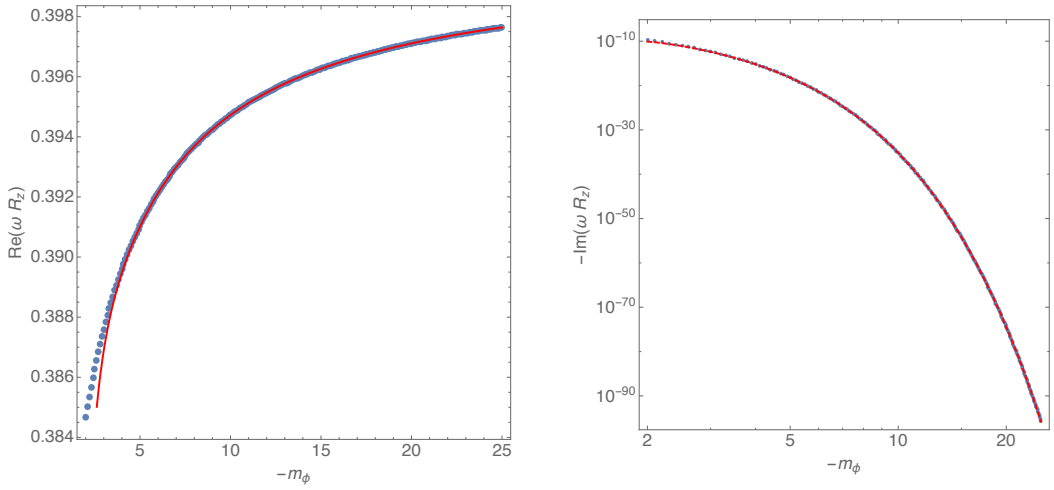


Fig. 2.2 *Left panel*: real part of $\tilde{\omega}$ as a function of $m_\phi < 0$. *Right panel*: imaginary part of $\tilde{\omega}$ as a function of m_ϕ. In both panels, the blue points are the numerical data, the solid red line is the analytic prediction for $\mathrm{Re}(\tilde{\omega})$ based on a quasimode construction (see Section 2.5), the dashed red line is the fit to (2.92), and both plots were generated with $\alpha_1 = \alpha_2 = n = -m_\psi = 1$.

Quasinormal modes grow exponentially at spatial infinity but they are well behaved at future null infinity. We can consider the behaviour of quasinormal modes on a surface of constant retarded time $u = t - r$, which extends to future null infinity. In Fig. 2.3 we plot the absolute value of the quasinormal mode as a function of w and x on such a surface for the smallest and largest value of m_ϕ we studied. The idea is to see if the quasinormal mode is localized near the corresponding null geodesic on $\mathscr{S}$, i.e., the geodesic with $p_\psi/p_\phi = m_\psi/m_\phi$

(represented in Fig. 2.3 by a black dot). We see that as m_ϕ increases, the maximum moves towards $x = 0$, as a consequence of the fact that m_ϕ is increasing, while m_ψ is kept constant, so the ratio m_ψ/m_ϕ decreases. Furthermore, the quasinormal mode localises more sharply around the geodesic prediction, as expected from geometric optics because $\ell = |m_\phi| + |m_\psi|$ is increasing.

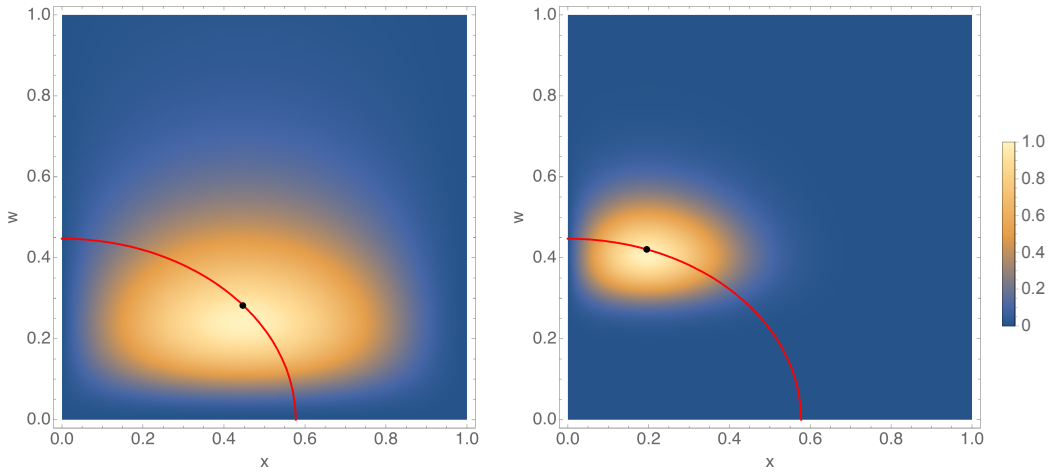


Fig. 2.3 Contour plot for $|\Phi|$ as a function of w and x on a surface extending to future null infinity. The red curve is the evanescent ergosurface $\mathscr{S}$. On the *left panel* we have $m_\phi = -4$ and on the *right panel* we have $m_\phi = -25$. Both panels were generated with $\alpha_1 = \alpha_2 = n = -m_\psi = 1$ and the normalization is $\max|\Phi| = 1$.

We have also considered a case in which both m_ϕ and m_ψ are *simultaneously* increasing with ℓ, while their ratio is kept fixed. In Fig. 2.4, we use $m_\phi = 4m_\psi$, and increase m_ϕ, with $\ell = |m_\phi| + |m_\psi|$. Since both m_ϕ and m_ψ are increasing, we expect the matched asymptotic expansion analysis to give a better approximation. We indeed see that this is the case: for $m_\psi = -1$ and $m_\phi = -4$, the matched asymptotic expansion result is barely discernible from the numerical data. Note that the colour coding in Fig. 2.4 is the same as in Fig. 2.2.

In Fig. 2.5, we plot the normalised quasinormal mode as a function of w and x, for the case $m_\phi = 4m_\psi$. As before, its peak is located exactly at the point predicted in section 2.2. Furthermore, the peak gets more and more sharp as we increase $\ell = |m_\phi| + |m_\psi|$.

One can go further, and determine the width of quasinormal mode around its maximum. It turns out to scale as $\sqrt{\ell}$, as expected from geometrical optics. This is best observed in Fig. 2.6, where we plot the contour lines of $|\Phi| = 1/5$, for several values of m_ϕ. The arrow in the plot indicates the direction of increasing $(-m_\phi)$, and the point in the middle indicates the geometric optics prediction for the location of the maximum of $|\Phi|$.

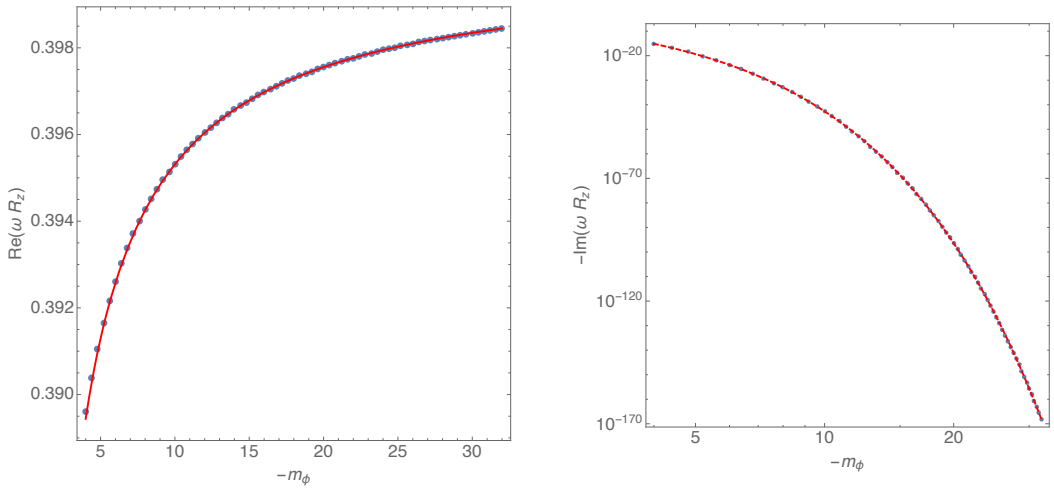


Fig. 2.4 *Left panel*: real part of $\tilde{\omega}$ as a function of $m_\phi < 0$. *Right panel*: imaginary part of $\tilde{\omega}$ as a function of m_ϕ. In both panels, the blue points are the numerical data, the dashed red line is the fit to (2.92), and both plots were generated with $\alpha_1 = \alpha_2 = n = 1$, with $m_\phi = 4m_\psi$.

2.4.6 Lower bound on decay rate

Proofs of nonlinear stability usually require first establishing *uniform* decay for linear perturbations. The first step is to establish decay of some *non-degenerate* energy. We consider some spacelike Cauchy surface Σ_0 and let Σ_t denote the surface obtained by translation of Σ_0 through parameter distance t w.r.t. the Killing field V. A non-degenerate energy $E_1(t)$ is an integral over Σ_t of some quantity quadratic in $\partial\Phi$, such that $E_1(t)$ is positive definite. Note that the *conserved* energy does not have this property because it degenerates on the evanescent ergosurface.

Ideally one would like to establish a quantitative uniform energy decay result of the form

$$E_1(t) \leq g(t)E_1(0) \tag{2.110}$$

for some function $g(t)$, independent of Φ, with $g(t) \to 0$ as $t \to \infty$. This is uniform because it applies to *any* perturbation Φ with g independent of the perturbation. If $g(t)$ decays fast enough (e.g. t^{-p} for large enough p) then one can hope to establish non-linear stability. However, when trapping is present, it is known that a decay result of this form does not exist [31]. Instead the best one can hope for is energy decay with "loss of a derivative", which means that one has

$$E_1(t) \leq g(t)E_2(0) \tag{2.111}$$

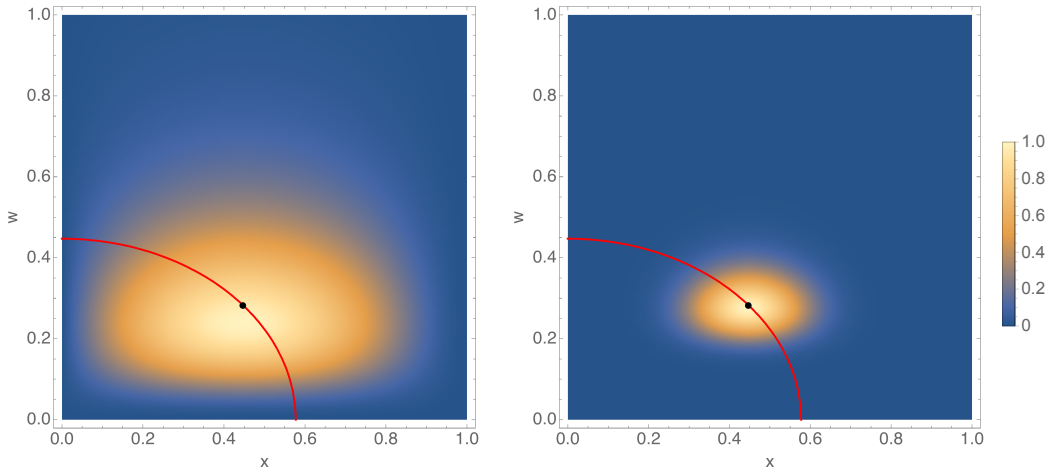


Fig. 2.5 Contour plot for $|\Phi|$ as a function of w and x: on the *left panel* we have $m_\phi = -4$ and on the *right panel* we have $m_\phi = -32$. Both panels were generated with $\alpha_1 = \alpha_2 = n = 1$ and $m_\phi = 4m_\psi$.

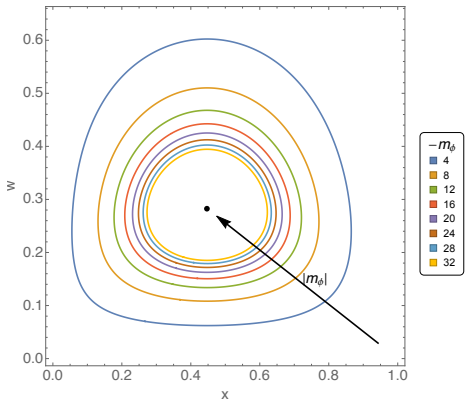


Fig. 2.6 Contour lines for $|\Phi| = 1/4$ at fixed $u = t - r$, as a function of w and x. All curves were generated with $\alpha_1 = \alpha_2 = n = 1$ and $m_\phi = 4m_\psi$.

where $E_2(t)$ is a *second order* energy, i.e., a positive functional of $\partial\Phi$ *and* $\partial^2\Phi$ defined as an integral over Σ_t. For example, the Schwarzschild solution, which exhibits unstable trapping at the photon sphere, admits a result of the above form with $g(t) \propto t^{-2}$ [119].

Energy-decay results of the above form have also been obtained for spacetimes with stable trapping, but the function $g(t)$ decays very slowly. For AdS black holes [32], and also for ultracompact neutron stars [33], results of the form (2.111) have been proved with $g(t) = (\log(2+t))^{-2}$. Moreover, in both of these examples, this result is sharp in the sense that if one picks $g(t)$ decaying faster than this then one can construct solutions which violate (2.111). In both cases, one can also obtain pointwise decay results for the field Φ.

We can now use our quasinormal modes to show that the decay is evern slower for the supersymmetric microstate geometries studied above. Quasinormal modes do not have finite

energy when defined on a surface of constant t in the coordinates of (2.36). This is because such modes diverge at spatial infinity. However, it is well known that quasinormal modes are finite at future null infinity. Therefore we will pick our Cauchy surfaces Σ_0 to extend to future null infinity.

Now consider a quasinormal mode with large ℓ. Since E_2 is quadratic in second derivatives of Φ, we expect $E_2(0) < C\ell^4$ for some $C > 0$ independent of ℓ. Hence if (2.111) holds we must have $E_1(t) < C\ell^4 g(t)$. On the LHS we have

$$E_1(t) \sim \ell^2 e^{2\omega_I t} \tag{2.112}$$

where $\omega = \omega_R + i\omega_I$. The factor of ℓ^2 comes from the fact that E_1 is quadratic in first derivatives of Φ. More precisely, we can find some constant $D > 0$, independent of ℓ, such that

$$E_1(t) > D\ell^2 e^{2\omega_I t} \tag{2.113}$$

hence if (2.111) holds then we must have

$$De^{2\omega_I t} < C\ell^2 g(t) \tag{2.114}$$

For example, consider $g(t) = (\log(2+t))^{-2}$ as for the examples discussed above. Set $t = e^{\tau\ell}$ for some $\tau > 0$. Then we need (using our result for ω_I)

$$D\exp(-2\beta e^{-2\ell\log\ell} e^{\tau\ell}) \lesssim \frac{C}{\tau^2} \tag{2.115}$$

where $\beta > 0$ is the coefficient in our large ℓ expression for ω_I derived above. Now taking the limit $\ell \to \infty$ gives $D \lesssim C/\tau^2$, which we can violate by taking τ large enough. This proves that a uniform decay result of the form (2.111) *cannot* exist with $g(t) = (\log(2+t))^{-2}$, so the decay in a supersymmetric microstate geometry is slower than for an AdS black hole or an ultracompact neutron star.

An example of a function $g(t)$ for which our quasinormal modes are consistent with (2.111) is given by

$$g(t) = \ell^{-2} \qquad \text{where} \qquad 2\ell\log\ell = \log(2+t) \qquad \text{for} \qquad \ell \gg 1 \tag{2.116}$$

Of course, we are not claiming that a result of the form (2.111) exists, merely that it is not ruled out by the behaviour of quasinormal modes. Such decay is much too slow to be of any use in establishing *nonlinear* stability.

The above analysis can be made rigorous by replacing quasinormal modes with *quasimodes*. These are *approximate* solutions of the wave equation which are compactly supported. In particular, they vanish in a neighbourhood at spatial infinity so one can work with a foliation of constant coordinate time t in the coordinates of (2.36) so the surfaces Σ_t extend to spatial infinity. Using quasimodes one can prove the following [105]

Theorem 2.4.1. *Let k_1, $k_2 > 0$. Let ℓ satisfy the following equation:*

$$\ell \log \ell = \log(2+t) \tag{2.117}$$

Then there exists a universal positive constant $C_{k_1,k_2} > 0$ such that the following holds: for solutions Φ to the linear wave equation $\Box_g \Phi = 0$,

$$\limsup_{t\to\infty} \sup_{\Phi \in H^{k_1+k_2}(\Sigma_0)} \frac{||\Phi||^2_{H^{k_1}(\Sigma_t)}}{||\Phi||^2_{H^{k_1+k_2}(\Sigma_0)}} \ell^{2k_2} \geq C_{k_1,k_2} \tag{2.118}$$

In particular, for any k_1, k_2 this gives sub-polynomial decay.

Here $||\Phi||^2_{H^{k_1}(\Sigma_t)}$ is the kth Sobolev norm associated to Σ_t, i.e., the norm involving an integral over Σ_t of the sum of squares of the first k derivatives of Φ. Our heuristic argument above corresponds to the case $k_1 = k_2 = 1$ of this theorem. In general, the theorem allows for a loss of k_2 derivatives.

2.5 Quasimode construction

Quasimodes are *approximate* solutions of the wave equation, with exponentially small error [32, 33]. Quasimodes can be used to study local features of potentials, and establish rigorous lower bounds on the uniform decay of fields. Even though one can envisage such a construction for generic backgrounds, it has only been firmly established for backgrounds that admit separable solutions [32, 33]. In such cases, the equations of motion governing how certain perturbations propagate on such backgrounds, become a set of coupled ordinary differential equations, for which potentials can be defined. Our geometries fall into that class.

Quasimodes are constructed as follows. One first restricts to a *finite* domain and impose boundary conditions at the edges of this domain. We choose our boundary conditions to be such that at the centre, $w = 0$, the quasimode is regular, and at a given radius, say $w = w_c$ we impose a Dirichlet boundary condition $\Phi = 0$. The choice of w_c is largely irrelevant, except we want to make it sufficiently large that any interesting feature in our potential lies in the interval $w \in (0, w_c)$. We solve this Dirichlet problem for $w < w_c$, which gives a set of normal

mode frequencies, and then set $\Phi = 0$ for $w > w_c$. The resulting solutions are not smooth at $w = w_c$; one defines quasimodes by applying a smoothing procedure near $w = w_c$, which means that one no longer has an exact solution to the wave equation: there is a small error near w_c.

We will perform the first part of this construction, i.e., solution of the Dirichlet problem. It turns out that the associated normal mode frequencies give an excellent fit to the real part of the frequences of our quasinormal modes. For the sake of presentation, we will only describe below the case in which we kept m_ψ fixed, but allow m_ϕ to become arbitrarily large. In addition, we will set $\ell = |m_\phi| + |m_\psi|$.

The idea is simple, we start with a consistent ansatz for the angular and radial eigenfunctions and eigenvalues. These take the following form:

$$X(x) = x^{|m_\psi|}(1-x^2)^{\frac{|m_\phi|}{2}} \sum_{k=0}^{+\infty} \frac{\widetilde{X}_k(x)}{|m_\phi|^k}, \qquad W(w) = e^{-|m_\phi|\widetilde{\phi}(w)} W_0(w) \left[1 + \sum_{k=1}^{+\infty} \frac{\widetilde{W}_k(w)}{|m_\phi|^k}\right],$$
$$A = (|m_\phi| + |m_\psi|)(|m_\phi| + |m_\psi| + 2) + \sum_{k=0}^{+\infty} \frac{\widetilde{A}_k}{|m_\phi|^k}, \qquad \text{and} \qquad \tilde{\omega} = \sum_{k=0}^{+\infty} \frac{\varpi_k}{|m_\phi|^k}.$$

Inputting these into the equations of motion, allows us to determine the coefficients

$$\{\widetilde{X}_k(x), \widetilde{W}_k(w), \widetilde{A}_k, \varpi_k\}$$

to any order in the expansion. For instance, keeping all parameters in the 3-charge microstate geometries gives

$$\widetilde{A}_0 = \alpha_1\,\alpha_2\,\eta\,n\,\varpi_0^2, \qquad \widetilde{A}_1 = -\alpha_1\,\alpha_2\,\eta\,\varpi_0\left[(2n+1)\varpi_0\left(|m_\psi|+1\right) - 2n\varpi_1\right],$$
$$\varpi_0 = 2\,\eta, \qquad \varpi_1 = -\frac{2\left(\alpha_1+\alpha_2+\alpha_1\alpha_2 n^2+\alpha_1\alpha_2 n\right)}{\left[(\alpha_1+\alpha_2)\,n^2+(\alpha_1+\alpha_2)\,n+1\right]^3}.$$

It is possible to go to higher orders in k, but the expressions become increasingly complicated. Progress can be made by choosing specific values for α_1, α_2, n and m_ψ. For instance, for $\alpha_1 = \alpha_2 = n = -m_\psi = 1$ (the parameters of Fig. (2.2)), we find

$$\tilde{\omega} = \frac{2}{5} - \frac{8}{125\left|m_\phi\right|} + \frac{424}{3125\left|m_\phi\right|^2} - \frac{21284}{78125\left|m_\phi\right|^3} + \frac{968684}{1953125\left|m_\phi\right|^4} - \frac{34114268}{48828125\left|m_\phi\right|^5} + O\left(\left|m_\phi\right|^{-6}\right).$$

Chapter 3

Geodesics in supersymmetric microstate geometries

This chapter is entirely my own work, based on my paper [132].

3.1 Introduction

In chapter 2 we argued that the supersymmetric microstate geometries are classically unstable. Recall the heuristic argument for instability in section 2.2.4; this is based on a massive particle near the evanescent ergosurface which cannot escape to infinity. In this chapter we consider whether or not we can actually set up the initial conditions for this process, that is, whether it is possible to have a massive particle that exists arbitrarily far away from the evanescent ergosurface but with positive binding energy, so that it does not escape to infinity.

The second argument in chapter 2 for non-linear instability of the supersymmetric microstate geometries is that solutions of the wave equation decay very slowly. The reason for this is that these spacetimes exhibit *stable trapping*: there are null geodesics that are trapped in some bounded region of space and for which initially close geodesics remain nearby, so they are stable to small perturbations. Since the null geodesics with tangent V on the evanescent ergosurface minimize the energy, they are stably trapped. Stable trapping can cause problems for the decay of solutions to the wave equation because it is possible to construct solutions that are localised near a null geodesic for an arbitrarily long time [31]. In the most symmetrical 2- and 3-charge microstate geometries, quasinormal mode solutions were found in chapter 2 that are localised near the stably trapped null geodesics and decay very slowly. This leads to a particularly slow rate of decay for solutions to the wave equation

in the supersymmetric microstate geometries, which was proven rigorously in [105]. This motivates the study of trapping in the supersymmetric microstate geometries.

The microstate geometries are supposed to be microstates of a black hole, so one might expect that they exhibit some similar properties to a black hole spacetime. It is therefore interesting to compare the geodesics of the microstate geometries to the geodesics around a Kerr black hole. The geodesics in the Kerr spacetime have been studied extensively, see for example [133], [134]. In Kerr, there are circular unstably trapped null geodesics in the equatorial plane, but no stable photon orbits: if perturbed, a photon will either fall into the black hole or escape to infinity. There are also null geodesics that are localised on spheres (not restricted to the equatorial plane) with radius between the radii of the unstably trapped circular geodesics in the equatorial plane. Since the microstate geometries have 5 non-compact dimensions, we should compare the geodesics in the microstate geometries to those around Myers-Perry black holes. These have been classified in [135]; it is particularly interesting to note that although there are unstable circular geodesics, there are no *stable* circular null (or timelike) geodesics in the equatorial plane. In the most symmetrical microstate geometries we will look for the equivalent of these orbits. In contrast to the black holes, we find that there are both stably and unstably trapped null geodesics (which can be circular) in the equatorial plane as well as other constant radius null geodesics that are not necessarily in the equatorial plane.

Strictly we should be comparing the geodesics in the supersymmetric microstate geometries to those of a supersymmetric black hole. The geodesics in the 5d supersymmetric, charged rotating and extremal BMPV [99] black hole spacetime (in supergravity) have been found in [136]. For our comparison, the most important point is that in the underrotating case describing extremal supersymmetric black holes there are no massive or massless bound orbits that are only in the outer region. This is in contrast to the supersymmetric microstate geometries. There are significant differences in the behaviour of geodesics in the supersymmetric microstate geometries and all of the black hole spacetimes discussed above.

It has been shown in [137] that for the most symmetrical 2-charge microstate geometries the Hamilton-Jacobi equation for null geodesics separates in certain coordinates; we find that the same happens for the most symmetrical 3-charge microstate geometries. This separability of the Hamilton-Jacobi equation is due to the fact that these spacetimes have a 'hidden' symmetry related to a conformal Killing tensor which also allows the wave equation to separate in both cases [126, 92].

We will characterise the null geodesics in the most symmetrical supersymmetric microstate geometries, in particular focussing on whether there exist stably trapped or unstably trapped null geodesics since these are important for decay of solutions to the wave equation.

Trapping is best understood on the tangent bundle, so we will study regions of the bundle for which trapping occurs. This means that we will find values of the conserved quantities arising from the various Killing vector fields (components of the tangent to the geodesic) that allow for trapping.

This chapter is organised in the following way: in section 3.2 we find that in general it is possible to have the massive particle for the instability mechanism discussed above for general microstate geometries. In section 3.3 we consider the 2-charge case: after separating the Hamilton-Jacobi equation we will investigate the null geodesics with zero momentum around the internal torus (these also correspond to null geodesics in 6 dimensions) in both of the submanifolds $\theta = \pi/2$ and $r = 0$ and show that there are both stable and unstable trapped geodesics. We will also find geodesics that remain at constant radius, and are thus trapped, but not necessarily in the equatorial plane. In addition, in section 3.4 we consider stable trapping and the relation to the binding energy for geodesics with non-zero momentum around the internal torus (these give massive particles in 5 or 6 dimensions) in the equatorial plane specifically for the 2-charge microstate geometries. The 3-charge case is more complicated, but in section 3.5 we find that the Hamilton-Jacobi equation separates and that in general there are both stable and unstable trapped geodesics, as well as constructing an example of the Penrose process to extract energy in these spacetimes. Finally, in section 3.6 we consider the implications of the trapped geodesics on quasinormal modes.

3.2 Geodesics of general microstate geometries

For a general supersymmetric microstate geometry, we start off in 10 dimensions before reducing to 6 dimensions by compactifying on the internal torus. Suppose we have a high energy graviton in 10d with non-zero momentum around the internal torus. Since it is high energy, we can use the geometric optics approximation to say that we expect it to be localised near a null geodesic and therefore investigate the null geodesics in 10d. If the geodesics have non-zero momentum on the internal torus they correspond to trajectories of massive particles after dimensional reduction to 6d. In the section 2.2.4 we described the mechanism for instability that involves a massive particle with a bound trajectory[1]. We will establish whether it is possible in a general supersymmetric microstate geometry to find such a particle with positive binding energy arbitrarily far away that does not escape to infinity. Note that this is not necessarily obvious, although gravity is an attractive force, because there are

[1] *Bound* means that the particle cannot escape to infinity; since there is no black hole for it to fall in to this is the same as the particle following a geodesic that is *trapped*. However, when we talk about massive particles in 6d we will use ‘bound’; for null geodesics we will use ‘trapped’.

various conformal factors involved in the dimensional reduction that have an effect on the particle. To find the geodesics in the spacetime we will use the Hamilton-Jacobi equation.

The Hamilton-Jacobi function S for geodesics satisfies

$$g^{ab}\nabla_a S\nabla_b S = -\mu^2. \tag{3.1}$$

where $\mu = 0$ for null geodesics or μ is the mass of the particle following a timelike geodesic. This implies that

$$P_a = \nabla_a S \tag{3.2}$$

is the momentum of a particle following a causal geodesic:

$$P^a\nabla_a P_b = P^a\nabla_a\nabla_b S = P^a\nabla_b\nabla_a S = P^a\nabla_b P_a = \frac{1}{2}\nabla_b(P^a P_a) = 0 \tag{3.3}$$

and $g^{ab}P_aP_b = -\mu^2$.

3.2.1 10d null geodesics

We will look for null geodesics in the full 10 dimensions of solutions to type IIB supergravity compactified on T^4. We can write the string frame metric as

$$(g^S_{10})_{\hat{\mu}\hat{\nu}}dx^{\hat{\mu}}dx^{\hat{\nu}} = (g_6)_{\mu\nu}dx^\mu dx^\nu + e^{2\Psi}\delta_{ij}dy^i dy^j \tag{3.4}$$

where $\mu, \nu = 0\ldots5$, $\hat{\mu}, \hat{\nu} = 0\ldots9$ and $i, j = 1\ldots4$, y^i are the coordinates on the internal torus and $\Psi(x)$ is some function independent of these coordinates. The upper index S refers to the fact that this metric is in the string frame. To obtain the 6d Einstein frame metric, one dimensionally reduces on T^4, which gives rise to one conformal factor, and then multiplies by a conformal factor to go from the string to Einstein frame; when we go from 10 to 6 dimensions it turns out that these conformal factors cancel. This means g_6, the 6d part of the 10d string frame metric in (3.4) is in fact the 6d Einstein frame metric.

The Hamilton-Jacobi equation for null geodesics in 10d is

$$0 = g^{\hat{\mu}\hat{\nu}}_{10}\partial_{\hat{\mu}}S\partial_{\hat{\nu}}S = g^{\mu\nu}_6\partial_\mu S\partial_\nu S + e^{-2\Psi}\delta^{ij}\partial_i S\partial_j S. \tag{3.5}$$

Let

$$S = \tilde{S}(x^\mu) + q_i y^i \tag{3.6}$$

where y^i are coordinates on the internal torus, x^μ are the other coordinates and $q_i = \left(\partial/\partial y^i\right)\cdot P$ are the quantities associated with the Killing vectors $\partial/\partial y^i$ that are conserved along a geodesic with momentum P_a.

Substituting (3.6) in (3.5) gives

$$g_6^{\mu\nu}\partial_\mu\tilde{S}\partial_\nu\tilde{S} = -e^{-2\Psi}\mu^2, \qquad \text{where } \mu^2 = \delta^{ij}q_iq_j \tag{3.7}$$

which implies that $\tilde{S}$ satisfies the Hamilton-Jacobi equation with metric $\tilde{g}_6$:

$$\tilde{g}_6^{\mu\nu}\partial_\mu\tilde{S}\partial_\nu\tilde{S} = -\mu^2, \tag{3.8}$$

where the rescaled 6d metric is

$$(\tilde{g}_6)_{\mu\nu} = e^{-2\Psi}(g_6)_{\mu\nu}. \tag{3.9}$$

If all the momenta around the internal torus are zero so that $\mu = 0$, this is exactly the Hamilton-Jacobi equation for null geodesics in 6d since the conformal factor in front of the Einstein frame metric has no effect. If $q_i \neq 0$, this is the Hamilton-Jacobi equation for timelike geodesics with mass μ, *not* for the 6d Einstein frame metric but for $\tilde{g}_6$, the Einstein frame metric multiplied by a conformal factor.

Now suppose we have a general isolated gravitating system in (5+1)-dimensions where one of the dimensions is compact and that near infinity it looks like a product of (4+1)-d Minkowski space with a Kaluza-Klein S^1. We show that for these spacetimes there are trajectories with $\mu \neq 0$ that start arbitrarily far from the evanescent ergosurface but cannot escape to infinity. This includes general supersymmetric microstate geometries, so we can then apply the argument for instability from section 2.2.4. In the full 10 dimensions, these geodesics correspond to null geodesics with non-zero momentum around the internal torus that are stably trapped.

We will assume that the spacetime is stationary and that $T = \partial/\partial t$ is the Killing vector field that is timelike near infinity. Then the energy $E = -T\cdot P = -P_t$ is conserved along geodesics. If the spacetime is a solution of supergravity there is also a globally null Killing vector field V. In the supersymmetric microstate geometries, $V = T + Z$ where z is the coordinate around the Kaluza-Klein S^1 and $Z = \partial/\partial z$ is a Killing vector field. We can define a positive conserved energy with respect to V both in 10d $E_{10} = -V\cdot P$, and in 6d, $E_6 = -V^aP_a = -P_t - P_z$. Note that in 6d P is tangent to a causal geodesic of $\tilde{g}_6$ but we still have $E_6 = E_{10}$, which is clear when we write $E_{10} = -V^aP_a$, as the conformal factors for the

metrics in different frames all cancel out and the energy is conserved from both the 10d and 6d perspectives. Note that in 6d the T energy is positive, $E \geq 0$, when $P_z = 0$.

To proceed, we show that under certain conditions it is always possible to find timelike vectors of the conformal metric $\tilde{g}_6 = e^{-2\Psi} g_6$ (g_6 is the Einstein frame metric) arbitrarily far out, that satisfy $\tilde{g_6}^{\mu\nu} P_\mu P_\nu = -\mu^2$ with $E < \mu$. If we can find such a vector at a point p in the spacetime then we can find a timelike geodesic (of $\tilde{g}$) through p with momentum P which is bound, since the binding energy $\mu - E$ is positive.

The asymptotic form of the Einstein frame metric for an isolated gravitating system in (4+1) dimensions is given in [138]. We can extend this to our case by including an extra compact dimension. To calculate the asymptotic form of the 6d metric we have to dimensionally reduce to 5d and then compare these metric components to those of the general asymptotic form in [138].

To reduce from the 6d to 5d Einstein frame metric,

$$g_6 = e^{2a_1\Phi} g_5 + e^{2a_2\Phi}(dz + \mathscr{A})^2 \tag{3.10}$$

where $a_1^2 = 1/40$, $a_2 = -4a_1$, z is the coordinate around the extra compact dimension, $\Phi(x)$ is a scalar field and $\mathscr{A}$ is a 1-form. We assume that

$$\Phi = \frac{b}{r^2} + O(r^{-3}), \qquad \mathscr{A}_\alpha = \frac{c_\alpha}{r^2} + O\left(\frac{1}{r^3}\right) \tag{3.11}$$

as $r \to \infty$ for some constants b, c_α, ($\alpha = 0 \ldots 4$), since asymptotically the spacetime is $\mathscr{M}_{4,1} \times S^1$.

In 5d, we have a general isolated gravitating system and can use the results of [138]. The system is asymptotically flat, so near infinity we write

$$(g_5)_{\alpha\beta} = \eta_{\alpha\beta} + \hat{h}_{\alpha\beta} \tag{3.12}$$

where $|\hat{h}_{\alpha\beta}| \ll 1$ and $\alpha, \beta \in \{t, i\}$, $i = 1 \ldots 4$. We have that

$$\hat{h}_{tt} = \frac{d_1}{r^2}, \qquad \hat{h}_{ij} = \frac{d_2}{r^2}\delta_{ij}, \qquad \hat{h}_{ti} = O\left(\frac{1}{r^3}\right) \tag{3.13}$$

for some positive constants d_1, d_2 that can be found in [138].

Now in 6d, we can write the metric near infinity as

$$(g_6)_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu} \tag{3.14}$$

where, from (3.10), at leading order,

$$h_{tt} = \frac{e_1}{r^2}, \qquad h_{ij} = \frac{e_2}{r^2}\delta_{ij}, \qquad h_{zz} = \frac{e_3}{r^2}, \qquad h_{\alpha z} = \frac{c_\alpha}{r^2}, \qquad h_{ti} = O\left(\frac{1}{r^3}\right) \tag{3.15}$$

where $e_1 = d_1 - 2a_2b$, $e_2 = d_2 + 2a_2b$, $e_3 = 2a_2b$, $i, j = 1 \ldots 4$ and $\alpha \in \{t, i\}$.

The inverse metric at this order is $g^{\mu\nu} = \eta^{\mu\nu} - h^{\mu\nu}$ where the indices on the right hand side are raised with $\eta^{\mu\nu}$.

We assume that for large r the conformal factor is,

$$e^{2\Psi} = 1 + \frac{f}{r^2} + O\left(\frac{1}{r^3}\right) \tag{3.16}$$

for some constant f.

Timelike vectors of $\tilde{g}$ satisfy

$$\tilde{g}^{\mu\nu} P_\mu P_\nu = e^{2\Psi} g^{\mu\nu} P_\mu P_\nu = -\mu^2. \tag{3.17}$$

Substituting (3.15) and (3.16) into (3.17) gives

$$(-1 - \frac{f+e_1}{r^2})P_t^2 + (1 + \frac{f-e_2}{r^2})\sum_{i=1}^{4} P_i^2 + (1 + \frac{f-e_3}{r^2})P_z^2 + 2(\frac{c_t}{r^2}P_t - \sum_{i=1}^{4}\frac{c_i}{r^2}P_i)P_z + O(r^{-3}) = -\mu^2. \tag{3.18}$$

If $P_z = 0$, solving for P_t gives

$$P_t^2 = \mu^2 - \frac{f+e_1}{r^2}\mu^2 - (1 - \frac{e_1+e_2}{r^2})\sum_{i=1}^{4} P_i^2 + O(r^{-3}). \tag{3.19}$$

Therefore, if

$$f + e_1 > 0 \tag{3.20}$$

it is possible to choose P_i small enough such that (3.19) is satisfied and that $P_t^2 < \mu^2$.

The condition that the particle has positive binding energy, $E^2 < \mu^2$, implies that the particle cannot escape to infinity and is stably trapped. To see this, take $r \to \infty$ in (3.18): the left hand side $LHS \geq -E^2$ while the right hand side $RHS < -E^2$, since these cannot be equal it is not possible to have timelike vectors with $E^2 < \mu^2$ as $r \to \infty$. Along a geodesic, E is conserved, so the geodesic can never go all the way out to infinity. Furthermore, even if the geodesic is perturbed slightly the binding energy is positive so it still will not escape to infinity. Thus the geodesic is stably trapped.

For the general 2-charge microstate geometries in [89],

$$f + e_1 = Q_1 > 0 \tag{3.21}$$

provided $Q_1 > 0$, which is indeed the case for the microstate geometries in [89]. For the supersymmetric 3-charge solutions in [92], $f + e_1 = Q_1 + Q_p > 0$, since $Q_1 > 0$ and $Q_p > 0$ in these solutions. This implies that it is always possible to find a stably trapped null (in 10d) geodesic with non-zero momenta around the internal torus in these geometries. Moreover, this is possible arbitrarily far out and thus we can set up the initial conditions for the heuristic argument of instability given in section 2.2.4 by the presence of stably trapped massive particles. If the asymptotics only depend on the charges, we expect this to also be the case for all of the 2-charge solutions.

For a BMPV black hole (starting from the 6d solution) $f = 0$ and $e_1 = M$, the mass of the black hole, implying that there are always bound geodesics with $E^2 < \mu^2$. This is in agreement with [136], where they find that geodesics with $E^2 < \mu^2$ are always bound orbits. Note that these orbits always cross the horizon and thus do not exist only in the outer region, which we cannot derive just from the above calculation.

3.3 2-charge microstate geometry

In this section we study the trapping of *null* geodesics in the most symmetrical supersymmetric microstate geometries. These geodesics are of interest for the second argument for instability in chapter 2, involving slow decay of solutions of the wave equation.

Recall the 2-charge microstate geometries introduced in section 2.3.3. The 10d string frame metric for this 2-charge D1-D5 microstate geometry (in the form given in [126]) is

$$\begin{aligned} ds_{10}^2 =& -\frac{1}{h}(dt^2 - dz^2) + hf\left(d\theta^2 + \frac{dr^2}{r^2+a^2}\right) - \frac{2a\sqrt{Q_1 Q_2}}{hf}\left(\cos^2\theta dz d\psi + \sin^2\theta dt d\phi\right) \\ &+ h\Big[\left(r^2 + \frac{a^2 Q_1 Q_2 \cos^2\theta}{h^2 f^2}\right)\cos^2\theta d\psi^2 + \left(r^2 + a^2 - \frac{a^2 Q_1 Q_2 \sin^2\theta}{h^2 f^2}\right)\sin^2\theta d\phi^2\Big] \\ &+ \sqrt{\frac{H_1}{H_2}}\Sigma_{i=1}^4 dx_i^2 \\ =& ds_6^2 + \sqrt{\frac{H_1}{H_2}}\Sigma_{i=1}^4 dx_i^2 \end{aligned} \tag{3.22}$$

where $r \geq 0$, $\theta \in [0, \pi/2]$, $0 \leq \phi, \psi \leq 2\pi$, $z \sim z + 2\pi R_z$,

$$f = r^2 + a^2 \cos^2\theta, \qquad H_i = 1 + \frac{Q_i}{f} \quad (i = 1, 2), \qquad h = (H_1 H_2)^{\frac{1}{2}} \tag{3.23}$$

and the solution is written in terms of the charges Q_1 and Q_2.

There is a globally null Killing vector field

$$V = \frac{\partial}{\partial t} + \frac{\partial}{\partial z}. \tag{3.24}$$

As in chapter 2, we define the evanescent ergosurface $\mathscr{S}$ in 6d to be the locus of points where $V \cdot Z = 0$, where $Z = \partial/\partial z$ is the Kaluza-Klein Killing vector field; in 5d V is timelike everywhere except on the evanescent ergosurface, where it is null. Therefore the evanescent ergosurface is given by $f = 0$, i.e. $r = 0$ and $\theta = \pi/2$, with topology S^1 at constant t.

This is a particularly important submanifold since it was shown in chapter 2 that there are stably trapped null geodesics with zero energy that stay on the evanescent ergosurface. Indeed, the only geodesics with zero Kaluza-Klein momentum that have zero energy are those with tangent V that are stably trapped on $\mathscr{S}$. It was also shown in chapter 2 that geodesics with tangent V through any point in the spacetime are stably trapped, but those away from $\mathscr{S}$ have non-zero Kaluza-Klein momentum.

3.3.1 Hamilton-Jacobi equation

The Hamilton-Jacobi equation for null geodesics (in 10d) is separable due to a conformal Killing tensor [137]. We can then separate the equation for geodesics in 6d, so in (3.8) let

$$\tilde{S} = p_I x^I + R(r) + \Theta(\theta). \tag{3.25}$$

where $p_I = \left(\partial/\partial x^I\right) \cdot P$.

Using the ansatz (3.25) equation (3.8) separates into equations for $R(r)$ and $\Theta(\theta)$. Then recall that $P_\mu = \nabla_\mu S$ is the momentum of a particle following a null geodesic so $P^\mu = \dot{x}^\mu(\lambda) = g_{10}^{\mu\nu}\nabla_\nu S$ where λ is an affine parameter along the geodesic, since these geodesics are null in 10d and we can rescale the affine parameter along the null geodesic. Then using

$$\dot{r}(\lambda) = \frac{dr}{d\lambda} = g_{10}^{r\mu}\partial_\mu S = g_{10}^{rr}\partial_r R, \qquad \dot{\theta}(\lambda) = \frac{d\theta}{d\lambda} = g_{10}^{\theta\mu}\partial_\mu S = g_{10}^{\theta\theta}\partial_\theta \Theta$$

where λ is an affine parameter along the geodesic, we find coupled first order differential equations for $\dot{\theta}$ and $\dot{r}$:

$$\dot{r}^2 + \frac{1}{(hf)^2} V_r(r) = 0 \tag{3.26}$$

$$\dot{\theta}^2 + \frac{1}{(hf)^2} V_\theta(\theta) = 0 \tag{3.27}$$

where the effective potentials are

$$\begin{aligned} V_r = (p_z^2 - p_t^2 + \mu^2) r^4 + \left(a^2(p_z^2 - p_t^2 + \mu^2) + \Lambda + \mu^2 Q_2\right) r^2 - (a p_\phi + \sqrt{Q_1 Q_2} p_t)^2 \\ + (a p_\psi + \sqrt{Q_1 Q_2} p_z)^2 + \Lambda a^2 + \mu^2 a^2 Q_2 + (a p_\psi + \sqrt{Q_1 Q_2} p_z)^2 \frac{a^2}{r^2} \end{aligned} \tag{3.28}$$

and

$$V_\theta = \frac{p_\phi^2}{\sin^2\theta} + a^2 \cos^2\theta (p_z^2 - p_t^2 + \mu^2) + \frac{p_\psi^2}{\cos^2\theta} + (Q_1 + Q_2)(p_z^2 - p_t^2) - \Lambda \tag{3.29}$$

where Λ is a constant arising from the separation of variables.

Equations (3.26) and (3.27) are coupled via the factor $hf(r,\theta)$ so that they are not in fact in the form of 1d equations of motion with an effective potential. However, this factor is strictly positive so we can still say something about the trapped geodesics by investigating the signs of V_r and V_θ since geodesics can only exist in regions where $V_r(r) \leq 0$ and $V_\theta(\theta) \leq 0$.

3.3.1.1 Trapping

For a geodesic to exist there must be some value of θ such that $V_\theta(\theta) \leq 0$. Assuming that this is the case, whether or not there is trapping depends only on the radial effective potential V_r.

Stable trapping occurs if there are some values $r_- \leq r_0 < r_+$ such that $V_r(r) \leq 0$ for $r_- \leq r \leq r_0$ but $V_r(r) > 0$ for $r_0 < r < r_+$; see Figure 3.1 for an example. The geodesics are then allowed to propagate in $r_- \leq r \leq r_0$ but cannot escape to infinity due to the potential barrier at $r_0 < r < r_+$. Suppose there is a geodesic that is trapped in this region. If we perturb this trapped geodesic (in the tangent bundle) the shape of the potential changes slightly but there will still be a potential barrier and the perturbed geodesic is also trapped. This is stable trapping since perturbing a trapped geodesic can only give another trapped geodesic.

On the other hand, there is unstable trapping if there is some value r_0 such that $V_r(r)$ has a (local) maximum at r_0 and $V_r(r_0) = 0$, as in Figure 3.2: the geodesic at r_0 stays there, but if there is a nearby geodesic at $r = r_0 + \varepsilon$ for some ε then that geodesic can escape away from r_0. Also, if we perturb the potential slightly there will no longer be a maximum when $V_r = 0$ and there is no reason for geodesics to stay near there: if the maximum moves in such a way to be at a negative value of V_r the perturbed geodesics are not trapped but can escape out

to infinity, so this is unstable trapping. Note that the unstably trapped geodesic will always remain at constant r, so that in the equatorial plane $\theta = \pi/2$ these orbits are all circular.

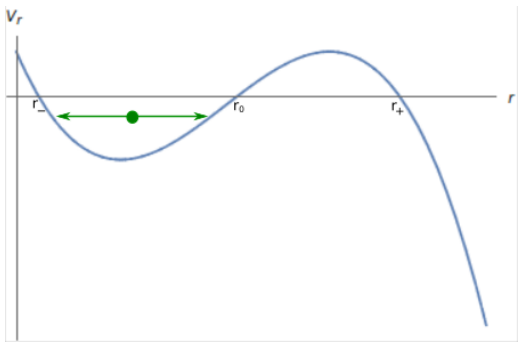


Fig. 3.1 Stable trapping

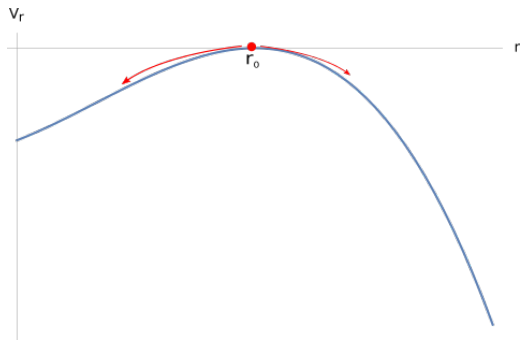


Fig. 3.2 Unstable trapping

We can distinguish between trapped null geodesics with positive or negative angular momentum (in the $\phi-$direction), $p_\phi > 0$ or $p_\phi < 0$, in the same way as for the unstably trapped circular null geodesics in the equatorial plane of Kerr black holes, see for example [139, 133]. In Kerr (taking $a = J/M > 0$), there are two possible radii for circular orbits in the equatorial plane: at $r = r_-$ there are *direct* or *prograde* circular orbits with positive angular momentum $p_\phi > 0$ while at $r = r_+$ there are *retrograde* circular orbits with negative angular momentum $p_\phi < 0$. Note that this definition of pro- and retrograde orbits applies when the angular momentum of the spacetime is positive ($a > 0$). Since we also have positive angular momentum (in the $\phi-$direction) $J_\phi > 0$, we will define prograde and retrograde orbits in the same way.

For geodesics tangent to V, we can calculate $p_\phi = -2a\sqrt{Q_1 Q_2}/hf < 0$ and see that all of these geodesics, including those with zero energy on the evanescent ergosurface, are retrograde. This is best explained in 5d: as V approaches a standard time translation at infinity the geodesics with tangent V do not rotate with respect to infinity. However, $J_\phi > 0$ so the spacetime has non-zero, positive angular momentum. This means that these geodesics are resisting the frame-dragging effect of the geometry and thus have angular momentum opposite to that of the spacetime.

3.3.2 6d null geodesics in equatorial plane $\theta = \pi/2$

We expect stable trapping near (in the tangent bundle) the zero energy null geodesics that are stably trapped on the evanescent ergosurface and we determine the region filled by such geodesics, which are retrograde. We also find that for certain microstate geometries, it is possible to have stably trapped prograde null geodesics.

We consider the null geodesics of the 6d metric that stay in the equatorial plane. Null geodesics in 10d with zero momentum around the internal torus are also null geodesics in 6d and have $\mu = 0$ in equations (3.8), (3.28) and (3.29).

Although equations (3.26) and (3.27) are coupled we can still find geodesics that stay at $\theta = \pi/2$, which we expect by symmetry. Differentiating equation (3.27) wrt λ and dividing through by $\dot{\theta}$ gives the second order equation for $\theta(\lambda)$:

$$2\ddot{\theta} - hf\dot{r}\sqrt{-V_\theta}\partial_r\left((hf)^{-2}\right) + V_\theta\,\partial_\theta\left((hf)^{-2}\right) + (hf)^{-2}\partial_\theta V_\theta = 0. \tag{3.30}$$

Suppose that for some λ_0 we have $\theta(\lambda_0) = \theta_0$ and $\dot{\theta}(\lambda_0) = 0$ so that $V_\theta(\theta_0) = 0$; if also $V'_\theta(\theta_0) = 0$ equation (3.30) implies that $\ddot{\theta}(\lambda_0) = 0$ and so the geodesic remains at constant θ_0.

For geodesics at $\theta = \pi/2$ we must have $V_\theta(\pi/2) = 0$, i.e.

$$p_\psi = 0, \quad \Lambda = p_\phi^2 + (Q_1 + Q_2)(p_z^2 - p_t^2). \tag{3.31}$$

Differentiating (3.29) wrt θ and substituting in $p_\psi = 0$, we find that we do indeed have $V'_\theta(\pi/2) = 0$ as required for the geodesics to stay at $\theta = \pi/2$.

3.3.2.1 Radial equation

We will find the geodesics in the submanifold $\theta = \pi/2$ with zero Kaluza-Klein momentum $p_z = 0$, since these geodesics will also correspond to massless particles after dimensional reduction to 5d.

Define the impact parameters

$$b_\phi = -\frac{p_\phi}{p_t}, \qquad b_\psi = -\frac{p_\psi}{p_t}, \qquad b_z = -\frac{p_z}{p_t}. \tag{3.32}$$

Due to the freedom to rescale the affine parameter along the geodesic, it is only these ratios that have any physical importance. We will look for values of b_ϕ that give either stable or unstable trapping ($b_\psi = 0$ for these geodesics with $\theta = \pi/2$, and by our choice $b_z = 0$).

As $\partial/\partial t$ is everywhere causal in the 2-charge microstate geometry, $p_t \leq 0$ for a future-directed null geodesic and an equivalent definition of pro/retro-grade in terms of $b_\phi = -p_\phi/p_t$ is that an orbit is direct or prograde if $b_\phi > 0$ and retrograde if $b_\phi < 0$. The zero energy null geodesics that are stably trapped on $\mathscr{S}$ have $b_\phi = -\infty$.

Substituting (3.31) into (3.28) we can write the potential as

$$\begin{gathered} V_r = p_t^2(-r^4+Br^2+C) \\ \text{where } B = b_\phi^2 - (Q_1+Q_2) - a^2, \quad C = 2a\sqrt{Q_1Q_2}(b_\phi - \xi) \end{gathered} \tag{3.33}$$

where

$$\xi = \left(Q_1Q_2 + a^2(Q_1+Q_2)\right)/(2a\sqrt{Q_1Q_2}). \tag{3.34}$$

Note that $p_t = 0$ gives the stably trapped zero energy null geodesics on the evanescent ergosurface. As discussed in section 3.3.1.1, trapping depends on the sign of V_r. To find regions where V_r is negative we note that $V_r \to -\infty$ as $r \to \infty$ so there is always an allowed region near infinity (unless $p_t = 0$) and then we only have to find the roots of V_r, which is a quadratic polynomial in r^2.

We find various different possibilities for the types of geodesic according to the values of B and C, which depend on b_ϕ. The bounds and equations for B and C which lead to the different types of trapping give conditions on b_ϕ.

Define

$$b_\phi^- = -a - \sqrt{Q_1} - \sqrt{Q_2}, \tag{3.35}$$

$$b_\phi^+ = \max\{-a + \sqrt{Q_1} + \sqrt{Q_2},\, a \pm (\sqrt{Q_1} - \sqrt{Q_2})\}, \tag{3.36}$$

$$\begin{aligned} r_-^2 &= \max\{0, -\sqrt{Q_1Q_2} + a|\sqrt{Q_1} - \sqrt{Q_2}|,\, \sqrt{Q_1Q_2} - a(\sqrt{Q_1} + \sqrt{Q_2})\} \\ r_+^2 &= \sqrt{Q_1Q_2} + a(\sqrt{Q_1} + \sqrt{Q_2}). \end{aligned} \tag{3.37}$$

and

$$2r_1^2 = b_\phi^2 - Q_1 - Q_2 - a^2 - \left((b_\phi^2 - Q_1 - Q_2 - a^2)^2 + 8a\sqrt{Q_1Q_2}(b_\phi - \xi)\right)^{1/2}. \tag{3.38}$$

There are always stably trapped retrograde geodesics with $b_\phi < b_\phi^-$, but what happens for larger b_ϕ depends on the background parameters. The two different cases, which depend on the value of R_z compared to Q_i, are:

1. $(Q_1+Q_2+a^2)^{\frac{1}{2}} < \xi$:
   This happens if $R_z > \sqrt{Q_1} + \sqrt{Q_2}$ or $0 < R_z < |\sqrt{Q_1} - \sqrt{Q_2}|$.

2. $(Q_1+Q_2+a^2)^{\frac{1}{2}} \geq \xi$:
   This requires $|\sqrt{Q_1} - \sqrt{Q_2}| \leq R_z \leq \sqrt{Q_1} + \sqrt{Q_2}$.

We summarize the possible geodesics in the two cases by giving the ranges of b_ϕ that give rise to different types of trapping:

1. $(Q_1+Q_2+a^2)^{\frac{1}{2}} < \xi$

   (a) $b_\phi < b_\phi^-$: retrograde stably trapped geodesics in a region $0 \le r \le r_1$ that includes the evanescent ergosurface at $r=0$ and r_1 is given in (3.38).

   (b) $b_\phi = b_\phi^-$: unstable trapping of retrograde geodesics at r_+.

   (c) $b_\phi^- < b_\phi < b_\phi^+$: no trapping, and geodesics from infinity can reach the evanescent ergosurface at $r=0$.

   (d) $b_\phi = b_\phi^+$: prograde unstably trapped geodesics at r_-.

   (e) $b_\phi^+ < b_\phi \le \xi$: stable trapping in the region $0 \le r < r_-$ that includes the evanescent ergosurface, and the geodesics are prograde. If $b_\phi = \xi$ the geodesics are stably trapped precisely on the evanescent ergosurface at $r=0$.

   (f) $b_\phi > \xi$: no trapping, and the geodesics are bounded away from the evanescent ergosurface.

2. $(Q_1+Q_2+a^2)^{\frac{1}{2}} \ge \xi$

   (a) $b_\phi < b_\phi^-$: retrograde stably trapped geodesics in a region $0 \le r \le r_1$ that includes the evanescent ergosurface.

   (b) $b_\phi = b_\phi^-$: unstable trapping of retrograde geodesics at $r_+ > 0$.

   (c) $b_\phi^- < b_\phi < \xi$: no trapping, and geodesics can reach the evanescent ergosurface.

   (d) $b_\phi = \xi$: unstable trapping of prograde geodesics on the evanescent ergosurface at $r=0$.

   (e) $b_\phi > \xi$: no trapping, but geodesics are bounded away from the evanescent ergosurface.

The effective potentials for the possible types of trapping are illustrated in Figures 3.3 and 3.4.

The stably trapped geodesics are in the region $0 \le r < r_1$ where r_1 is given in (3.38). For the prograde stably trapped geodesics this decreases as b_ϕ increases and so the maximum radius for these geodesics is r_-, the radius of the unstable trapped orbits. For the retrograde stably trapped orbits, we have $\frac{dr_1^2}{db_\phi} \to \infty$ as $b_\phi \to b_{\phi *}$ and $\frac{dr_1^2}{db_\phi} \to 0_+$ as $b_\phi \to -\infty$; generically we expect that $\frac{dr_1^2}{db_\phi} > 0$ for $-\infty < b_\phi \le b_{\phi_*}$ and thus r_1 decreases as b_ϕ decreases and the maximum radius of these orbits is r_+, the radius of the unstably trapped retrograde orbit.

We have so far taken b_ϕ to be finite. However, the stably trapped geodesics on the evanescent ergosurface in chapter 2 have zero energy. If we take the limit $p_t \to 0$ in (3.33)

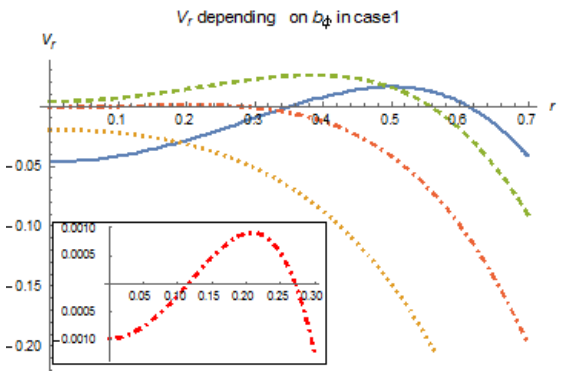


Fig. 3.3 V_r when $(Q_1+Q_2+a^2)^{\frac{1}{2}} < \xi$ with $a = 1/8$, $Q_1 = Q_2 = 1/8$. In order of increasing b_ϕ, the blue solid line is case (a), the orange dotted line is (c), the red dot-dashed line is (e) and the green dashed line is (f). The inset shows (e) zoomed in near the origin.

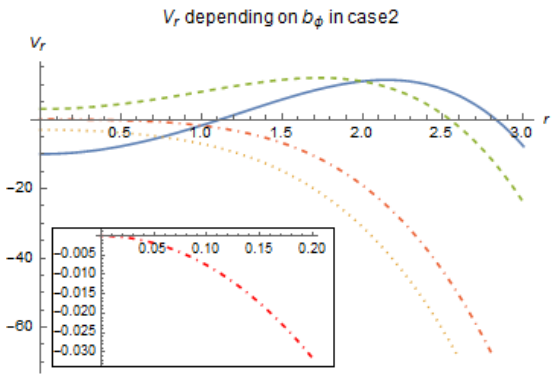


Fig. 3.4 V_r when $(Q_1+Q_2+a^2)^{\frac{1}{2}} \geq \xi$ with $a = 1$, $Q_1 = Q_2 = 1$. In order of increasing b_ϕ, the blue solid line is case (a), the orange dotted line is (c), the red dot-dashed line is (d) and the green dashed line is (e). The inset shows (d) zoomed in near the origin.

the potential has the form

$$V_r = p_\phi^2 r^2 - 2a p_\phi p_t + \dots \tag{3.39}$$

where the dots represent terms for which the coefficients are $O(p_t^2)$. If $p_\phi < 0$ so that we are taking $b_\phi \to -\infty$ the geodesics are trapped in a region near $r = 0$ that becomes smaller in the limit until we have the retrograde zero energy geodesics that are stably trapped exactly at the evanescent ergosurface (this is included in case 1(a) or 2(a)). If $p_\phi > 0$ there could only be geodesics when $p_t = 0$ which implies that these are tangent to V, but these have $p_\phi < 0$.

3.3.3 Geodesics at $r = 0$

We now look for geodesics that stay in the submanifold $r = 0$. Since these geodesics are by definition trapped as they remain inside a bounded region and cannot escape to infinity, we will discuss whether or not there are geodesics that are restricted to some range of θ and if they can reach the evanescent ergosurface at $\theta = \pi/2$ (and $r = 0$).

In a similar way to the geodesics in the submanifold $\theta = \pi/2$, if $V_r(r_0) = 0$ and $V_r'(r_0) = 0$ at some $r_0 = r(\lambda_0)$ then $\dot{r}(\lambda_0) = 0 = \ddot{r}(\lambda_0)$ and the geodesic stays at constant r_0. Substituting $r = 0$ into V_r in (3.28) we see that $V_r|_{r=0} = 0$ and $V_r'(0) = 0$ as required if

$$p_z = 0, \qquad p_\psi = 0, \qquad \frac{\Lambda}{p_t^2} = \left(\frac{\sqrt{Q_1 Q_2}}{a} - b_\phi\right)^2. \tag{3.40}$$

We now investigate the sign of the angular potential (3.29) to find the 'allowed' and 'forbidden' regions for the geodesics. Substituting the appropriate values (3.40) into (3.29)

gives

$$V_\theta = \frac{p_t^2}{\sin^2\theta}\left[a^2\sin^4\theta - \left((b_\phi - \frac{\sqrt{Q_1Q_2}}{a})^2 + Q_1 + Q_2 + a^2\right)\sin^2\theta + b_\phi^2\right]. \tag{3.41}$$

Define

$$x = \sin^2\theta, \qquad x \in [0,1]. \tag{3.42}$$

To find the regions where geodesics can exist we need to find how many roots of

$$W_\theta(x) = a^2x^2 - \left((b_\phi - \frac{\sqrt{Q_1Q_2}}{a})^2 + Q_1 + Q_2 + a^2\right)x + b_\phi^2 \tag{3.43}$$

there are in the range $x \in [0,1]$.

By examining the coefficients of (3.43) we find that $W_\theta|_{x=0} \geq 0$ and that there are always two real roots x_-, x_+ that are both positive (and $x_- = x_+$ is only possible if $Q_1 = Q_2$). We then have two possibilities:

i) $b_\phi > \xi$. This splits into two subcases, neither of which give geodesics in $r = 0$:

- $\left(b_\phi - \frac{\sqrt{Q_1Q_2}}{a}\right)^2 > a^2 - Q_1 - Q_2$
  Both roots $x_-, x_+ > 1$ and the effective potential is strictly positive for $x \in [0,1]$. This implies that there are no geodesics with this range of values of the impact parameter in the submanifold $r = 0$.
- $\left(b_\phi - \frac{\sqrt{Q_1Q_2}}{a}\right)^2 < a^2 - Q_1 - Q_2$
  This only happens if the background parameters of the microstate geometry satisfy

$$1 < \sqrt{\frac{Q_1Q_2}{a^4}} + \sqrt{1 - \left(\frac{Q_1}{a^2} + \frac{Q_2}{a^2}\right)} \equiv G\left(\frac{Q_1}{a^2}, \frac{Q_2}{a^2}\right).$$

  However, this is not actually possible. Let $X = Q_1/a^2 + Q_2/a^2$, note that $\sqrt{Q_1Q_2}/a^2 < \frac{1}{2}X$ and so $G\left(Q_1/a^2, Q_2/a^2\right) < \frac{1}{2}X + \sqrt{1-X} = F(X)$. But $F'(X) < 0$ (assuming the charges are positive and $0 < X < 1$) so $F(X) < F(0) = 1$. Thus we in fact have

$$G\left(\frac{Q_1}{a^2}, \frac{Q_2}{a^2}\right) < F(X) < F(0) = 1.$$

ii) $b_\phi \leq \xi$

There is one root $x_- \in [0,1]$ and the other root $x_+ \geq 1$. The geodesics can exist in $\theta \in [\theta_-, \pi/2]$ and therefore are allowed in a region that includes the evanescent ergosurface.

The zero energy null geodesics of chapter 2 that are trapped exactly on the evanescent ergosurface can seen by taking $p_t \to 0$ in (3.41). If p_ϕ is negative this is the limit $b_\phi \to -\infty$ and we have case (ii) with the root $x_- \to 1$ so the geodesics are trapped exactly at $\theta = \pi/2$.

Geodesics in the submanifold $r = 0$ can reach the evanescent ergosurface at $\theta = \pi/2$ when $b_\phi \leq \xi$. We can compare this to geodesics in the equatorial plane, which reach $r = 0$ when $C \leq 0$ in (3.33): this happens when $b_\phi \leq \xi$ so we in fact have the same range of b_ϕ for which geodesics reach the evanescent ergosurface in either submanifold. However, the geodesics in the equatorial plane that can reach $r = 0$ do not cross from $\theta = \pi/2$ to the submanifold $r = 0$ (and θ arbitrary), which are in an orthogonal surface. This can be seen either by noting that the geodesics in each submanifold have different separation constants and so the equivalent of the Carter constant is different for each family, or that there is an S^3 which shrinks to zero size at $r = 0$, $\theta = \pi/2$ and in coordinates that are regular near the evanescent ergosurface it can be seen that this prevents the geodesics from crossing here[2].

3.3.4 Geodesics at constant r

We mentioned in section 3.3.1.1 that in the Kerr spacetime there are null geodesics that follow unstable circular orbits in the equatorial plane with radius r_- or r_+. In Kerr, there are also spherical photon orbits that remain at constant r but for which θ varies, and the radius r of these orbits is in the range $r_- \leq r \leq r_+$ [134]. We will look for an analogue of these spherical photon orbits in the 2-charge microstate geometries, and find that they do exist but that in contrast to Kerr there is in general no restriction on the radius of these orbits; however, if we set $p_z = 0 = p_\psi$ the constant-r geodesics are restricted to the range $r_- \leq r \leq r_+$ where r_-, r_+ are the radii of the unstable trapped orbits in the equatorial plane.

A geodesic at constant $r = r_0$ must have $V_r(r_0) = 0$ and $V_r'(r_0) = 0$ so that, from (3.26), if $r(\lambda_0) = r_0$, $\dot{r}(\lambda_0) = 0$ then $\dot{r}(\lambda) = 0$ and $\ddot{r}(\lambda) = 0$ and the geodesic remains at r_0 for all values of the affine parameter λ. The geodesic then takes values of θ for which $V_\theta \leq 0$; this must be the case for at least one value $\theta_0 \in [0, \pi/2]$ for the geodesic to exist at all. We therefore find values r_0 such that $V_r(r_0) = 0$, $V_r'(r_0) = 0$ and $V_\theta(\theta) \leq 0$ for some values of the parameters $b_\psi, b_\phi, b_z, \Lambda/p_t^2$ and some θ.

Define $p_t^2 W_r = r^2 V_r$ so that W_r is a cubic polynomial in r^2 and

$$\{W_r(r_0) = 0, \frac{dW_r}{dr}(r_0) = 0\} \Rightarrow \{V_r(r_0) = 0, \frac{dV_r}{dr}(r_0) = 0\} \text{ or } r = 0.$$

[2]Such coordinates can be obtained via a coordinate transformation similar to that in [140], and it turns out that the surfaces $r = 0$ and $\theta = \pi/2$ are orthogonal to each other.

For ease of notation, let

$$\tilde{\Lambda} = \Lambda/p_t^2, \qquad \mu = 1 - b_z^2 \leq 1, \qquad \nu = ab_\psi + \sqrt{Q_1 Q_2} b_z \tag{3.44}$$

We will find values of r for which it is possible to have geodesics at constant r by using the equations

$$W_r(r) = 0, \qquad \frac{dW_r}{dr} = 0 \tag{3.45}$$

to find the parameters $\tilde{\Lambda}$ and b_ϕ in terms of r^2, b_ψ and b_z. Then substituting these into the requirement that $V_\theta \leq 0$ for some value of θ gives an inequality of the form $F(r^2, b_\psi, b_z, \theta) \geq 0$: given values of b_ψ, b_z and θ we can use this to find the possible range of r.

Solving equations (3.45) for $\tilde{\Lambda}$ and b_ϕ gives:

$$\begin{aligned} \tilde{\Lambda} =& 2\mu r^2 + \mu a^2 + a^2 \nu^2 \frac{1}{r^4} \\ b_\phi =& \frac{\sqrt{Q_1 Q_2}}{a} \pm \left(\frac{a}{r^2} + \frac{1}{a}\right) \sqrt{\nu^2 + \mu r^4}. \end{aligned} \tag{3.46}$$

Note that we must have $\nu^2 + \mu r^4 \geq 0$ so that b_ϕ is real. For the parameters to be such that $V_\theta \leq 0$ for some θ, from (3.29),

$$\tilde{\Lambda} \geq \frac{b_\phi^2}{1-u^2} + \frac{b_\psi^2}{u^2} - \mu(Q_1 + Q_2) - \mu a^2 u^2 \tag{3.47}$$

where $u = \cos\theta$, $0 \leq u \leq 1$. Substituting in $\tilde{\Lambda}(r^2, b_z, b_\psi)$ and $b_\phi(r^2, b_z, b_\psi)$ from (3.46) gives the inequality

$$F(r^2, b_z, b_\psi, u) \geq 0 \tag{3.48}$$

where

$$\begin{aligned} F(r^2, b_z, b_\psi, u) =& 2\mu r^6 + \mu(a^2 + Q_1 + Q_2) r^4 + a^2 \nu^2 - \frac{\nu^2}{1-u^2} \frac{(a^2 + r^2)^2}{a^2} - \frac{1}{1-u^2} \frac{(a^2 + r^2)^2}{a^2} \mu r^4 \\ &\mp \frac{1}{1-u^2} \frac{2\sqrt{Q_1 Q_2}}{a^2} r^2 (a^2 + r^2) \sqrt{\nu^2 + \mu r^4} - \frac{1}{1-u^2} \frac{Q_1 Q_2}{a^2} r^4 - \left(\frac{(\nu + Q_1 Q_2 \sqrt{1-\mu})^2}{a^2 u^2} - \mu a^2 u^2 \right) r^4 \end{aligned} \tag{3.49}$$

which gives a constraint on the values of r for which it is possible to have geodesics at constant r for certain values of b_ψ, b_z and u. Alternatively, (3.48) could be used to find the range of θ for a geodesic at constant r given r, b_ψ and b_z.

Instead of attempting to explicitly solve $F \geq 0$ to find the allowed radii of the constant-r geodesics in terms of b_z, b_ψ and u we will simply explain why in general these geodesics

can actually exist at any radius: in chapter 2 it is found that there is are stably trapped geodesics with tangent V through every point in the spacetime. As $V = \partial/\partial t + \partial/\partial z$ these geodesics remain at constant r (and θ) and hence must satisfy (3.48) with $\mu = 0$ and b_ψ, b_z as appropriate and we have geodesics at constant r for *any* radius.

The radius at which these geodesics can exist in general depends on θ. It is interesting to consider the specific case $p_\psi = 0$ so that there are geodesics which lie entirely within the submanifold $\theta = \pi/2$. If we also set $p_z = 0$ then we know the radii of the unstably trapped geodesics in the equatorial plane from section 3.3.2. In this case we actually find something similar to the spherical photon orbits of Kerr [134]: the geodesics at constant r with $p_z = 0$, $p_\psi = 0$ can only exist in the range $r_- \leq r \leq r_+$ where $r_\pm$ are the radii of the unstable photon orbits in $\theta = \pi/2$ (it is possible that $r_- = 0$).

To see this, set $p_z = 0 = p_\psi$ in (3.49) and multiply through by $(1-u^2)r^{-4}$, so the inequality $F \geq 0$ becomes

$$G(u,r) = -a^2u^4 - (2r^2 + Q_1 + Q_2)u^2 + \mathscr{Q} \geq 0 \tag{3.50}$$

where we use the notation

$$\mathscr{Q} = 2r^2 + a^2 + Q_1 + Q_2 - \frac{1}{a^2}\left(\sqrt{Q_1Q_2} \pm (a^2 + r^2)\right)^2$$

for comparison to the geodesics in Kerr [134]. Observe that $G(1,r) < 0$ and that the coefficient of u^2 is negative so that for geodesics to exist we must have $\mathscr{Q} \geq 0$ for $G(u,r)$ to be positive for some $u \in [0,1]$. This gives a range of allowed values for r that we can calculate explicitly: $\mathscr{Q}$ is a quadratic polynomial in r^2 that is negative for $r^2 \to \pm\infty$. We can therefore calculate the roots $r_\pm$ of $\mathscr{Q}$, and the allowed range of r is between these roots. When $\mathscr{Q} = 0$, $G(0,r_\pm) = 0$ and $u = 0$ is a maximum of $G(u,r_\pm)$: this implies that the geodesics are at $\theta = \pi/2$ and they are stable to perturbations in θ. It turns out that the roots $r_\pm$ such that $\mathscr{Q}(r_\pm) = 0$ are precisely the radii of the unstable orbits in the equatorial plane, and so the geodesics at constant r are only possible in the range $r_- \leq r \leq r_+$, with $r_\pm$ are given in (3.37).

If we only set $p_z = 0$ we again find that it is not possible to have constant-r geodesics everywhere. Expanding F for large r, one finds that there are no values for b_ψ for which F is positive; therefore for large r there are no constant radius orbits.

3.4 Null geodesics with momentum around the internal torus

Similarly to Section 3.2, we will now investigate null geodesics in 10d that have non-zero momenta around the internal torus and which are timelike geodesics of the 6d metric $(\tilde{g}_6)_{\mu\nu}$ in (3.9) (*not* of the Einstein frame metric). In particular we find that some of these geodesics are stably trapped, and these can then be used in the argument for instability in section 2.2.4 that involves a massive particle.

The equations for these geodesics are given by the two coupled 1d equations of motion in (3.27) where μ is the mass of the particle (in 6d), $\sqrt{H_2/H_1}g_6^{\mu\nu}P_\mu P_\nu = -\mu^2$. The effective potentials are given in (3.28) and (3.29).

3.4.1 Geodesics in the equatorial plane

In an analogous way to section 3.3.2, there are geodesics that stay in the equatorial plane $\theta = \pi/2$ if

$$p_\psi = 0, \qquad \Lambda = p_\phi^2 + (p_z^2 - p_t^2)(Q_1 + Q_2) \tag{3.51}$$

and for simplicity we will investigate the case $p_z = 0$.

In this case the radial equation reduces to

$$\dot{r}^2 + \frac{1}{h^2 f^2} U_r = 0 \tag{3.52}$$

where

$$\frac{1}{p_t^2} U_r = (m^2 - 1)r^4 + (B + m^2(a^2 + Q_2))r^2 + (C + m^2 a^2 Q_2) \tag{3.53}$$

and B, C are given in (3.33) and $m = -\mu/p_t$.

We will briefly discuss whether or not it is possible to have stable trapping in terms of the binding energy. There are two cases:

i) Positive binding energy, $\mu > -p_t$.
Geodesics cannot even exist near infinity so it is only possible to have bound orbits, and the only possibility is that the geodesics are stably trapped. If $b_\phi \leq \xi - m^2 a\sqrt{Q_2/4Q_1}$ the geodesics are trapped in a region which includes the evanescent ergosurface.

ii) Negative binding energy, $\mu \leq -p_t$.
It is not obvious in this case that there are stably trapped geodesics, but in fact there are geodesics that are stably trapped near the evanescent ergosurface; this can be seen by taking $b_\phi \to \infty$. If we do this by leaving p_t finite but taking p_ϕ to be large and negative then the terms involving b_ϕ are much larger than those which depend on m. Since the

potential is then almost the same as for the null geodesics in 6d in section 3.3.2, we know there will be stable trapping for b_ϕ sufficiently large and negative. We also expect there to be other ranges of the parameters that give rise to stable trapping.

It is interesting to observe that, depending on the background parameters, for particles with positive binding energy it can be shown that it is possible to find geodesics in the equatorial plane that are stably trapped in a region that does not include the evanescent ergosurface at $r = 0$. This is in contrast to the results of [141], where it is shown that in a Schwarzschild spacetime in $n+1$ dimensions, if $n > 3$ there are no stable bound orbits. Similarly, around a 5d Myers-Perry black hole there are no bound orbits outside of the event horizon [135], in contrast to the supersymmetric microstate geometries.

3.5 3-charge microstate geometries

Recall now the 3-charge microstate geometries introduced in section 2.3.1. The 10d string frame metric is:

$$\begin{aligned}
ds^2 =& -\frac{1}{h}(dt^2 - dz^2) + \frac{Q_p}{hf}(dt-dz)^2 + hf\Big(\frac{dr^2}{r^2+(\tilde\gamma_1+\tilde\gamma_2)^2\eta} + d\theta^2\Big) \\
&+ h\Big(r^2 + \tilde\gamma_1(\tilde\gamma_1+\tilde\gamma_2)\eta - \frac{(\tilde\gamma_1^2-\tilde\gamma_2^2)\eta Q_1 Q_2\cos^2\theta}{h^2 f^2}\Big)\cos^2\theta d\psi^2 \\
&+ h\Big(r^2 + \tilde\gamma_2(\tilde\gamma_1+\tilde\gamma_2)\eta + \frac{(\tilde\gamma_1^2-\tilde\gamma_2^2)\eta Q_1 Q_2\sin^2\theta}{h^2 f^2}\Big)\sin^2\theta d\phi^2 \\
&+ \frac{Q_p(\tilde\gamma_1+\tilde\gamma_2)^2\eta^2}{hf}(\cos^2\theta d\psi + \sin^2\theta d\phi)^2 \\
&- 2\frac{\sqrt{Q_1Q_2}}{hf}\Big(\tilde\gamma_1\cos^2\theta d\psi + \tilde\gamma_2\sin^2\theta d\phi\Big)(dt-dz) \\
&- 2\frac{(\tilde\gamma_1+\tilde\gamma_2)\eta\sqrt{Q_1Q_2}}{hf}\Big(\cos^2\theta d\psi + \sin^2\theta d\phi\Big)dz + \sqrt{\frac{H_1}{H_2}}\Sigma_{i=1}^4 dx_i^2 \\
=& ds_6^2 + \sqrt{\frac{H_1}{H_2}}\Sigma_{i=1}^4 dx_i^2
\end{aligned} \tag{3.54}$$

where

$$\eta = \frac{Q_1Q_2}{Q_1Q_2 + Q_1Q_p + Q_2Q_p}, \tag{3.55}$$

$$\tilde\gamma_1 = -an,\ \ \tilde\gamma_2 = a(n+1), \tag{3.56}$$

$$f = r^2 + a^2\eta(-n\sin^2\theta + (n+1)\cos^2\theta), \tag{3.57}$$

$$H_1 = 1 + \frac{Q_1}{f}, \;\; H_2 = 1 + \frac{Q_2}{f} \;\text{ and }\; h = \sqrt{H_1 H_2}, \tag{3.58}$$

where $\theta \in [0, \pi/2]$, $r > 0$ and $0 \leq \phi, \psi \leq 2\pi$.

The angular momenta of these geometries are

$$J_\psi = -n n_1 n_5, \qquad J_\phi = (n+1) n_1 n_5. \tag{3.59}$$

The 3-charge solution reduces to the 2-charge supersymmetric microstate geometry in the previous section if we set $n = 0$.

The evanescent ergosurface is given by $f = 0$, where the globally null Killing vector field

$$V = \frac{\partial}{\partial t} + \frac{\partial}{\partial z} \tag{3.60}$$

is orthogonal to the Kaluza-Klein Killing vector field $Z = \partial/\partial z$. This gives the submanifold defined by

$$r^2 = a^2 \eta \left(n \sin^2\theta - (n+1)\cos^2\theta \right)$$

which has topology $S^1 \times S^3$ at constant t [93].

3.5.1 Null geodesics

The Hamilton-Jacobi equation for null geodesics (3.1) in the supersymmetric 3-charge microstate geometries separates into coupled first order equations of motion. We can use the effective potentials to find regions where geodesics are either 'allowed' or 'forbidden' and from this we can see whether there are stable or unstable trapped null geodesics.

We will not restrict to any particular values of the conserved quantities or to any particular submanifold as we did for the 2-charge microstate geometry in section 3.3. The results here are therefore more general, but we do not find exact bounds on the values of the impact parameters for which the different types of trapping occur.

3.5.1.1 Equations of motion

The behaviour of a geodesic depends on its conserved quantities. We will factor out the energy $-p_t$ because we can rescale the affine parameter along the null geodesics so only the ratios of the momenta have any physical meaning. We define the impact parameters as for the 2-charge case:

$$b_\phi = -\frac{p_\phi}{p_t}, \qquad b_\psi = -\frac{p_\psi}{p_t}, \qquad b_z = -\frac{p_z}{p_t}. \tag{3.61}$$

Due to the extra symmetry associated to a conformal Killing tensor that allows one to separate the wave equation, the Hamilton-Jacobi equation for null geodesics also separates. If we substitute the ansatz (3.6) and the inverse of the metric (3.54) (which can be found in ref. [92]) into the Hamilton-Jacobi equation (3.1) we obtain first order equations for $R(r)$ and $\Theta(\theta)$. In the same way as for the 2-charge microstate geometry in section 3.3, this allows us to find the coupled first order equations of motion for $r(\lambda)$ and $\theta(\lambda)$:

$$\dot{r}^2 + \frac{1}{(hf)^2} U_r(r) = 0 \tag{3.62}$$

$$\dot{\theta}^2 + \frac{1}{(hf)^2} U_\theta(\theta) = 0 \tag{3.63}$$

where $\dot{} = \frac{d}{d\lambda}$, λ is an affine parameter along the null geodesic. The effective potentials are

$$\begin{aligned} U_\theta = a^2\eta\Big(-n\sin^2\theta + (n+1)\cos^2\theta\Big)(p_z^2 - p_t^2 + \mu^2) + (Q_1+Q_2)(p_z^2 - p_t^2) \\ - Q_p(p_t+p_z)^2 + \frac{p_\psi^2}{\cos^2\theta} + \frac{p_\phi^2}{\sin^2\theta} - \Lambda \end{aligned} \tag{3.64}$$

$$\begin{aligned} U_r = \frac{p_t^2}{r^2}\Bigg[-\eta r^2\Big(\frac{\sqrt{Q_1Q_2}}{\eta} - \frac{Q_1+Q_2}{\sqrt{Q_1Q_2}}a^2 n(n+1)b_z - a(n+1)b_\phi + anb_\psi\Big)^2 + r^4(r^2+a^2\eta)(b_z^2-1) \\ + (r^2+a^2\eta)\eta\Big(-\sqrt{Q_1Q_2}b_z - a(n+1)b_\psi + anb_\phi\Big)^2 + \frac{\Lambda}{p_t^2}r^2(r^2+a^2\eta) \\ + \frac{\mu^2}{p_t^2}r^2\big(r^4 + (Q_2+a^2\eta)r^2 + Q_2a^2\eta\big)\Bigg] \end{aligned} \tag{3.65}$$

where $\mu^2 = \delta^{ij}q_iq_j$, $i,j = 1\ldots4$, q_i is the conserved momenta around the internal torus and Λ is a constant arising from the separation of variables. Note that the coupling factor $(hf)^{-2}$ is strictly positive so the 'allowed' regions for geodesics only depend on the signs of the effective potentials.

3.5.1.2 Trapped geodesics

We will describe the possibilities for trapping when $\mu = 0$, since these correspond to null geodesics in 6d.

Geodesics can only exist in regions where both of the effective potentials are negative. Therefore there must be some value $\theta \in [0, \pi/2]$ such that U_θ is negative; this implies that

the separation constant Λ must satisfy

$$\Lambda \geq a^2\eta\Big(-n\sin^2\theta+(n+1)\cos^2\theta\Big)(p_z^2-p_t^2)+\frac{p_\psi^2}{\cos^2\theta_0}+\frac{p_\phi^2}{\sin^2\theta_0} \tag{3.66}$$

for some $\theta_0 \in [0,\pi/2]$ so that the geodesic can exist at least at θ_0.

If we let $x=\cos\theta$ we can write $x^2(1-x^2)U_\theta(x)$ as a cubic polynomial in x^2 which is positive at $x=0$ and $x=1$. Therefore there can be at most one region in the range $0\leq x\leq 1$ where geodesics are allowed; it is not possible to have disjoint regions for the geodesics for the same values of the impact parameters.

Assuming we choose Λ such that (3.66) holds, the problem is then to find the regions in r where geodesics can propagate, i.e. where U_r is negative. To do this we consider the terms in square brackets in U_r in (3.65), which give a cubic polynomial in r^2 and so can have up to 3 positive roots. We write this expression as

$$W_r=(b_z^2-1)r^6+\alpha r^4+\beta r^2+\gamma \tag{3.67}$$

for some α,β,γ that are given by the coefficients in (3.65) and depend on b_ϕ, b_ψ, b_z and Λ/p_t^2, although it is useful to note that for any values of the parameters $\gamma\geq 0$. As W_r depends on these parameters, whether or not there are trapped geodesics also depends on b_ϕ, b_ψ, b_z and Λ/p_t^2.

If $p_z^2-p_t^2<0$, which in particular includes geodesics with zero Kaluza-Klein momentum, W_r can have up to 3 positive real roots: the potential is positive at $r=0$ because $\gamma\geq 0$ and $p_z^2-p_t^2<0$ implies that $W_r(r)\to-\infty$ as $r\to\infty$. We therefore have the following possibilities for geodesics:

i) W_r has 3 positive real roots, $\{r_-, r_0, r_+\}$.
If all the roots are distinct, $r_-<r_0<r_+$, then for the values of $b_\phi, b_\psi, p_z/p_t$ and Λ/p_t^2 that allow for such roots there are stably trapped geodesics in the region $r_-\leq r\leq r_0$ as well as geodesics only allowed in $r\geq r_+$ that can escape to infinity.

If $r_0=r_+$ there is unstable trapping at r_0. This might occur for several different values of $b_\phi, b_\psi, \frac{p_z}{p_t}$ and Λ/p_t^2; as r_0 depends on these parameters this unstable trapping could occur at various values of r.

If $r_-=r_0<r_+$ there is stable trapping with the geodesics localised exactly at r_0.

ii) W_r has 1 real root, r_0.
For these values of b_ϕ, b_ψ, b_z and Λ/p_t^2 there are only geodesics in the region $r\geq r_0$ which can escape out to infinity and therefore no trapping.

Figure 3.5 plots W_r in these cases with appropriate values of b_ϕ, b_ψ, p_z/p_t and Λ/p_t^2. On this plot the 'allowed' and 'forbidden' regions for these geodesics are clear and it is easy to see whether or not the geodesics are trapped.

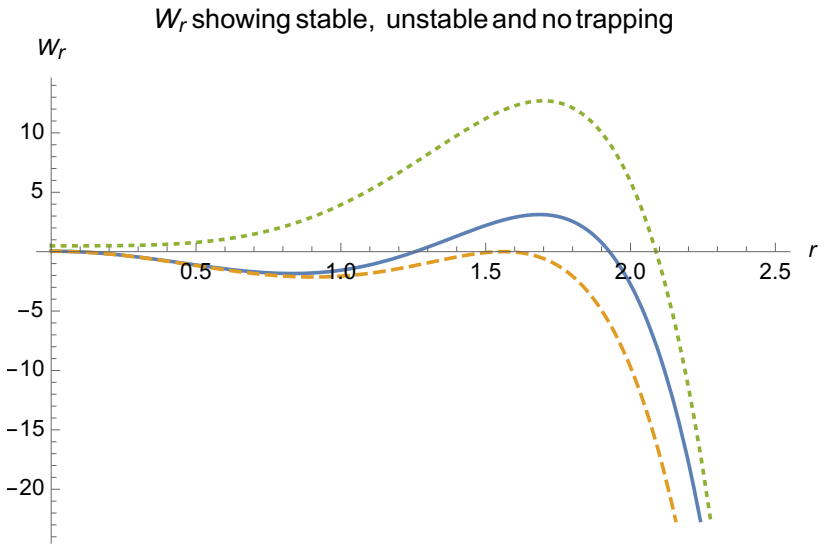


Fig. 3.5 Plot of W_r showing that for different values of the impact parameters there is stable trapping (solid blue line) , unstable trapping (dashed orange line) or no trapping (dotted green line).

If $p_z^2 - p_t^2 > 0$ there can either be two distinct positive roots, one double root or W_r is always positive. If there are two roots $r_- < r_+$ the geodesics are stably trapped in the region $r_- \leq r \leq r_+$. If $r_- = r_+$ the geodesics are stably trapped at exactly r_-. The only other possibility is that there are no real roots so W_r is always strictly positive and no geodesics can exist.

Finally, if $p_z^2 = p_t^2$ there are stably trapped geodesics if b_ϕ, b_ψ, b_z and Λ/p_t^2 are such that $\alpha > 0$, $\beta^2 \geq 4\alpha\gamma$ and $\beta \leq 0$. If instead b_ϕ, b_ψ, $b_z$and Λ/p_t^2 give $\alpha \leq 0$, the geodesics only exist in the region $r \geq r_0$ for some r_0; if neither of these is the case then there are no geodesics with those values of the parameters.

Note that the above does not guarantee such geodesics exist; for all of these cases to happen it must be possible to find roots of W_r for some values of the impact parameters subject to the restriction on Λ in (3.66) which ensures there is an 'allowed' value of θ for the geodesic. Even in the cases where there is stable trapping there is never more than one region in which the geodesics can be trapped for a given set of parameters. However, we know from the results of chapter 2 that there is a stably trapped geodesic through each point of the spacetime (with tangent V), so at least some of the above geodesics do exist.

3.5.1.3 Trapping and the evanescent ergosurface

The same argument from chapter 2, given in section 3.3 for the 2-charge case, for the stable trapping of the zero-energy null geodesics also applies to these 3-charge microstate geometries. Therefore there are zero-energy null geodesics with tangent vector V that are stably trapped on the evanescent ergosurface.

This can be seen using the equations of motion and the effective potentials by setting $p_t = 0$ and $p_z = 0$ in (3.65) and (3.64). In this case there is a minimum of U_θ at θ_0 where $\tan^2\theta_0 = p_\phi/p_\psi$, and by an appropriate choice of Λ these geodesics are localised exactly at θ_0. Substituting this into the radial potential, we find that there is a minimum at $r_0^2 = a^2\eta|(n+1)\cos^2\theta - n\sin^2\theta|$ and $U_r'(r_0) = 0$ so the geodesics are stably trapped exactly on the evanescent ergosurface.

In general, the trapped region does not always include the evanescent ergosurface. Indeed, it was shown in chapter 2 that there are stably trapped geodesics with tangent V through every point in the spacetime. If these geodesics have non-zero energy they are not localised on the evanescent ergosurface, and these have $p_z \neq 0$ in general.

3.5.2 Penrose process

The Penrose process [45] is a method of extracting energy from a Kerr black hole. It is possible because there is an ergoregion, where the Killing vector field T, which is timelike at infinity, becomes spacelike outside of the event horizon. This means that the energy of a physical particle with momentum P (P is future-directed and causal), which is given by $E = -T\cdot P$, can become negative in the ergoregion. In the Kerr spacetime, it is possible to set up a situation in which a particle with positive energy in the ergoregion decays into two other particles that follow geodesics, one with negative energy that falls into the black hole and one with energy greater than that of the initial particle that escapes back out to infinity, thus extracting energy from the black hole.

There is also an ergoregion in the 3-charge microstates geometries, so it interesting to ask whether a similar process can happen here. We have only considered geodesics that are null in the full 10 dimensions; within this class of geodesics it is not possible to replicate the Penrose process exactly to find one particle that decays into two other particles since momentum conservation (in 10d) would require writing one null vector as a sum of two non-parallel null vectors, which is not possible.

Instead we will look for two particles sent in from infinity with positive energy interacting within the ergoregion to produce two other particles, one with negative energy that stays trapped in some region and one that escapes to infinity with energy greater than the sum of

the energies of the two initial particles. To set up the initial conditions for this to happen, the particles that interact should be able to fall in from infinity, and all the particles should follow geodesics.

Note that the ergoregion is defined to be the region where $T^2 > 0$, which is the region where $f < Q_p$ (note that the evanescent ergosurface lies within the ergoregion). To simplify matters, we can restrict attention to geodesics that remain in the equatorial plane $\theta = \pi/2$ by setting

$$p_\psi = 0, \qquad \Lambda = -a^2\eta n(p_z^2 - p_t^2 + \mu^2) + (Q_1+Q_2)(p_z^2 - p_t^2) - Q_p(p_t+p_z)^2 + p_\phi^2 \quad (3.68)$$

in equation (3.64). If we substitute this into (3.65) we find the radial effective potential for geodesics in the equatorial plane.

For our Penrose-like process, let the two particles with positive energy that interact and could come in from infinity have momenta P and Q. Let the particle with negative energy that becomes trapped have momentum R and the particle that escapes out to infinity with greater energy than the initial energy have momentum S. By momentum conservation, we must have

$$P + Q = R + S. \quad (3.69)$$

In this example of energy extraction, we will set $a = 1, Q_1 = 1 = Q_2$ and $n = 1$. Let $\mu_P = \sum_{i=1}^4 q_i^2$ be the sum of the components of the momentum around the internal torus and E_P (for example) be the energy of the particle with momentum P, so $E_P = -T \cdot P = -P_t$ and so by (3.69), $E_P + E_Q = E_R + E_S$.

All of P, Q, R and S must be future-directed and causal. Recall that the Killing vector field $V = \partial/\partial t + \partial/\partial z$ is future-directed and globally null, so the momentum of a physical particle must satisfy $V \cdot P \leq 0$, i.e. $P_t + P_z \leq 0$ to ensure it is future-directed.

Since all the geodesics are in the equatorial plane, the $\theta, \psi-$components of the momenta are zero. From the equation for geodesics (3.62) we can find the radial components of the momenta in terms of the conserved $t, \phi, z-$components:

$$(P^r)^2 = \frac{-U_r(r)}{h^2 f^2}. \quad (3.70)$$

In the equatorial plane, the ergoregion is given by $r^2 < 6/5$. We will assume that the process happens well inside the ergoregion, at $r = 0.2$. We are free to pick most of the components of the momenta then use momentum conservation and the equation for the radial component to find the other. We do this by specifying the conserved components of P (particle follows a geodesic that comes in from infinity), R, the particle with negative

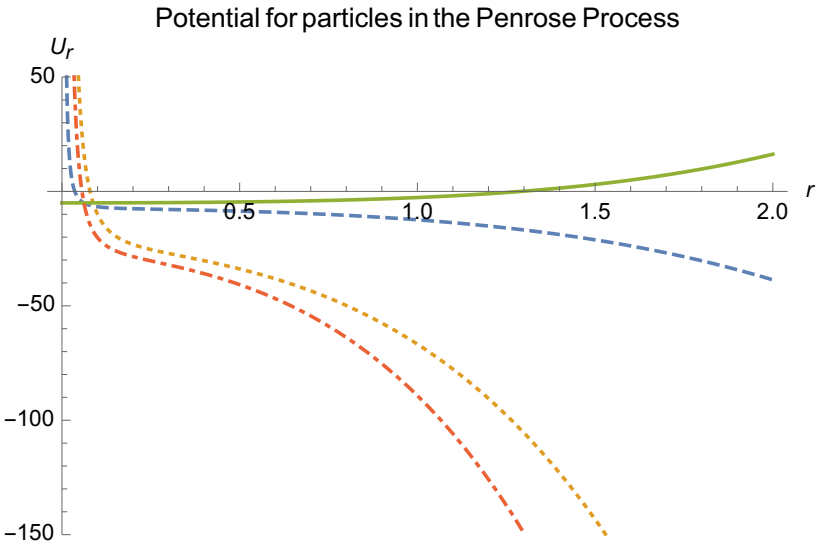


Fig. 3.6 The potentials for the particles above: the green solid line is for the particle R that becomes stably trapped with negative energy, while the blue dashed, orange dotted and red dot-dashed are for P, Q and S respectively.

energy that becomes trapped and all but one of the conserved components of S (the other is calculated from the condition arising from (3.69) and (3.70)). Equation (3.69) then gives Q.

We can do this with particles that are null in 6d, *i.e.* with $\mu = 0$. Values of the momenta of the particles that allow such a process are as follows:

$$\begin{aligned} P_t &= -1, & P_\phi &= -\frac{1}{2}, & P_z &= 0 & &\Rightarrow hf|_{r=0.2} P^r = 2.75\ldots \\ Q_t &= -3, & Q_\phi &= 2.02\ldots, & Q_z &= 0 & &\Rightarrow hf|_{r=0.2} Q^r = 4.82\ldots \\ R_t &= 0.1, & R_\phi &= -1, & R_z &= -1 & &\Rightarrow hf|_{r=0.2} R^r = 2.23\ldots \\ S_t &= -4.1, & S_\phi &= 2.52\ldots, & S_z &= 1 & &\Rightarrow hf|_{r=0.2} S^r = 5.34\ldots. \end{aligned} \tag{3.71}$$

The potentials $U_r(r)$ corresponding to each of these particles are shown in Figure 3.6.

Note that the potential for R shows that the particle is stably trapped in some region since it is positive at infinity, and that $E_R = -0.1 < 0$ so this particle is indeed trapped with negative energy. The particle with momentum S has energy $E_S > E_P + E_Q$ as required and it can be seen that it can escape back out to infinity.

It is interesting to observe that, despite the ergoregion and the existence of timelike geodesics with negative energy within it, there is no Friedman instability [142]. One might expect that, due to the ergoregion, there could be solutions of the Klein-Gordon equation which are not uniformly bounded and indeed even grow. Such an instability was found to be present in certain *non*-supersymmetric microstate geometries in [98]. However, in the

supersymmetric microstate geometries, this should be prevented for the same reason that all solutions of the massless wave equation are bounded in these geometries despite the presence of the ergoregion [105]: there is a globally null Killing vector field V that provides us with a conserved, but degenerate, energy that we can use to bound the non-degenerate energy.

3.6 Implications for quasinormal modes

Recall from chapter 2 that we can find quasinormal mode solutions (see 1.2.1) related to trapped geodesics using the geometric optics approximation, from which we can learn something about the rate of decay of solutions to the wave equation. The various types of trapping have different effects on the rate of decay. We summarise them briefly here and then discuss the implications of the geodesics we have found in this chapter. Note that ω is the quasinormal frequency, and the solution is in the form

$$\Phi(t,r,\theta,\phi,\psi,z) = e^{-i\omega t+im_\psi\psi+im_\phi\phi+i\lambda z}e^{iq_jz^j}\Phi_r(r)\Phi_\theta(\theta). \tag{3.72}$$

Around a Kerr black hole there are unstable circular photon orbits for which the geodesics have non-zero energy. In the limit $\ell\to\infty$ there are quasinormal mode solutions localised near these null geodesics with $\omega_R = O(\ell)$ and $\omega_I = O(1)$, $\omega_I < 0$ [52].

In ultracompact neutron stars, which are fluid objects with a photon sphere but no horizon, and in Kerr-AdS there are stably trapped null geodesics with non-zero energy. For Kerr-Ads this leads to quasinormal modes with $\omega_R = O(\ell)$ and $\omega_I = O(e^{-\gamma\ell})$ for some positive constant γ and $\omega_I < 0$ [143, 101]. This stable trapping is the reason that the rate of decay of solutions to the wave equation is very slow in both Kerr-AdS and ultracompact neutron stars, which was proved in [32] and [33] respectively.

In the supersymmetric 2 and 3-charge microstate geometries that we discussed in the previous section, in chapter 2 we found quasinormal modes localised near the zero-energy null geodesics that are stably trapped on the evanescent ergosurface that have $\omega_R = O(1)$ (this is because the geodesics have zero energy) and $\omega_I = O(e^{-2\ell\log\ell})$, $\omega_I < 0$. There are also quasinormal modes localised near the geodesics with tangent V that are not on the evanescent ergosurface and thus have non-zero momentum around the Kaluza-Klein direction. These have $(\omega_R - \lambda) = O(1)$ in (3.72) and $\omega_I = O(e^{-\ell\log\ell})$, $\omega_I < 0$. It was shown in chapter 2 and proved rigorously in [105] that this leads to even slower decay for solutions of the wave equation than in the cases where the stably trapped geodesics have non-zero energy.

We now summarise the other geodesics we have found in the supersymmetric microstate geometries and the implications for quasinormal modes. In all the following we assume

$\omega_I < 0$ since the modes eventually disperse out to infinity, and that $\alpha_i > 0$ are some positive constants. Some of the possible cases are as follows:

- From section 3.3.2 we know there are geodesics with zero momentum around the internal torus ($q_i = 0$ and also $\lambda = 0$ in (3.72)) that have non-zero energy and can either be stably or unstably trapped. From the discussion above, we expect there to be quasinormal modes localised near the stably trapped null geodesics with $\omega_R = O(\ell)$, $\omega_I = O(e^{-\alpha_1 \ell})$. In case 1 of section 3.3.2 the stably trapped geodesics can be either pro- or retrograde; this implies that m_ϕ/ω can be either positive or negative. However, in case 2 all the stably trapped null geodesics are retrograde, so we expect the quasinormal modes to have $m_\phi/\omega < 0$.

  We also expect there to be quasinormal modes with $\omega_R = O(\ell)$, $\omega_I = O(1)$ localised near the unstably trapped null geodesics. This also applies for the geodesics at constant r in section 3.3.4, but we expect these quasinormal modes to be localised near a single value of r as opposed to some finite region. In case 1 of section 3.3.2 the unstably trapped geodesics are retrograde so for these quasinormal modes we expect $m_\phi/\omega < 0$, whilst in case 2 these geodesics can be either pro-or retrograde so m_ϕ/ω can have either sign.

- If the momentum around the internal torus is non-zero, we showed in section 3.2 that we can always find stably trapped null geodesics in 10d. In terms of quasinormal modes, this means that there are solutions with $q_i \neq 0$ in (3.72) that are localised in some finite region. In section 3.4 we found stable and unstable trapped null geodesics (in 10d); we might expect that there are quasinormal modes localised near these with $\omega_R = O(\ell)$ and $\omega_I = O(e^{-\alpha_2 \ell})$ or $\omega_I = O(1)$ respectively.

We expect analogous consequences due to the stably and unstably trapped null geodesics found in section 3.5 for quasinormal modes in the 3-charge supersymmetric microstate geometries.

3.7 Discussion

In section 3.2 we verified the existence of the type of massive particle orbit needed for the argument for instability, so it is interesting to review the possible endpoint for such an instability. It was suggested in chapter 2 that the endpoint could be an almost supersymmetric black hole with less angular momenta but the same charges as the microstate geometry. The endpoint of the instability was investigated in more detail in [144]. Here, entropic

arguments and a toy model were used to suggest that the instability mechanism causes the angular momentum to decrease until the density of states is largest and the final state is then a microstate geometry which is not smooth. At this point, stringy corrections to the supergravity approximation become important and these could in fact stabilize the instability, although at this scale the geometry is indistinguishable from a black hole in the supergravity approximation.

We have mostly considered the most symmetric supersymmetric microstate geometries. In recent years, many other supersymmetric microstate geometries with fewer isometries have been constructed, such as bubbling geometries (see for example [94]) and microstate geometries with a long AdS_2 throat [113] among others. If these geometries are supersymmetric we know from chapter 2 that they will have a stably trapped null geodesic through every point in the spacetime. Providing these microstate geometries satisfy the necessary conditions of section 3.2, the general results here will apply to these geometries and show that it is possible to have a massive particle with positive binding energy arbitrarily far out that does not escape to infinity in these spacetimes. This will imply that there are slowly decaying quasinormal modes localised near these geodesics, which could lead to a non-linear instability. Note that for the type of instability to be the same as for the most symmetrical microstate geometries there must be an evanescent ergosurface, which is the case for the examples mentioned here.

It is interesting to compare these geodesics to the geodesics around a BMPV black hole [136]. As mentioned before, there are no bound orbits of either massive or massless particles that exist only in the outer region of the BMPV black hole black hole; our analysis shows that this is different to the supersymmetric microstate geometries and also to more general microstate geometries satisfying the conditions in section 3.2. According to the fuzzball proposal, the black hole is supposed to be described by an ensemble of microstate geometries. Once a measurement is made the state collapses to one particular microstate. Observing a test particle near the black hole would be an example of such a measurement; after this the state would possibly collapse to a microstate where there are bound orbits far away from the black hole, and the orbits of particles would look very different to those that are expected from the black hole. The fuzzball proposal is only conjectured to apply to supersymmetric black holes, but if there was an analogous description for astrophysical black holes in terms of an ensemble of microstate geometries, we might also expect some of the orbits of objects (such as stars) around them to be changed in a similar way to above.

Chapter 4

Strong Cosmic Censorship in de Sitter space

This chapter is essentially [145], which was done in collaboration with Oscar Dias, Harvey Reall and Jorge Santos.

4.1 Introduction

The strong cosmic censorship conjecture [64] asserts that, for generic asymptotically flat initial data for Einstein's equation, the maximal Cauchy development is inextendible in a suitable sense, *i.e.*, Cauchy horizons do not form. We discussed this in some detail in section 1.3 of the Introduction for asymptotically flat black holes, where it is well-known that the presence of a Cauchy horizon inside a charged or rotating black hole does not constitute a violation of strong cosmic censorship because of an infinite blue shift at the Cauchy horizon, which renders it unstable and therefore non-generic [146, 68, 73, 75]. Some time ago, it was observed that the mechanism behind this instability is weaker when the cosmological constant Λ is positive [82]. This is because there is a redshift of late time perturbations entering the black hole, arising from the fact that these perturbations have to climb out of the gravitational potential well associated with the cosmological horizon. In this chapter we will look at strong cosmic censorship in spacetimes with a positive cosmologival constant $\Lambda > 0$.

Early calculations (reviewed in Ref. [147]) indicated that, for charged or rotating black holes sufficiently close to extremality, a violation of strong cosmic censorship would indeed be possible with positive Λ. However, subsequent work argued that the decay of scalar field perturbations outside the black hole was still sufficiently slow to ensure that the gradient

of the scalar field would diverge at the Cauchy horizon, with backreaction then causing a curvature divergence, and so strong cosmic censorship would be respected [83].

Recent interest in this topic has been stimulated by the recognition that by 'inextendible' we can mean as a *weak* solution of the Einstein equations, leading to Christodoulou's formulation of strong cosmic censorship [76] discussed in section 1.3.

For $\Lambda = 0$, it seems very likely that this conjecture is true (see Ref. [70] for a detailed discussion). However, with $\Lambda > 0$ it was observed in [70] that calculations similar to those of [83] suggest that Christodoulou's version of the strong cosmic censorship conjecture may be false for near-extremal Reissner-Nordström-de Sitter and Kerr-de Sitter black holes. Very recently, Ref. [148] has presented compelling evidence that this is indeed the case for near-extremal Reissner-Nordström de Sitter. The argument is based on recent mathematical developments in the study of black holes with positive Λ, as we will now explain.

The behaviour of perturbations at the Cauchy horizon depends on the rate of decay of perturbations along the event horizon [149]. Faster decay along the event horizon gives a milder instability of the Cauchy horizon. With positive cosmological constant, it has been proved that perturbations decay *exponentially* along the event horizon. Specifically, for massless scalar field perturbations Φ of Reissner-Nordström-de Sitter, or slowly rotating Kerr-de Sitter, black holes it has been proved that, there exist constants Φ_0 and $C, \alpha > 0$ such that, outside the black hole [35–40]

$$|\Phi - \Phi_0| \leq Ce^{-\alpha t} \tag{4.1}$$

where t labels a foliation by spacelike hypersurfaces that extend from the future event horizon to the future cosmological horizon (e.g. the surface Σ of Fig. 4.1), with the hypersurfaces related by the time translation symmetry of the black hole. The best constant α for which (4.1) is true is called the spectral gap. The spectral gap can be determined by looking at the most slowly decaying *quasinormal modes* of the black hole: α is the largest number such that $\alpha \leq -\mathrm{Im}(\omega)$ for all quasinormal frequencies ω.

If α is known then one can hope to determine the behaviour of generic perturbations at the Cauchy horizon and hence ascertain whether or not strong cosmic censorship is violated. And α can be determined by looking at quasinormal modes of the black hole. This is what was done in Ref. [148] for Reissner-Nordström-de Sitter black holes. By numerically finding the most slowly decaying quasinormal modes, the value of α was determined. For black holes sufficiently close to extremality, the value of α was sufficiently large to indicate that, when nonlinearities are included (e.g. using results of Ref. [150]), it would be possible to extend the solution across the Cauchy horizon as a weak solution of the equations of motion, in violation of the strong cosmic censorship conjecture.

It was suggested in [151] that a charged scalar field may actually save strong cosmic censorship for Reissner-Nordström-de Sitter. However this is not the case, since it was shown in [152] that although the charged scalar field improves the situation, it is always possible to find a black hole close enough to extremality that violates strong cosmic censorship.

The scalar field is a toy model for the linearised coupled electromagnetic and gravitational perturbations. These were studied in [153], where it was shown that these actually violate strong cosmic censorship in Reissner-Nordström-de Sitter in a much worse way than the scalar field: perturbations can be in C^r, for *any* r, for a black hole that is both large enough and sufficiently close to extremality.

Quasinormal modes only determine the late time behaviour for *smooth* initial data. If instead we require only that the initial data has finite local energy on the initial hypersurface, then the solution to the wave equation generically has infinite energy at the Cauchy horizon [154]. This implies that a slightly modified version of strong cosmic censorship, in which we start with some minimal allowed regularity (finite energy) but lose this at the Cauchy horizon, does in fact hold for Reissner-Nordström-de Sitter.

It was pointed out in [153] that the type of regularity of the initial data provides the explanation for the claim in [83] that strong cosmic censorship was respected in Reissner-Nordström-de Sitter: they (unknowingly) considered initial data that was not smooth, which led to the required blow up at the Cauchy horizon.

All of the evidence above for a violation of SCC in Reissner-Norström-de Sitter black holes comes from the *linear* version of SCC, the study of wave equations on a fixed background. Although this is usually thought to give a reasonable indication of the full non-linear results, one might wonder whether non-linear effects could somehow save SCC. However, the nonlinear Einstein-Maxwell system with a positive cosmological constant, in spherical symmetry, was evolved numerically in [155]. The results of this were in agreement with the linear problem, providing more evidence that for near-extremal black holes SCC is violated.

Reissner-Nordström-de Sitter black holes are not very relevant physically. However, they are often viewed as a toy model for the much more physical case of Kerr-de Sitter black holes. The massless scalar field can be viewed as a toy model for linearized gravitational perturbations. So the results of Ref. [148] suggest that maybe there is a violation of strong cosmic censorship for nearly extremal Kerr-de Sitter black holes in vacuum. Indeed this was conjectured in Ref. [70]. That is what we will investigate in this chapter.

Our approach is the following. We will study linear perturbations of a non-extremal Kerr-de Sitter black hole. These perturbations could be either a massless scalar field or linearized gravitational perturbations. Such a linear perturbation will source a second order metric perturbation. The linear perturbation will be continuous but not necessarily differentiable

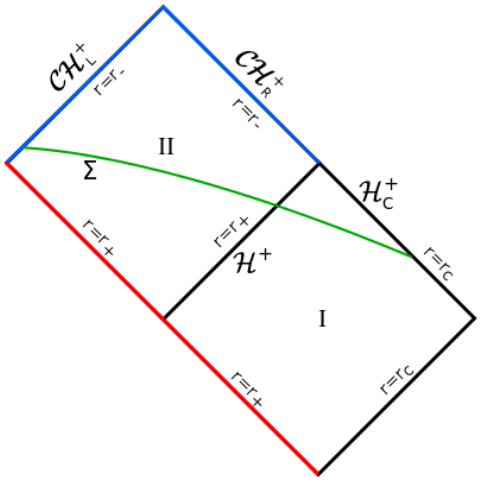


Fig. 4.1 Penrose diagram for Kerr-de Sitter. The event and cosmological horizons are $\mathscr{H}^+$ and $\mathscr{H}_C^+$, respectively, the blue lines are the left and right Cauchy horizons $\mathscr{CH}_{L,R}^+$, the green line is a spacelike hypersurface extending from the cosmological horizon across the event horizon and the left Cauchy horizon. Quasinormal modes blow up along the white hole horizon (red line) and also along the past cosmological horizon.

at the Cauchy horizon. However, in order to extend beyond the Cauchy horizon, the linear solution needs to be sufficiently regular that the equation of motion for the second order perturbation can be satisfied in a weak sense at the Cauchy horizon. As we will explain, this leads to the criterion that the scalar field, or linearized metric perturbation, must have a locally square integrable derivative, *i.e.*, it should belong to the Sobolev space $H^1_{\rm loc}$. This was also the criterion used in Ref. [148].

Consider a scalar field quasinormal mode in a non-extremal Kerr-de Sitter spacetime. Such a solution has definite frequency and satisfies ingoing boundary conditions at the future event horizon $\mathscr{H}^+$, and outgoing boundary conditions at the future cosmological horizon $\mathscr{H}_C^+$ (see Fig. 4.1). Working in coordinates regular across $\mathscr{H}^+$, a quasinormal mode can be analytically continued into the black hole interior (region II of Fig. 4.1). We will determine how such a quasinormal mode behaves at the Cauchy horizon $\mathscr{CH}_R^+$ of Fig. 4.1. It is straightforward to show that it belongs to $H^1_{\rm loc}$ if, and only if, minus the imaginary part of the quasinormal frequency exceeds a certain value, *i.e.*, the mode decays fast enough.

We will use geometric optics and numerics to show that there always exist “photon sphere" quasinormal modes whose decay is slow enough that, when continued inside the black hole, they do not belong to $H^1_{\rm loc}$ at the Cauchy horizon $\mathscr{CH}_R^+$. We can now prove that strong cosmic censorship is respected as follows. Assume that one is given initial data on the surface Σ shown in Fig. 4.1, for a linearized perturbation which belongs to $H^1_{\rm loc}$ at $\mathscr{CH}_R^+$.

Now “perturb this perturbation" by adding the initial data for our quasinormal mode, with an arbitrary amplitude. This produces a new perturbation which does not belong to $H^1_{\rm loc}$ at $\mathscr{C}\mathscr{H}^+_R$. Hence a *generic* perturbation[1] does not belong to $H^1_{\rm loc}$ and so it cannot be extended beyond $\mathscr{C}\mathscr{H}^+_R$ consistently with the equations of motion. Hence strong cosmic censorship is respected.[2]

For linearized gravitational perturbations, we exploit the fact that there exists a gauge invariant component of the Weyl tensor which satisfies a decoupled equation of motion. If the linearized metric perturbation belongs to $H^1_{\rm loc}$ in some gauge then the blow up of this Weyl component at $\mathscr{C}\mathscr{H}^+_R$ cannot exceed a certain rate[3]. However, for some photon sphere quasinormal modes we find that the blow up does exceed this rate. This means that there exists no gauge in which the linearized metric perturbation is in $H^1_{\rm loc}$. This provides evidence to suggest that strong cosmic censorship is respected by gravitational perturbations of any non-extremal Kerr-de Sitter black hole.

4.2 Weak solutions

We will be discussing linear perturbations which are continuous, but not necessarily differentiable, at the Cauchy horizon. The fundamental question that needs to be addressed is whether there is any sense in which such a perturbation can satisfy the equations of motion at the Cauchy horizon. Moreover, we are primarily interested in answering this question for *nonlinear* perturbations. We will explain why this leads to the condition that linear perturbations should belong to $H^1_{\rm loc}$.

Consider a scalar field Φ satisfying $\Box\Phi = 0$. Treat this as a first order perturbation, sourcing a second order metric perturbation $h^{(2)}_{\mu\nu}$. Then $h^{(2)}_{\mu\nu}$ will satisfy

$$\mathscr{L} h^{(2)}_{\mu\nu} = 8\pi T_{\mu\nu}[\Phi]\,, \tag{4.2}$$

where $\mathscr{L}$ is a certain second order differential operator and $T_{\mu\nu}[\Phi]$ is the energy momentum of the scalar field. Now assume that Φ and $h^{(2)}_{\mu\nu}$ are not necessarily continuously differentiable. One can still make sense of the above equation by multiplying by a smooth, compactly

[1] See the paragraph after Theorem 1.1 of [154] for more detail on what we mean by “generic".

[2] Note that we do not need to assume the validity of equation (4.1), which is just as well because (4.1) has been established only for slowly rotating black holes.

[3] Technically, this Weyl component must belong to $H^{-1}_{\rm loc}$ in a regular tetrad. $H^{-1}_{\rm loc}$ is the dual space of $H^1_{\rm loc}$: it consists of all functions f such that $(1+|x|)^{-1/2}\hat{f}(x) \in L^2$, where $\hat{f}(x)$ is the Fourier transform of f (or f multiplied by an arbitrary smotth, compactly supported test function).

supported, symmetric tensor, $\psi^{\mu\nu}$ and integrating by parts:[4]

$$\int d^4x\sqrt{-g}\left(h^{(2)}_{\mu\nu}\mathscr{L}^\dagger\psi^{\mu\nu} - 8\pi\psi^{\mu\nu}T_{\mu\nu}\right) = 0\,, \tag{4.3}$$

where $\mathscr{L}^\dagger$ is the adjoint of $\mathscr{L}$ arising from the integration by parts. If this equation is satisfied for *any* smooth, compactly supported symmetric $\psi^{\mu\nu}$ then we say that we have a *weak solution* of (4.2). In order for this equation to make sense, the terms involving the scalar field must be finite, which is guaranteed by demanding that Φ belongs to $H^1_{\rm loc}$. This is the space of functions defined by the condition that $\psi(\Phi^2 + \partial_\mu\Phi\partial_\mu\Phi)$ is integrable, for any smooth compactly supported function ψ.

Similarly, if one starts from a linearized gravitational perturbation $h_{\mu\nu}$ one can consider the second order perturbation $h^{(2)}_{\mu\nu}$ sourced by the linear perturbation. This satisfies an equation analogous to (4.2) where the RHS is quadratic in first derivatives of $h_{\mu\nu}$. So repeating the above argument, the minimum regularity required of $h_{\mu\nu}$ in order for the equation for $h^{(2)}_{\mu\nu}$ to be satisfied weakly is that, in some gauge, $h_{\mu\nu}$ should belong to $H^1_{\rm loc}$.

We can relate this to the criterion for weak solutions of the full nonlinear vacuum Einstein equation. Applying the above procedure to the Einstein equation results in the criterion that, in some chart, the Christoffel symbols should be locally square integrable [76]. In such a chart, perform a perturbative expansion of the metric $g_{\mu\nu} = \bar{g}_{\mu\nu} + h_{\mu\nu} + h^{(2)}_{\mu\nu} + \ldots$ and consider the integral of the sum of squares of the Christoffel symbols. At first order this will give terms linear in $h_{\mu\nu}$ and its first derivative. So at first order the minimum regularity required is that $h_{\mu\nu}$ and its first derivative be integrable. However, at second order, terms quadratic in first derivatives $h_{\mu\nu}$ will arise, and so we will need first derivatives of $h_{\mu\nu}$ to be square integrable and hence we will need $h_{\mu\nu}$ to belong to $H^1_{\rm loc}$. Continuing to higher orders does not give anything new because all terms are at most quadratic in first derivatives of $h_{\mu\nu}$.

4.3 Kerr-de Sitter

4.3.1 Coordinates

We will write the Kerr-de Sitter metric [156] as follows [157]

$$ds^2 = \rho^2\left[\frac{\mathrm{d}r^2}{\Delta_r} + \frac{\mathrm{d}\chi^2}{\Delta_\chi}\right] + \frac{1}{\rho^2\Xi^2}\left[\Delta_\chi\left(\mathrm{d}t - \frac{\sigma_r}{a}\mathrm{d}\phi\right)^2 - \Delta_r\left(\mathrm{d}t - \frac{\sigma_\chi}{a}\mathrm{d}\phi\right)^2\right] \tag{4.4}$$

[4]The test function $\psi^{\mu\nu}$ allows the integration to be done in a finite volume and permits integrating by parts without introducing boundary terms.

where

$$\sigma_r = a^2 + r^2\,, \qquad \sigma_\chi = a^2 - \chi^2\,, \qquad \rho^2 = r^2 + \chi^2\,, \qquad \Xi = 1 + \frac{a^2}{L^2} \tag{4.5}$$

and

$$\Delta_r = \sigma_r \left(1 - \frac{r^2}{L^2}\right) - 2Mr\,, \qquad \Delta_\chi = \sigma_\chi \left(1 + \frac{\chi^2}{L^2}\right)\,, \qquad \Lambda = \frac{3}{L^2}\,. \tag{4.6}$$

In these coordinates, $\phi \in [0, 2\pi)$ and $\chi \in [-|a|, |a|]$. It is convenient to define

$$\Omega(r) = \frac{a}{r^2 + a^2}\,. \tag{4.7}$$

We assume that the solution describes a non-extremal black hole, which implies that there are three real positive roots of Δ_r, satisfying $r_- < r_+ < r_c$. These correspond to the Cauchy horizon, event horizon and cosmological horizon, respectively. The angular velocities of the horizons will be denoted by

$$\Omega_- = \Omega(r_-), \qquad \Omega_+ = \Omega(r_+), \qquad \Omega_c = \Omega(r_c). \tag{4.8}$$

Starting from the above metric with $r_+ < r < r_c$, which we call region I (see Fig. 4.1), we define ingoing coordinates (v, r, χ, ϕ') as follows:

$$\mathrm{d}t = \mathrm{d}v - \frac{\Xi\,\sigma_r}{\Delta_r}\mathrm{d}r \qquad \mathrm{d}\phi = \mathrm{d}\phi' - \frac{a\,\Xi}{\Delta_r}\mathrm{d}r\,. \tag{4.9}$$

In the ingoing coordinates, we can extend across $r = r_+$ into a new region, region II (see Fig. 4.1), with $r_- < r < r_+$. In the new coordinates $g_{rr} = 0$ so $\partial/\partial r$ is globally null. In fact $\partial/\partial r$ is also geodesic and shear free: it is one of the repeated principal null directions of the solution; $-\partial/\partial r$ is tangent to ingoing null geodesics.

In region II we can re-introduce the original coordinates (t, r, χ, ϕ) using (4.9). The metric in these coordinates takes the same form as (4.4). Now, in region II, we introduce outgoing coordinates (u, r, χ, ϕ'') defined by

$$\mathrm{d}t = \mathrm{d}u + \frac{\Xi\,\sigma_r}{\Delta_r}\mathrm{d}r \qquad \mathrm{d}\phi = \mathrm{d}\phi'' + \frac{a\,\Xi}{\Delta_r}\mathrm{d}r\,. \tag{4.10}$$

This lets us analytically continue the metric across the "right" Cauchy horizon $\mathscr{CH}^+_R$ in region II into a new region with $r < r_-$. In these coordinates, $-\partial/\partial r$ is null, geodesic and shear free, and future-directed. It is the second repeated principal null direction of the

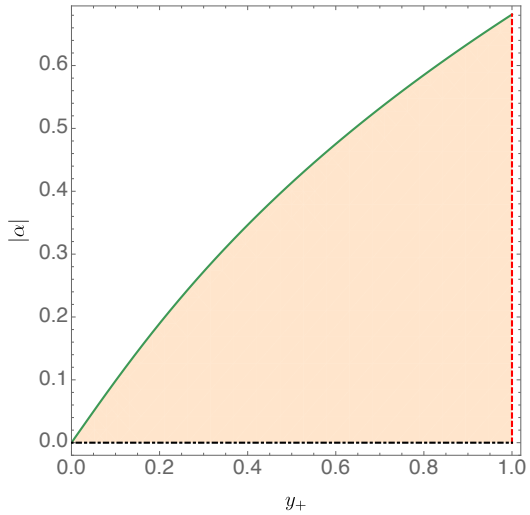


Fig. 4.2 Moduli space of solutions in the (y_+, α) plane: the green solid curve corresponds to extremality where $r_+ = r_-$, the red dotted-dashed line corresponds to the limit where the black hole horizon coincides with the cosmological horizon and the black dashed line to a Schwarzschild-de Sitter black hole.

solution. It is tangent to outgoing null geodesics in region II, *i.e.*, null geodesics which cross the Cauchy horizon.

We will parametrise Kerr-de Sitter solutions using the dimensionless quantities $\{y_+, \alpha\} \equiv \{r_+/r_c, a/r_c\}$. These variables are in one-to-one correspondence with members of the Kerr-de Sitter family of solutions and mean we essentially normalise all our quantities to r_c. The moduli space of solutions is shown in Fig. 4.2. Kerr-de Sitter black holes have three distinct extremal limits: $r_+ = r_-$, $r_+ = r_c$ and $r_+ = r_- = r_c$. The first two are marked as the green line and red dotted-dashed line in Fig. 4.2, respectively. For completeness, we also show in Fig. 4.2 the Schwarzschild limit marked as a black dashed line. When $r_+ = r_-$, we have

$$|a| = |a_{\rm ext}| \equiv \frac{r_c}{\sqrt{2}} \sqrt{(1+y_+)\sqrt{1+2y_++9y_+^2} - y_+(2+3y_+) - 1}\,. \tag{4.11}$$

4.3.2 Tetrad

When we study gravitational perturbations of Kerr-de Sitter black holes, it will be useful to introduce a null tetrad $\{\boldsymbol{\ell}, \boldsymbol{n}, \boldsymbol{m}, \bar{\boldsymbol{m}}\}$ satisfying the following orthogonality relations

$$\boldsymbol{\ell}\cdot\boldsymbol{n} = -1\,, \quad \bar{\boldsymbol{m}}\cdot\boldsymbol{m} = 1 \tag{4.12}$$

with all remaining combinations of inner products giving zero, and $g_{\mu\nu} = -2\boldsymbol{\ell}_{(\mu}\boldsymbol{n}_{\nu)} + 2\boldsymbol{m}_{(\mu}\overline{\boldsymbol{m}}_{\nu)}$.

There is obviously a lot of freedom in choosing such tetrad, and some choices make the equations governing gravitational perturbations of Kerr-de Sitter black holes easier than others. Here we will choose the Chambers-Moss null tetrad $\{\boldsymbol{\ell},\boldsymbol{n},\boldsymbol{m},\overline{m}\}$ [157], which in $\{t,r,\chi,\phi\}$ coordinates, reads:

$$\begin{aligned}
\boldsymbol{\ell}^\mu\partial_\mu &= \frac{1}{\sqrt{2}\sqrt{r^2+\chi^2}}\left(\Xi\frac{a^2+r^2}{\sqrt{\Delta_r}}\,\partial_t+\sqrt{\Delta_r}\,\partial_r+\frac{a\,\Xi}{\sqrt{\Delta_r}}\,\partial_\phi\right),\\
\boldsymbol{n}^\mu\partial_\mu &= \frac{1}{\sqrt{2}\sqrt{r^2+\chi^2}}\left(\Xi\frac{a^2+r^2}{\sqrt{\Delta_r}}\,\partial_t-\sqrt{\Delta_r}\,\partial_r+\frac{a\,\Xi}{\sqrt{\Delta_r}}\,\partial_\phi\right),\\
\boldsymbol{m}^\mu\partial_\mu &= -\frac{i}{\sqrt{2}\sqrt{r^2+\chi^2}}\left(\Xi\frac{a^2-\chi^2}{\sqrt{\Delta_\chi}}\,\partial_t+i\sqrt{\Delta_\chi}\,\partial_\chi+\frac{a\,\Xi}{\sqrt{\Delta_\chi}}\,\partial_\phi\right),
\end{aligned}\tag{4.13}$$

and $\overline{m}$ is the complex conjugate of $\boldsymbol{m}$.

We will need to investigate the regularity of such a tetrad across the Cauchy horizon. So we need to write it in outgoing coordinates $\{u,r,\chi,\phi''\}$:

$$\begin{aligned}
\boldsymbol{\ell}^\mu\partial_\mu &= \frac{\sqrt{\Delta_r}}{\sqrt{2}\sqrt{r^2+\chi^2}}\partial_r,\\
\boldsymbol{n}^\mu\partial_\mu &= \frac{\sqrt{2}}{\sqrt{r^2+\chi^2}}\left(\Xi\frac{a^2+r^2}{\sqrt{\Delta_r}}\,\partial_u-\frac{\sqrt{\Delta_r}}{2}\,\partial_r+\frac{a\,\Xi}{\sqrt{\Delta_r}}\,\partial_{\phi''}\right),\\
\boldsymbol{m}^\mu\partial_\mu &= -\frac{i}{\sqrt{2}\sqrt{r^2+\chi^2}}\left(\Xi\frac{a^2-\chi^2}{\sqrt{\Delta_\chi}}\,\partial_u+i\sqrt{\Delta_\chi}\,\partial_\chi+\frac{a\,\Xi}{\sqrt{\Delta_\chi}}\,\partial_{\phi''}\right).
\end{aligned}\tag{4.14}$$

However the tetrad (4.14) is not regular when $\Delta_r=0$ (for instance at the Cauchy horizon $r=r_-$) since $\boldsymbol{n}$ blows up there. To fix this, we change to a new tetrad where

$$\widetilde{\boldsymbol{\ell}}=\frac{1}{\sqrt{\Delta_r}}\boldsymbol{\ell},\qquad \widetilde{\boldsymbol{n}}=\sqrt{\Delta_r}\boldsymbol{n}\qquad\text{and}\qquad \widetilde{\boldsymbol{m}}=\boldsymbol{m},\tag{4.15}$$

which is now smooth when $\Delta_r=0$.

4.4 Scalar field quasinormal modes

4.4.1 Preliminaries

Consider a scalar field Φ obeying the wave equation $\Box\Phi = 0$. The wave equation separates for Kerr-de Sitter so quasinormal modes are solutions of the form

$$\Phi = e^{-i\omega t} e^{im\phi} S_{\omega\ell m}(\chi) R_{\omega\ell m}(r) \tag{4.16}$$

where $\ell = 0, 1, 2, \ldots$, $|m| \leq \ell$ and the frequency ω is determined in terms of ℓ, m and an "overtone" number $n = 0, 1, 2, \ldots$. As discussed in section 1.2.1, the solution obeys ingoing boundary conditions as $r \to r_+$. The outer boundary of the exterior of the black hole is not null infinity (as for asymptotically flat black holes) but the cosmological horizon. Therefore the outer boundary condition is that the solution is outgoing at $r = r_c$. These boundary conditions require that the solution is smooth at the future event horizon $\mathscr{H}^+$ and at the future cosmological horizon $\mathscr{H}^+_C$. Despite providing the motivation for the link between quasinormal modes and the geodesics on the photon sphere [53, 48], these boundary conditions are not actually explicitly needed for the geometric optics approximation, but are used in the numerics.

If we use ingoing coordinates (v, r, χ, ϕ'), regular in regions I and II of Fig. 4.1, then a quasinormal mode is an analytic function of the coordinates in region I and can be analytically continued into region II. In the ingoing coordinates, a quasinormal mode has time dependence $e^{-i\omega v}$, so it will diverge as $v \to -\infty$, *i.e.*, along the red line on Fig. 4.1. We are interested in the behaviour of the mode at the Cauchy horizon $\mathscr{C}\mathscr{H}^+_R$. To investigate regularity there we need to convert to outgoing coordinates in the black hole interior.

In region II, we can convert from the ingoing coordinates to coordinates (t, r, χ, ϕ) and the quasinormal mode will again take the form (4.16). Now converting (4.16) to outgoing coordinates (u, r, χ, ϕ'') in region II gives

$$\Phi = e^{-i\omega u} e^{im\phi''} S_{\omega\ell m}(\chi) \tilde{R}_{\omega\ell m}(r) \tag{4.17}$$

for some function $\tilde{R}_{\omega\ell m}$. Near the right Cauchy horizon $\mathscr{C}\mathscr{H}^+_R$, there are two independent solutions of this form, which behave as follows

$$\Phi^{(1)} = e^{-i\omega u} e^{im\phi''} S_{\omega\ell m}(\chi) \hat{R}^{(1)}_{\omega\ell m}(r)\,, \tag{4.18a}$$

$$\Phi^{(2)} = e^{-i\omega u} e^{im\phi''} S_{\omega\ell m}(\chi) (r - r_-)^{i(\omega - m\Omega_-)/\kappa_-} \hat{R}^{(2)}_{\omega\ell m}(r)\,, \tag{4.18b}$$

where $\hat{R}^{(1,2)}$ denote smooth functions which are non-zero at $r = r_-$, and $\Omega_- = \Omega(r_-)$. Notice that $\mathrm{Im}(\omega) < 0$ implies that $\Phi^{(2)}$ vanishes at $r = r_-$. However $\Phi^{(2)}$ is not smooth at $r = r_-$. At the Cauchy horizon, our quasinormal mode will be some linear combination of the above two solutions. In principle there is no reason why either of the coefficients in this linear combination should vanish[5]. Hence the regularity of the quasinormal mode is determined by the non-smooth solution $\Phi^{(2)}$. What is the condition for $\Phi^{(2)}$ to be locally square integrable? We have $\Phi^{(2)} \sim (r - r_-)^p$ with $p = i(\omega - m\Omega_-)/\kappa_-$. Hence $\partial_r \Phi^{(2)} \sim (r - r_-)^{p-1}$ which is square integrable if, and only if, $2(\beta - 1) > -1$ where $\beta = \mathrm{Re}(p)$. In other words the condition for our quasinormal mode to belong to H^1_{loc} at the Cauchy horizon is

$$\beta > \frac{1}{2} \qquad \text{where} \qquad \beta \equiv -\frac{\mathrm{Im}(\omega)}{\kappa_-}\,. \tag{4.19}$$

Therefore if we can find a quasinormal mode with $\beta < 1/2$ then the scalar field cannot be extended across the Cauchy horizon in H^1_{loc} and so strong cosmic censorship is respected. On the other hand if *all* quasinormal modes have $\beta > 1/2$ then strong cosmic censorship may be violated. Ref. [148] argued that the latter is what happens for nearly extremal Reissner-Nordström-de Sitter black holes.

4.4.2 Geometric optics

In the eikonal limit, also known as geometric optics limit, where $\ell \gg 1$ ($\ell \sim |m| \gg 1$ for spinning backgrounds) there are quasinormal mode frequencies − known as "photon sphere" quasinormal modes − which are related to the properties of the unstable circular photon orbits in the equatorial plane. Recall from section 1.2.2 that the real part ω_R of the frequency is proportional to the Keplerian frequency Ω_c of the circular null orbit and the imaginary part of the frequency is proportional to the Lyapunov exponent λ of the orbit [53, 48, 54–59, 52]. The latter describes how quickly a null geodesic congruence on the circular orbit increases its cross section under infinitesimal radial deformations.

These photon sphere quasinormal modes turn out to play a fundamental role in our discussion. Therefore, in this section we will use geometric optics to compute these modes for the Kerr-de Sitter background. In the next section we will find that the resulting analytical expression for the frequency matches extremely well the values that we find numerically already for values of $\ell = m$ as low as 10.

The geodesic equation, describing the motion of point-like particles around a Kerr-de Sitter black hole, is known to lead to a set of quadratures. This is perhaps an unexpected

[5] However, it has not actually been proven that neither coefficient vanishes. This is an important open problem, but for now we will assume that this is indeed the case.

result, since Kerr-de Sitter only possesses two Killing fields, given in our coordinate system as $K = \partial/\partial_t$ and $M = \partial/\partial_\phi$ and thus seems one short of leading to an integrable system. However, there is another conserved quantity, the Carter constant, associated to a Killing tensor K_{ab}, which saves the day [50].

The most direct way to see this integrable structure is to look at the Hamilton-Jacobi equation for null geodesics [50]:

$$\frac{\partial S}{\partial x^\mu}\frac{\partial S}{\partial x^\nu}g^{\mu\nu} = 0\,, \tag{4.20}$$

where S is the principal function, as in section 1.2.2.1.

We then take a separation *ansatz* of the form

$$S = -\tilde{E}\,t + \tilde{L}\,\phi + R(r) + X(\chi)\,, \tag{4.21}$$

which gives the following coupled ordinary differential equations for $R(r)$ and $X(\chi)$

$$\Delta_r^2(\partial_r R)^2 - \Xi^2\left(\tilde{E}\sigma_r - a\tilde{L}\right)^2 + \left[\mathscr{Q} + \Xi^2(\tilde{L} - a\tilde{E})^2\right]\Delta_r = 0\,, \tag{4.22a}$$

$$\Delta_\chi^2(\partial_\chi X)^2 - \Xi^2\left(\tilde{E}\sigma_\chi - a\tilde{L}\right)^2 - \left[\mathscr{Q} + \Xi^2(\tilde{L} - a\tilde{E})^2\right]\Delta_\chi = 0\,, \tag{4.22b}$$

where $\mathscr{Q}$ is a separation constant known as the Carter constant. The constants $\tilde{E}$ and $\tilde{L}$ are the conserved charges associated with the Killing fields K and M[6] via

$$\tilde{E} \equiv -K_\mu \dot{x}^\mu \qquad \text{and} \qquad \tilde{L} \equiv M_\mu \dot{x}^\mu\,. \tag{4.23}$$

We write a tilde on $\tilde{L}$ to distinguish it from L defined by $\Lambda = 3/L^2$.

Eqs. (4.22) translate into a statement about the particle trajectories via (1.12) (in the Introduction) and (4.21). In particular, for $\dot{\chi}$, we find

$$(r^2+\chi^2)^2\frac{\dot{\chi}^2}{\Delta_\chi} = \mathscr{Q} - \tilde{E}^2\Xi^2\left[\frac{(ab-\sigma_\chi)^2}{\Delta_\chi} - (b-a)^2\right], \tag{4.24}$$

where we define the geodesic impact parameter by

$$b \equiv \frac{\tilde{L}}{\tilde{E}}\,. \tag{4.25}$$

[6]For massive particles, these coincide with the energy and angular momentum of the particle, but for massless particles $\tilde{E}$ and $\tilde{L}$ have no physical meaning since they can be rescaled. The ratio $\tilde{L}/\tilde{E}$, however, is invariant under such rescallings.

Since we are interested in matching the behaviour of geodesics with that of quasinormal modes with large values of $\ell = m$, we can restrict attention to the equatorial plane for which $\chi = 0$. This can only be the case if initially $\chi(0) = \dot{\chi}(0) = 0$ and $\mathscr{Q} = 0$. The equation governing the radial motion now gives

$$\dot{r}^2 = V(r;b)\,, \tag{4.26}$$

where

$$V(r;b) = \frac{\tilde{L}^2\Xi^2}{b^2}\left\{1 + \frac{(a-b)^2}{L^2} + \frac{(a-b)}{r^2}\left[a+b+\frac{a^2}{L^2}(a-b)\right] + \frac{2M(a-b)^2}{r^3}\right\}. \tag{4.27}$$

We are now interested in finding the photon sphere (region where null particles are trapped on circular unstable orbits), *i.e.* the values of $r = r_s$ and $b = b_s$, such that

$$V(r_s, b_s) = 0 \quad \text{and} \quad \partial_r V(r,b)|_{r=r_s, b=b_s} = 0. \tag{4.28}$$

From the second equation above we get

$$b_s(r_s) = a\,\frac{a^2 r_s + L^2\,(3M + r_s)}{a^2 r_s + L^2\,(3M - r_s)}\,, \tag{4.29}$$

while from the first we get:

$$a^4 r_s^3 + a^2\left[2L^2 r_s^2\,(3M + r_s) - 4L^4 M\right] + L^4 r_s\,(r_s - 3M)^2 = 0\,. \tag{4.30}$$

Its real positive roots are

$$r_s^{\pm} = \frac{2M}{\Xi^2}\left\{\gamma^- + \gamma\cos\left[\frac{2}{3}\arccos\left(\mp\sqrt{\frac{1}{2} - \frac{\gamma^-\gamma^+}{2\gamma^3} + \frac{a^2\Xi^4}{M^2\gamma^3}}\right)\right]\right\} \tag{4.31}$$

where

$$\gamma \equiv \sqrt{1 - \frac{14a^2}{L^2} + \frac{a^4}{L^4}}\,, \quad \gamma^+ \equiv 1 + \frac{34a^2}{L^2} + \frac{a^4}{L^4} \quad \text{and} \quad \gamma^- \equiv 1 - \frac{a^2}{L^2}\,. \tag{4.32}$$

The signs are chosen such that r_s^+ corresponds to prograde orbits, *i.e.* $b_s^+ \equiv b_s(r_s^+) > 0$ and r_s^- to retrograde orbits, *i.e.* $b_s^- \equiv b_s(r_s^-) < 0$.

We can now compute the orbital angular velocity (aka Kepler frequency) of our null circular photon orbit, which is simply given by

$$\Omega_c^{\pm} \equiv \frac{\dot{\phi}}{\dot{t}} = \frac{1}{b_s^{\pm}}. \tag{4.33}$$

On an orbit with impact parameter $b = b_s^{\pm}$, the radial potential (4.27) simplifies considerably,

$$V(r; b_s^{\pm}) = \frac{\tilde{L}^2 \Xi^2}{(b_s^{\pm})^2} (\beta_s^{\pm})^2 \left(1 - \frac{r_s^{\pm}}{r}\right)^2 \left(1 + \frac{2 r_s^{\pm}}{r}\right), \tag{4.34}$$

where we defined

$$(\beta_s^{\pm})^2 = 1 + \frac{(a - b_s^{\pm})^2}{L^2}. \tag{4.35}$$

The final step in our calculation is to compute the largest Lyapunov exponent λ, measured in units of t, associated with infinitesimal fluctuations around photon orbits with $r(\tau) = r_s^{\pm}$. This can be readily done by perturbing the geodesic equation (4.26) with the simplified potential (4.34) and setting $r(\tau) = r_s^{\pm} + \delta r(\tau)$. One finds that small deviations obey

$$\delta r(t) = \exp\left[+\frac{\sqrt{3}}{\Xi} \beta_s^{\pm} \frac{a^2 - a\, b_s^{\pm} + (r_s^{\pm})^2}{(b_s^{\pm} - a)\, b_s^{\pm}\, r_s^{\pm}} t\right] + C_+, \tag{4.36}$$

and

$$\delta r(t) = \exp\left[-\frac{\sqrt{3}}{\Xi} \beta_s^{\pm} \frac{a^2 - a\, b_s^{\pm} + (r_s^{\pm})^2}{(b_s^{\pm} - a)\, b_s^{\pm}\, r_s^{\pm}} t\right] + C_-, \tag{4.37}$$

where $C_{\pm}$ are integration constants. The largest Lyapunov exponent is simply given by

$$\lambda^{\pm} = \left| \frac{\sqrt{3}}{\Xi} \beta_s^{\pm} \frac{a^2 - a\, b_s^{\pm} + (r_s^{\pm})^2}{(b_s^{\pm} - a)\, b_s^{\pm}\, r_s^{\pm}} \right|. \tag{4.38}$$

One reconstructs the approximate spectrum of the photon sphere family of quasinormal modes with $\ell = |m| \gg 1$ using [53, 48, 54–59, 52]

$$\omega_{\rm WKB}^{\pm} \approx m\, \Omega_c^{\pm} - i \left(n + \frac{1}{2}\right) \lambda^{\pm}, \quad \text{where} \quad n = 0, 1, 2, \cdots, \tag{4.39}$$

is the radial overtone.

In Fig. 4.3 we plot $\beta_{\rm WKB} \equiv -{\rm Im}(\omega_{\rm WKB}^{+})/\kappa_{-}$. For all the range of $(y_+, |\alpha|)$ we find that $\beta_{\rm WKB} \leq 1/2$, with $\beta_{\rm WKB} = 1/2$ saturated only at extremality (represented by the dashed

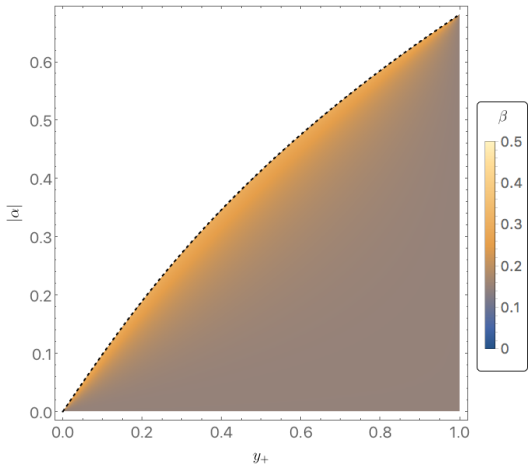


Fig. 4.3 β computed in the WKB approximation (using co-rotating photon sphere geodesics) for all values of (y_+, α). $\beta = 1/2$ is saturated at extremality, but is otherwise smaller than $1/2$. The extremal curve is represented here by the dashed black line.

'diagonal' black line in Fig. 4.3). *This shows that scalar field perturbations of any non-extremal Kerr-de Sitter black hole respect the strong cosmic censorship conjecture.*

Of course this calculation was based on approximate (geometric optics/WKB) methods and so one could ask whether corrections to these results might push the true value of β above $1/2$, especially near extremality. However, these corrections are of order $1/|m|$ so, for any fixed background, the corrections can be made arbitrarily small by taking $|m|$ large enough. So the WKB results should be reliable for sufficiently large $|m|$. In the next section we will determine the quasinormal frequencies numerically and find that, for large enough $|m|$, the WKB result is always in excellent agreement with the exact result. Moreover, we will find that accuracy of the WKB approximation actually gets better as we approach extremality.

4.4.3 Numerics

In this section we will compute numerically the quasinormal modes of a Kerr-de Sitter black hole and make a matching with the analytic results of section 4.4.2. We first note that the massless scalar wave equation admits separable solutions of the form (4.16), with $S_{\omega\ell m}(\chi)$ and $R_{\omega\ell m}(r)$ obeying the following two-parameter coupled eigenvalue problem

$$\partial_\chi \left[\Delta_\chi(\chi)\partial_\chi S_{\omega\ell m}(\chi)\right] - \left[\frac{\Xi^2}{\Delta_\chi(\chi)}\left(am - \sigma_\chi \omega\right)^2 - K\right] S_{\omega\ell m}(\chi) = 0\,, \tag{4.40a}$$

$$\partial_r \left[\Delta_r(r)\partial_r R_{\omega\ell m}(r)\right] + \left[\frac{\Xi^2}{\Delta_r(r)}\left(am - \sigma_r \omega\right)^2 - K\right] R_{\omega\ell m}(r) = 0\,. \tag{4.40b}$$

The symmetry exhibited by the above two equations is only achieved for the particular coordinates used in the line element (4.4). The eigenvalues to be determined are (ω, K) where K arises as a separation constant. Before describing the numerical method we used, we first comment on the thorny issue of boundary conditions. Both equations have regular singular points when Δ_r and Δ_χ vanish, so we can use Frobenius method to determine their behaviour there.

For the angular equation, we find

$$S_{\omega\ell m}(\chi) = (|a|-\chi)^{\pm\frac{|m|}{2}} \sum_{n=0}^{+\infty} (|a|-\chi)^n S^{(n,+)}_{\omega\ell m} \tag{4.41}$$

at $\chi = |a|$. Regularity then demands choosing the $+$ sign. A similar behaviour is found at $\chi = -|a|$:

$$S_{\omega\ell m}(\chi) = (|a|+\chi)^{\pm\frac{|m|}{2}} \sum_{n=0}^{+\infty} (|a|+\chi)^n S^{(n,-)}_{\omega\ell m}\,. \tag{4.42}$$

Again the upper sign leads to the physically meaningful solution. We thus conclude that we can factor out all non-analytic behaviour of $S_{\omega\ell m}$ by setting

$$S_{\omega\ell m}(\chi) = (a^2-\chi^2)^{\frac{|m|}{2}} \tilde{S}_{\omega\ell m}(\chi)\,, \tag{4.43}$$

and solving for the smooth eigenfunction $\tilde{S}_{\omega\ell m}(\chi)$.

For the radial coordinate, we have to distinguish the cosmological horizon from the black hole horizon. At the black hole horizon a Frobenius expansion yields

$$R_{\omega\ell m}(r) = (r-r_+)^{\pm\frac{i}{2\kappa_+}(\omega-m\Omega_+)} \sum_{n=0}^{+\infty} (r-r_+)^n R^{(n,+)}_{\omega\ell m} \tag{4.44}$$

and regularity at the black hole event horizon, which stems from demanding a smooth expansion around $r = r_+$ in ingoing coordinates (v, r, χ, ϕ') at the black hole horizon, demands choosing the lower sign. For the cosmological horizon we find

$$R_{\omega\ell m}(r) = (r_c-r)^{\pm\frac{i}{2\kappa_c}(\omega-m\Omega_c)} \sum_{n=0}^{+\infty} (r_c-r)^n R^{(n,c)}_{\omega\ell m}\,, \tag{4.45}$$

and again imposing outgoing boundary conditions at the cosmological horizon demands selecting the minus sign in the expression above. We thus consider the following field redefinition:

$$R_{\omega\ell m}(r) = (r_c-r)^{-\frac{i}{2\kappa_c}(\omega-m\Omega_c)} (r-r_+)^{-\frac{i}{2\kappa_+}(\omega-m\Omega_+)} \tilde{R}_{\omega\ell m}(r) \tag{4.46}$$

where $\tilde{R}_{\omega\ell m}(r)$ should now be a smooth function with a regular Taylor series at each of the horizons.

The procedure is now clear, we take the field redefinitions (4.43) and (4.46) and input them into Eqs. (4.40). The resulting equations are still quadratic in ω and K, and form a coupled eigenvalue problem with eigenfunctions $\left(\tilde{S}_{\omega\ell m}(\chi),\tilde{R}_{\omega\ell m}(r)\right)$ and eigenvalues (ω,K). The boundary conditions for $\tilde{S}_{\omega\ell m}(\chi)$ and $\tilde{R}_{\omega\ell m}(r)$ are then found by Taylor expanding the equations of motion close to either boundary, and turn out to be of the Robin type, *i.e.*

$$\mathscr{F}^{1,\pm}(\omega,K)\tilde{S}'_{\omega\ell m}(\pm|a|)=\mathscr{F}^{0,\pm}(\omega,K)\tilde{S}_{\omega\ell m}(\pm|a|) \tag{4.47}$$

and

$$\mathscr{Q}^{1,+}(\omega,K)\tilde{R}'_{\omega\ell m}(r_+)=\mathscr{Q}^{+,0}(\omega,K)\tilde{R}_{\omega\ell m}(r_+)\,,\quad \mathscr{Q}^{c,1}(\omega,K)\tilde{R}'_{\omega\ell m}(r_c)=\mathscr{Q}^{c,0}(\omega,K)\tilde{R}_{\omega\ell m}(r_c)\,. \tag{4.48}$$

with $\mathscr{F}^{1,\pm}(\omega,K)$, $\mathscr{F}^{0,\pm}(\omega,K)$, $\mathscr{Q}^{1,+}(\omega,K)$, $\mathscr{Q}^{0,+}(\omega,K)$ $\mathscr{Q}^{1,-}(\omega,K)$ and $\mathscr{Q}^{-,+}(\omega,K)$ being known functions which are at most second order polynomials in ω and K. For the numerical procedure, it is also useful to consider coordinates whose range do not depend on the parameters of the solution. To achieve this, we make the following simple linear coordinate transformations

$$x=\frac{|a|+\chi}{2\,|a|}\qquad\text{and}\qquad y=\frac{1-\frac{r_+}{r}}{1-\frac{r_+}{r_c}}\,. \tag{4.49}$$

The resulting equations are then solved using a Newton-Raphson algorithm, on a unit length Chebyshev grid, as first proposed in [131] and recently detailed in [158].

Our results are shown in Fig. 4.4 where we take $y_+=1/4,1/2,3/4$ (from left to right) and plot β as a function of $a/a_{\rm ext}$. Since we are interested in tracking photon sphere modes, we will take $m=\ell=10$. For most of the moduli space of solutions $\beta\ll 1/2$, and β only gets close to $1/2$ near extremality. This is why in Fig. 4.4 we restricted the range of the horizontal axis to $a/a_{\rm ext}\in[9/10,999/1000]$. Also showing in Fig. 4.4 are the analytic WKB photon sphere predictions of section 4.4.2, see (4.39), denoted by the solid black lines. For $m=\ell=50$ (not shown in Fig. 4.4) we see a maximum deviation between the analytic and numerical data which is not larger than 10^{-6} anywhere in parameter space. To sum up, our numerical results corroborate the analytic analysis performed in section 4.4.2. For the specific value of $y_+=1/2$, we have pushed our numerical scheme to $1-a/a_{\rm ext}=10^{-6}$ and see no deviation from the WKB result. Furthermore, the WKB approximation seems to become more accurate as we approach extremality.

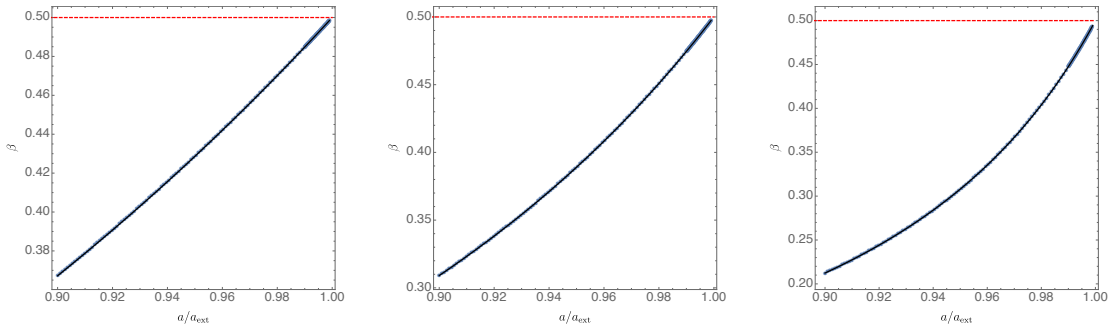


Fig. 4.4 β as a function of $a/a_{\rm ext}$ plotted for fixed several values $y_+ = 1/4, 1/2, 3/4$ (from the left to the right panel) and fixed $\ell = m = 10$.

4.5 Gravitational quasinormal modes

4.5.1 Teukolsky equation

The Kerr-de Sitter black hole is a Petrov type D solution. Therefore, gravitational perturbations of this geometry can be studied using the Teukolsky equation, which uses the Newman-Penrose (NP) framework [159–161, 50]. We will study perturbations using the Chambers-Moss null tetrad (4.13). For quasinormal modes we assume a separable *Ansatz* for the (gauge invariant) perturbed Weyl scalars

$$\psi_0 \equiv \boldsymbol{\ell}^\mu \boldsymbol{m}^\nu \boldsymbol{\ell}^\rho \boldsymbol{m}^\alpha \delta C_{\mu\nu\rho\alpha} = e^{-i\omega t + im\phi} \frac{R^{(+2)}_{\omega\ell m}(r) S^{(+2)}_{\omega\ell m}(\chi)}{(r - i\chi)^2}\,, \tag{4.50a}$$

$$\psi_4 \equiv \boldsymbol{n}^\mu \bar{\boldsymbol{m}}^\nu \boldsymbol{n}^\rho \bar{\boldsymbol{m}}^\alpha \delta C_{\mu\nu\rho\alpha} = e^{-i\omega t + im\phi} \frac{R^{(-2)}_{\omega\ell m}(r) S^{(-2)}_{\omega\ell m}(\chi)}{(r - i\chi)^2}\,, \tag{4.50b}$$

where $\delta C_{\mu\nu\rho\alpha}$ are the components of the Weyl tensor perturbation. The Teukolsky equation then reduces to the following two sets of two-parameter eigenvalue problems[7]

$$\begin{cases} \left[\mathscr{D}_{-1} \Delta_r \mathscr{D}_1^\dagger - 6 \left(\dfrac{r^2}{L^2} - i\,\Xi\,\omega\, r \right) - K^{(+2)} \right] R^{(+2)}_{\omega\ell m}(r) = 0\,, \\ \left[\mathscr{L}_{-1} \Delta_\chi \mathscr{L}_1^\dagger - 6 \left(\dfrac{\chi^2}{L^2} + \Xi\,\omega\,\chi \right) + K^{(+2)} \right] S^{(+2)}_{\omega\ell m}(\chi) = 0\,, \end{cases} \tag{4.51}$$

[7] The reader interested on a complete but concise overview that discusses how the solutions of (4.51)-(4.52) allow to get information about other variables can see section 2 and appendix A of [162] (with the trading $L^2 \to -L^2$).

and

$$\begin{cases} \left[\mathscr{D}_{-1}^{\dagger}\Delta_r\mathscr{D}_1 - 6\left(\dfrac{r^2}{L^2} + i\,\Xi\,\omega\,r\right) - K^{(-2)}\right] R^{(-2)}_{\omega\ell m}(r) = 0\,, \\ \left[\mathscr{L}_{-1}^{\dagger}\Delta_\chi\mathscr{L}_1 - 6\left(\dfrac{\chi^2}{L^2} - \Xi\,\omega\,\chi\right) + K^{(-2)}\right] S^{(-2)}_{\omega\ell m}(\chi) = 0\,, \end{cases} \tag{4.52}$$

where $K^{(\pm 2)}$ are separation constants and we defined the operators [50, 157]:

$$\begin{aligned} \mathscr{D}_n &= \partial_r + i\,\frac{\Xi}{\Delta_r}\,(ma - \omega\sigma_r) + n\,\frac{\partial_r\Delta_r}{\Delta_r}, \qquad & \mathscr{D}_n^{\dagger} &= \partial_r - i\,\frac{\Xi}{\Delta_r}\,(ma - \omega\sigma_r) + n\,\frac{\partial_r\Delta_r}{\Delta_r}\,, \\ \mathscr{L}_n &= \partial_\chi + \frac{\Xi}{\Delta_\chi}\,\big(ma - \omega\sigma_\chi\big) + n\,\frac{\partial_\chi\Delta_\chi}{\Delta_\chi}, \qquad & \mathscr{L}_n^{\dagger} &= \partial_\chi - \frac{\Xi}{\Delta_\chi}\,\big(ma - \omega\sigma_\chi\big) + n\,\frac{\partial_\chi\Delta_\chi}{\Delta_\chi}\,. \end{aligned} \tag{4.53}$$

Equations (4.51) and (4.52) are isospectral[8], that is to say, once appropriate boundary conditions are imposed, they give the same values of ω and $K^{(+2)} = K^{(-2)}$. So in the following section, we shall focus on the pair $\{R^{(+2)}_{\omega\ell m}(r), S^{(+2)}_{\omega\ell m}(\chi)\}$ with eigenvalues $\{\omega, K^{(+2)}\}$.

For our discussion of strong cosmic censorship, we need to determine the behaviour of the Weyl scalar ψ_0 defined in (4.50a) at the Cauchy horizon. For that we use the outgoing coordinates (u, r, χ, ϕ'') that extend the solution across $r = r_-$. The radial equation for $R^{(2)}_{\omega\ell m}(r)$ has a regular singular point when $\Delta_r = 0$ and thus a Frobenius analysis yields the two possible behaviours at the Cauchy horizon $r = r_-$. We find that the most general solution for ψ_0 near $r = r_-$ is a linear combination of $\psi_0^{(1)}$ and $\psi_0^{(2)}$ where

$$\psi_0^{(1)} = e^{-i\omega u}e^{im\phi''}S^{(+2)}_{\omega\ell m}(\chi)(r - i\chi)^{-2}(r - r_-)\hat{R}^{(+2)(1)}_{\omega\ell m}(r)\,, \tag{4.54a}$$

$$\psi_0^{(2)} = e^{-i\omega u}e^{im\phi''}S^{(+2)}_{\omega\ell m}(\chi)(r - i\chi)^{-2}(r - r_-)^{-1+i(\omega - m\Omega_-)/\kappa_-}\hat{R}^{(+2)(2)}_{\omega\ell m}(r)\,, \tag{4.54b}$$

where $\Omega_- = \Omega(r_-)$ and $\hat{R}^{(+2)(1)}_{\omega\ell m}$, $\hat{R}^{(+2)(2)}_{\omega\ell m}$ are smooth functions of r that are non-zero at $r = r_-$.

This gives the behaviour in the Chambers-Moss tetrad (4.14). This tetrad is not regular at the Cauchy horizon so we need to convert our results to a regular tetrad. Consider the Weyl scalar $\widetilde{\psi}_0 \equiv \widetilde{\boldsymbol{\ell}}^{\mu}\widetilde{\boldsymbol{m}}^{\nu}\widetilde{\boldsymbol{\ell}}^{\rho}\widetilde{\boldsymbol{m}}^{\alpha}\delta C_{\mu\nu\rho\alpha}$ defined using the regular null tetrad $\{\widetilde{\boldsymbol{\ell}}, \widetilde{\boldsymbol{n}}, \widetilde{\boldsymbol{m}}, \overline{\widetilde{\boldsymbol{m}}}\}$ defined in (4.15). We now have $\widetilde{\psi}_0 = \psi_0/\Delta_r$ and hence, near the Cauchy horizon $\widetilde{\psi}_0$ is a linear

[8] We have explicitly checked this is the case, by computing the corresponding sets of quasinormal modes associated with each of the equations. It is likely to be related to the generalisation of the Teukolsky-Starobinsky identities to Kerr-de Sitter [157].

combination of $\widetilde{\psi}_0^{(1)}$ and $\widetilde{\psi}_0^{(2)}$, where

$$\widetilde{\psi}_0^{(1)} = e^{-i\omega u} e^{im\phi''} S^{(+2)}_{\omega\ell m}(\chi)(r-i\chi)^{-2}\tilde{R}^{(+2)(1)}_{\omega\ell m}(r)\,, \tag{4.55a}$$

$$\widetilde{\psi}_0^{(2)} = e^{-i\omega u} e^{im\phi''} S^{(+2)}_{\omega\ell m}(\chi)(r-i\chi)^{-2}(r-r_-)^{-2+i(\omega-m\Omega_-)/\kappa_-}\tilde{R}^{(+2)(2)}_{\omega\ell m}(r)\,, \tag{4.55b}$$

and $\tilde{R}^{(+2)(i)}_{\omega\ell m} \equiv f\hat{R}^{(+2)(i)}_{\omega\ell m}$ $(i=1,2)$ where $f \equiv (r-r_-)/\Delta_r$ is smooth and non-vanishing at the Cauchy horizon. It follows that the $\tilde{R}^{(+2)(i)}_{\omega\ell m}$ are smooth and non-vanishing at the Cauchy horizon.

The solution $\widetilde{\psi}_0^{(1)}$ is smooth and non-vanishing at the Cauchy horizon. However, the solution $\widetilde{\psi}_0^{(2)}$ diverges at the Cauchy horizon. A quasinormal mode solution will be a linear combination of these two solutions and there is no reason why either coefficient in this linear combination should vanish. It follows that $\widetilde{\psi}_0$ diverges at the Cauchy horizon. Defining $p = i(\omega - m\Omega_-)/\kappa_-$, the behaviour, in the regular tetrad, of a quasinormal mode near the Cauchy horizon is

$$\widetilde{\psi}_0 \sim (r-r_-)^{p-2}. \tag{4.56}$$

We now define $\beta = \mathrm{Re}(p) = -\mathrm{Im}(\omega)/\kappa_-$ as before. We will show that if the quasinormal mode corresponds to a linearized metric perturbation that, in some gauge, is in H^1_{loc} then we must have $\beta > 1/2$.

The easiest way to see this is as follows. If the linearized metric perturbation is in H^1_{loc} in some gauge then the corresponding (gauge invariant) $\widetilde{\psi}_0$ must be in H^{-1}_{loc}. For (4.56) to belong to H^{-1}_{loc} we need $\beta > 1/2$.

A less rigorous, argument goes as follows. Assume that, in the coordinates (u, r, χ, ϕ'') each component of the metric perturbation behaves (near the Cauchy horizon) as $(r-r_-)^q$ where $q = q_R + i\left(\mathrm{Re}(\omega) - m\Omega_-\right)/\kappa_- = q_R + i\,\mathrm{Im}(p)$. The value of the real part q_R may be different for different components. The condition that the perturbation belongs to H^1_{loc} is that each component must have $q_R > 1/2$. Since the Weyl tensor perturbation involves two derivatives of the metric, it follows that the Weyl scalar must be at least as smooth as $(r-r_-)^{\min(q_R)+i\,\mathrm{Im}(p)-2}$ and hence from (4.56) we must have $\mathrm{Re}(p) > \min(q_R) > 1/2$, *i.e.*, $\beta > 1/2$.

We conclude that if a quasinormal mode corresponds to a linearized metric perturbation that, in some gauge, belongs to H^1_{loc} then the mode must have

$$\beta > \frac{1}{2} \qquad \text{where} \qquad \beta \equiv -\frac{\mathrm{Im}(\omega)}{\kappa_-}. \tag{4.57}$$

Hence if *all* gravitational quasinormal modes have $\beta > 1/2$ then strong cosmic censorship might be violated. However, if we can find one quasinormal mode with $\beta < 1/2$ then the

linearized gravitational field cannot be extended across the Cauchy horizon in H^1_{loc} and strong cosmic censorship holds.

We can use geometric optics to calculate the frequencies of "photon sphere" gravitational quasinormal modes with $\ell = |m| \gg 1$. The calculation is exactly as in section 4.4.2. As explained in [59] (see Eq. (51) of [59]), the spin dependence of the WKB approximation of quasinormal frequencies with $\ell = |m|$ only comes at order $1/m$, for both co-rotating and counter-rotating modes. This makes sense, since in the WKB limit we are taking $\ell = |m|$ to be large, while keeping the spin fixed (either to zero, in the scalar case, or to two in the gravitational case). Hence for $\ell = |m| \gg 1$, the gravitational quasinormal frequencies are, to leading order, exactly the same as the scalar field quasinormal frequences, as computed in section 4.4.2. Furthermore, the subleading terms can be made arbitrarily small by taking $\ell = |m|$ large enough. We conclude that for any non-extremal Kerr-de Sitter black hole, there are gravitational quasinormal modes with $\beta < 1/2$. So *gravitational perturbations of any non-extremal Kerr-de Sitter black hole respect the strong cosmic censorship conjecture*.

In the next section, we will check the accuracy of the geometric optics/WKB approximation for gravitational perturbations by computing the quasinormal frequencies numerically. Just as for the scalar field case, we will find that the geometric optics approximation is always very accurate for $\ell = m \gg 1$ and gets better near to extremality.

4.5.2 Numerics

We write the perturbation for the Weyl scalar ψ_0 as in (4.50a). Our task now is to find $S^{(+2)}_{\omega\ell m}(\chi)$, $R^{(+2)}_{\omega\ell m}(r)$ and the eigenvalues $\{\omega, K^{(+2)}\}$ by solving (4.51). As explained before, (4.52) is isospectral to (4.51) and thus we do not consider it further.

This section follows *mutatis mutandis* section 4.4.3, so we will only point out the differences. Regularity at the poles, located at $\chi = \pm|a|$ now demands that

$$S^{(+2)}_{\omega\ell m}(\chi) = (a^2 - \chi^2)^{\frac{|m-2|}{2}} \tilde{S}^{(+2)}_{\omega\ell m}(\chi)\,, \tag{4.58}$$

where $\tilde{S}^{(+2)}_{\omega\ell m}(\chi)$ is a smooth function of χ for all values of m. He have discarded the irregular solution $(a^2-\chi^2)^{-\frac{|m-2|}{2}}$. Demanding outgoing boundary conditions at the cosmological horizon $-$ *i.e.* that the solution is regular at $r = r_c$ in outgoing coordinates (u, r, χ, ϕ'') $-$ and ingoing boundary conditions at the black hole horizon $-$ *i.e.* that the solution is regular at $r = r_+$ in ingoing coordinates (v, r, χ, ϕ') $-$ now motivates the following field redefinition:

$$R^{(+2)}_{\omega\ell m}(r) = (r_c - r)^{-\frac{i}{2\kappa_c}(\omega - m\Omega_c)+1} (r - r_+)^{-\frac{i}{2\kappa_+}(\omega - m\Omega_+)-1} \tilde{R}^{(+2)}_{\omega\ell m}(r)\,, \tag{4.59}$$

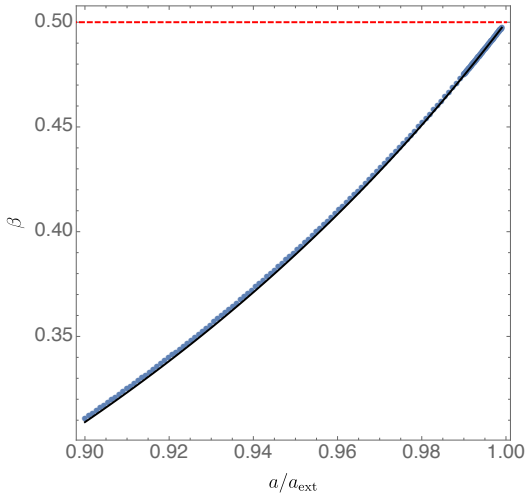


Fig. 4.5 β for gravitational perturbations as a function of $a/a_{\rm ext}$ plotted for fixed $y_+ = 1/2$ and $\ell = m = 10$. The blue dots are the numerical results while the black line is the analytic approximation.

where again $\tilde{R}^{(+2)}_{\omega\ell m}(r)$ is a smooth function at both $r = r_+$ and $r = r_c$.

The numerical results are displayed in Fig. 4.5, where we have choosen $y_+ = 1/2$, $a/a_{\rm ext} \in [9/10, 999/1000]$ and $m = \ell = 10$. As expected, at large enough $\ell = m$, the spin is irrelevant, and the analytic approximation of section 4.4.2 is excellent. The only difference worth noticing is that it seems we need to get to larger values of $m = \ell$ in order for the analytic approximation to be as accurate as for the scalar field case. We note, however, that the approximation gets increasingly better as we approach extremality.

4.6 Discussion

It is worth emphasising that spherically symmetric, charged black holes are usually expected to have similar properties to rotating, uncharged black holes. However, this chapter shows that there is a fundamental difference between Reissner-Nordström-de Sitter and Kerr-de Sitter black holes. For *any* non-extremal Kerr-de Sitter black hole the photon sphere modes decay slowly enough for us to conclude that the linear version of Christodoulou's formulation of the strong cosmic censorship conjecture is respected. On the other hand, the analogous modes in near-extremal Reissner-Nordström-de Sitter black holes decay much faster, and we cannot draw the same conclusion. In terms of the equations, the difference comes about simply because one of the terms in the radial component of the wave equation on Kerr-de Sitter vanishes when $a = 0$. Physically, the null geodesics on the photon sphere that spread out slowly in Kerr-de Sitter simply aren't there in Reissner-Nordström-de Sitter black holes; all the photon sphere orbits spread out more quickly in that case.

As mentioned in section 4.4.1, it is crucial that both coefficients in the linear combination of solutions at the Cauchy horizon are non-zero. We simply assumed this to be the case, as there is no particular reason why either should vanish. However, it would be considerably better to rigorously prove that this is indeed the case, and this remains an open problem.

The analytic calculation in this chapter relied entirely on the geometric optics approximation. This could be improved by possibly using the WKB method or some kind of asymptotic matching in the limit $m \to \infty$ (similar to that of chapter 2) to obtain the quasinormal frequencies. Even better would be to find them mathematically as the poles of the Green's function of the associated operator, perhaps in a similar way to [37], which does this for slowly rotating Kerr-de Sitter black holes.

Chapter 5

Predictability of subluminal and superluminal wave equations

This chapter is a slightly modified version of [163], which was done in collaboration with Harvey Reall and Jan Sbierski and has been accepted for publication in Communications in Mathematical Physics. The theorems in section were all contributed by Jan Sbierski, so I have omitted the proofs and instead related them back to the numerical example in section 5.3.7, which I constructed.

5.1 Introduction

Many Lorentz invariant classical field theories permit superluminal propagation of signals around non-trivial background solutions. It is sometimes claimed that such theories are unviable because the superluminality can be exploited to construct causality violating solutions, i.e., “time machines". The argument for this is to consider two lumps of non-trivial field with a large relative boost: it is claimed that there exist solutions of this type for which small perturbations will experience closed causal curves [164]. However, this argument is heuristic: the causality-violating solution is not constructed, it is simply asserted to exist. This means that the argument is open to criticism on various grounds [165–167].

The reason that causality violation would be problematic is that it implies a breakdown of predictability. In this chapter, rather than focusing on causality violation, we will investigate predictability. Our aim is to determine whether there is any qualitative difference in predictability between Lorentz invariant classical theories which permit superluminal propagation and those that do not.

We will consider quasilinear scalar wave equations for which causality is determined by a metric $g(u,du)$ which depends on the scalar field u and its first derivative du. In the initial value problem we specify initial data (S,u,du) where S is the initial hypersurface and u, du are chosen on S such that S is spacelike w.r.t. $g(u,du)$. We can now ask: what is the largest region M of spacetime in which the solution is uniquely determined by the data on S? Uniqueness requires that (M,g) should be globally hyperbolic with Cauchy surface S, i.e., the solution should be a globally hyperbolic development (GHD) of the data on S. This suggests that the "largest region in which the solution is unique" will be a GHD that is inextendible as a GHD, i.e., it is a *maximal* globally hyperbolic development (MGHD).

Our aim, then, is determine whether there is any qualitative difference between MGHDs for subluminal and superluminal equations.

In section 5.2, we will introduce the class of scalar wave equations that we will study, and define what we mean by "subluminal" and "superluminal" equations. Note that the standard linear wave equation is both subluminal and superluminal according to our definition.

In section 5.3 we will study an example of a Lorentz invariant equation in $1+1$ dimensions, namely the Born-Infeld scalar field. The general solution of this equation is known [168, 169]. This equation can be formulated in either a subluminal or superluminal form. One can consider the interaction of a pair of wavepackets in these theories. If the amplitude of the wavepackets is not too large then the wavepackets merge, interact, and then separate again [169]. In the subluminal theory they emerge with a time delay, in the superluminal theory there is a time advance. The MGHD is the entire 2d Minkowski spacetime in both cases.

For larger amplitude, it is known that the solution can form singularities in the subluminal theory [169]. Singularities can also form in the superluminal theory. In both cases, the formation of a singularity leads to a loss of predictability because MGHDs are extendible across a Cauchy horizon, and the solution is not determined uniquely beyond a Cauchy horizon. However, there is a qualitative difference between the subluminal and superluminal theories. In the superluminal theory there is a *unique* MGHD. However, in the subluminal theory, MGHDs are not unique: *there can exist multiple distinct MGHDs arising from the same initial data*.

This is worrying behaviour. Given a solution defined in some region U, we can ask: in which subset of U is the solution determined uniquely by the initial data? In the superluminal case, this region is simply the intersection of U with the unique MGHD, or, equivalently, the domain of dependence of the initial surface within U. This can be determined *from the solution itself*. However, in the subluminal case there is, in general, no such method of

determining the appropriate subset of U. To determine the region in which the solution is unique, one has to construct all other solutions arising from the same initial data!

In section 5.4 we will discuss the existence and uniqueness of MGHDs for a large class of quasilinear wave equations (in any number of dimensions). We start by giving a theorem (proven in [163]) asserting that two GHDs defined in regions U_1 and U_2 will agree in $U_1 \cap U_2$ *provided* $U_1 \cap U_2$ *is connected*. Thus if one can show that $U_1 \cap U_2$ is *always* connected then one always has uniqueness. This is the case for any equation with the property that there exists a vector field which is timelike w.r.t. $g(u, du)$ for all (u, du). For such an equation, and for a suitable initial surface, there exists a unique MGHD. Note that *any* superluminal equation admits such a vector field so *for any superluminal equation there exists a unique MGHD*.

Our Born-Infeld example demonstrates that one cannot expect a unique MGHD for a general subluminal equation. One can define the maximal region in which solutions are unique, which we call the maximal unique globally hyperbolic development (MUGHD). Unfortunately, as mentioned above, there is no simple characterization of the MUGHD: given a solution defined in a region U, there is no simple general method for determining which part of U belongs to the MUGHD. As we will show, one can establish some partial results e.g. for a solution defined in U, the solution is unique in the subset of U corresponding to the domain of dependence of the initial surface detemined w.r.t. the *Minkowski* metric. However, this is rather a weak result especially for equations with a speed of propagation considerably less than the speed of light.

An important application of the notion of a MGHD is Christodoulou's work on shock formation in relativistic perfect fluids [170]. Given that this work concerns subluminal equations, one might wonder whether the MGHD constructed in Ref. [170] suffers from the lack of uniqueness dicussed above. However, if a MGHD "lies on one side of its boundary" then it is unique. This provides a method for demonstrating uniqueness of a MGHD once it has been constructed. In particular, this implies that there is a unique MGHD for the initial data considered in Ref. [170]. However, we emphasize that the equations of Ref. [170] are likely to exhibit non-uniqueness of MGHDs for more complicated choices of initial data.

Of course we have not answered the question which motivated the present work, namely whether it is possible to "build a time machine" in any Lorentz invariant theory which admits superluminal propagation. However, this chapter does show that the object that one would have to study in order to address this question, namely the MGHD, is well-defined in a superluminal theory. Smooth formation of a time machine would require that there exist generic initial data belonging to some suitable class (e.g. smooth, compactly supported, data specified on a complete surface extending to spatial infinity in Minkowski spacetime)

for which the MGHD is extendible, with a compactly generated [171] Cauchy horizon.[1] In the section 5.3.2 we explain why this is not possible in $1+1$ dimensions. Whether this is possible in a higher dimensional superluminal theory (let alone *all* such theories) is an open question.

5.2 General scalar equation

5.2.1 Subluminal and superluminal equations

Consider a scalar field $u : \mathbb{R}^{d+1} \to \mathbb{R}$ in $(d+1)$-dimensional Minkowski spacetime. Assume that the field satisfies a quasilinear equation of motion[2]

$$g^{\mu\nu}(u,du)\partial_\mu\partial_\nu u = F(u,du) \tag{5.1}$$

where F is a smooth[3] function and (5.1) is written with respect to the canonical coordinates x^μ on $\mathbb{R}^{d+1}$.

We will say that (M,u) is a *hyperbolic solution* if M is a connected open subset of $\mathbb{R}^{d+1}$ and $u : M \to \mathbb{R}$ is a smooth solution of the above equation for which $g^{\mu\nu}(u,du)$ has Lorentzian signature. For such a solution we can define $g_{\mu\nu}(u,du)$ as the inverse of $g^{\mu\nu}$ and then (M,g) is a spacetime. Causality for the scalar field is determined by the metric g so we will be studying the causal properties of the spacetime (M,g).

Now assume that we have a Minkowski metric $m_{\mu\nu}$ on $\mathbb{R}^{d+1}$ (i.e. a flat, Lorentzian metric), with inverse $m^{\mu\nu}$. We call the above equation *subluminal* if, whenever $g^{\mu\nu}$ is Lorentzian, every vector that is causal w.r.t. $g_{\mu\nu}$ is also causal w.r.t. $m_{\mu\nu}$ (so the null cone of $g_{\mu\nu}$ lies on, or inside, the null cone of $m_{\mu\nu}$). We call the equation *superluminal* if, whenever $g^{\mu\nu}$ is Lorentzian, every vector that is causal w.r.t. $m_{\mu\nu}$ is also causal w.r.t. $g_{\mu\nu}$ (so the null cone of $g_{\mu\nu}$ lies on, or outside, the null cone of $m_{\mu\nu}$).

Most equations are neither subluminal nor superluminal e.g. because the null cones of $g_{\mu\nu}$ and $m_{\mu\nu}$ may not be nested or because the relation between the null cones of $g_{\mu\nu}$ and $m_{\mu\nu}$ may be different for different field configurations. Note also that the standard wave equation ($g_{\mu\nu} = m_{\mu\nu}$) is both subluminal and superluminal according to our definitions.

Clearly these definitions depends on the choice of $m_{\mu\nu}$. There are infinitely many Minkowski metrics on $\mathbb{R}^{d+1}$. An equation might be subluminal w.r.t. one choice of $m_{\mu\nu}$

[1] The word “generic" is included to reflect the condition that the time machine should be stable under small perturbations of the initial data.

[2] Everything we say in the next few sections applies also to a quasilinear *system*, where u denotes a N-component vector of scalar fields.

[3] Here, and throughout this paper, ‘smooth’ means C^∞.

and superluminal w.r.t. some other choice. However, for many equations there exists no $m_{\mu\nu}$ such that the equation is either subluminal or superluminal. In physics applications one usually has a preferred choice of $m_{\mu\nu}$, i.e., $m_{\mu\nu}$ is "the" spacetime metric. In particular, this is the case for the class of Lorentz invariant equations (defined below).

Since M is a subset of $\mathbb{R}^{d+1}$ it follows that M is orientable because an orientation $(d+1)$-form of $\mathbb{R}^{d+1}$ can be restricted to M. In the superluminal case, any vector field T^μ that is timelike w.r.t. $m_{\mu\nu}$ must also be timelike w.r.t. $g_{\mu\nu}$. It follows that (M,g) is time orientable in the superluminal case. In the subluminal case, note that the null cone of $g^{\mu\nu}$ lies on or outside the null cone of $m^{\mu\nu}$ hence the 1-form dx^0 (for inertial frame coordinates x^μ) is timelike w.r.t. $g^{\mu\nu}$. Therefore $T^\mu = -g^{\mu\nu}(dx^0)_\nu = -g^{0\mu}$ defines a time orientation so (M,g) is time orientable. Furthermore, this shows that x^0 is a global time function which implies that (M,g) is stably causal in the subluminal case [165].

5.2.2 The initial value problem

Let's now discuss the initial value problem for an equation of the form (5.1). Consider prescribing smooth initial data (S,u,du) where S is a hypersurface in $\mathbb{R}^{d+1}$ and (u,du) are specified on S. Local well-posedness of the initial value problem requires that initial data is chosen so that $g(u,du)$ is Lorentzian and that S must be spacelike w.r.t. $g(u,du)$. Given such data, one expects a unique hyperbolic solution of (5.1) to exist locally near S.[4]

We'll say that a hyperbolic solution (M,u) is a *development* of the data on S if $S \subset M$ and the solution (M,u) is consistent with the data on S. To discuss predictability, we would like to know whether (M,u) is *uniquely* determined by the initial data (S,u,du). A necessary condition for such uniqueness is that (M,g) should be globally hyperbolic with Cauchy surface S. If (M,g) is not globally hyperbolic then the solution in the region of M beyond the Cauchy horizons $H^\pm(S)$ is not determined uniquely by the data on S. We will say that a hyperbolic solution (M,u) is a *globally hyperbolic development* (GHD) of the initial data iff (M,g) is globally hyperbolic with Cauchy surface S.

A GHD (M,u) is *extendible* if there exists another GHD (M',u') with $M \subsetneq M'$ and $u = u'$ on M. We say that (M,u) is a *maximal* globally hyperbolic development (MGHD) of the initial data if (M,u) is not extendible as a GHD of the specified data on S. Note that a MGHD might be extendible but the extended solution will not be a GHD of the data on S: it will exhibit a Cauchy horizon for S.

MGHDs play an important role in General Relativity. In General Relativity, given initial data for the Einstein equation, there exists a *unique* (up to diffeomorphisms) MGHD of

[4] In fact, for a general equation this is expecting too much. We will discuss this in section 5.4.1 and Proposition 5.4.1.

the data [172]. This MGHD is therefore the central object of interest in GR because it is the largest region of spacetime that can be uniquely predicted from the given initial data. Any well-defined question in the theory, such as the strong cosmic censorship conjecture in chapter 4, can be formulated as a question about the MGHD.

Surprisingly, the subject of maximal globally hyperbolic developments for equations of the form (5.1) has not received much attention.[5] By analogy with the Einstein equation one might expect a unique MGHD for such an equation. We will see that this is indeed the case for superluminal equations but it is not always true for subluminal equations. The reason that this does not occur for the Einstein equation is that solving the Einstein equation involves constructing the background manifold (which gives flexibility) whereas in solving (5.1) the background manifold is fixed. It is this rigidity which leads to non-uniqueness of MGHDs for subluminal equations.

5.3 Born-Infeld scalar in two dimensions

5.3.1 Two dimensions

Let's now consider *Lorentz invariant* equations. By this we mean that we pick a Minkowski metric $m_{\mu\nu}$ on $\mathbb{R}^{d+1}$, with constant components in the canonical coordinates x^μ, and we demand that isometries of $m_{\mu\nu}$ map solutions of the equation to solutions of the equation. We will assume that our equation has the form (5.1) where now $g = g(m,u,du)$ and $F = F(m,u,du)$ depend on the choice of m.

The two-dimensional case is special because if m is a Minkowski metric then so is

$$\hat{m} = -m \tag{5.2}$$

Using this fact we can relate subluminal and superluminal equations. Define

$$\hat{g}(m,u,du) = -g(-m,u,du) \tag{5.3}$$

and

$$\hat{F}(m,u,du) = -F(-m,u,du) \tag{5.4}$$

[5]The only exceptions we are aware of are the sketches in [170], Chapter 2, page 40, and in [85], Section 1.4.1, which both do not mention the subtleties arising in the case of general wave equations, namely that for two GHDs $u_1 : U_1 \to \mathbb{R}$ and $u_2 : U_2 \to \mathbb{R}$ of the same initial data posed on a connected hypersurface we do not need to have that $U_1 \cap U_2$ is connected. For more on this see our detailed discussion in Section 5.4.2.

Now u satisfies (5.1) if, and only if, it satisfies

$$\hat{g}^{\mu\nu}(\hat{m},u,du)\partial_\mu\partial_\nu u = \hat{F}(\hat{m},u,du) \tag{5.5}$$

We view this equation as describing a scalar field in 2d Minkowski spacetime with metric $\hat{m}$. It is easy to see that if (5.1) is a subluminal equation then (5.5) is superluminal, and vice-versa.

Since the above transformation reverses the overal sign of m and g, it maps timelike vectors to spacelike vectors and vice-versa, i.e., the causal "cones" of the two theories are the complements of each other. This means that any solution of a superluminal equation arises from a solution of the corresponding subluminal equation simply by interchanging the definitions of timelike and spacelike. For example, if one draws a spacetime diagram for a solution of the subluminal equation, with time running from bottom to top, then the same diagram describes a solution of the superluminal equation, with time running from left to right (or right to left: one still has the freedom to choose the time orientation).

5.3.2 Superluminal equations in two dimensions

In this section we will consider causal properties of superluminal equations in $1+1$ dimensions. The low dimensionality imposes strong restrictions on the causal structure of solutions. We will review some results on causality in $1+1$ dimensions and explain why it is not possible to violate causality in a smooth way in a finite region of spacetime.

Assume we have a hyperbolic solution u defined on some open subset M of $\mathbb{R}^2$. If M is simply connected then M is homeomorphic to $\mathbb{R}^2$, which implies that (M,g) is stably causal [173, Theorem 3.43]. Hence a violation of stable causality requires that M is not simply connected i.e. M must have holes or punctures. We assume that (M,u) is *inextendible*, i.e., it is not possible to extend u as a hyperbolic solution onto a connected open set strictly larger than M. Hence the non-trivial topology of M must be associated with u developing some pathological feature when we attempt to extend to points of ∂M, for example, u or its derivative might blow up, or g might fail to be Lorentzian at such points.

This looks bad for the possibility of smoothly violating causality (i.e. "forming a time machine"). But maybe the above pathological features are *consequences* of the time machine, i.e., they lie to the future of the causality violating region. This is not the case: we will explain why some such pathology must occur *before* causality is violated. One cannot form a time machine smoothly in a two-dimensional superluminal theory.

Pick inertial coordinates (t,x) for Minkowski spacetime so that

$$m = -dt^2 + dx^2 \tag{5.6}$$

Now dx is spacelike w.r.t. $m^{\mu\nu}$, which implies that it is also spacelike w.r.t. $g^{\mu\nu}$ (because, for a superluminal equation, the null cone of $g^{\mu\nu}$ lies on, or inside, that of $m^{\mu\nu}$). Hence x is a *global space function* for g, i.e., a function with everywhere spacelike (non-zero) gradient. The transformation of the previous subsection relates it to a global time function of the corresponding subluminal equation.

Consider a null geodesic of g. Then x must be monotonic along the geodesic. To see this, let V be tangent to the geodesic. Then $V^x = V^\mu (dx)_\mu$ and this cannot vanish because V is null w.r.t. g and dx is spacelike w.r.t. g.[6] Non-vanishing of V^x implies that x is monotonic along the geodesic. It follows that *a null geodesic of g cannot be closed and cannot intersect itself.* (Again this is easy to understand using the transformation of the previous section.)

It is also easy to see that there cannot be a smooth closed future-directed causal curve (w.r.t. g) which is *simple*, i.e., does not intersect itself. This is because there will be a point on any such curve at which the tangent vector is timelike and past directed w.r.t. the Minkowski metric $m_{\mu\nu}$ and hence also timelike and past directed w.r.t. $g_{\mu\nu}$, contradicting the fact that the curve is future-directed. Hence *a closed future-directed causal curve must be non-smooth or non-simple.*

Now let S be a partial Cauchy surface, i.e., a surface (actually a line) which, viewed as a subset of (M,g), is closed, achronal and edgeless [10]. The future domain of dependence of S is $D^+(S)$ and the future Cauchy horizon is $H^+(S) = \overline{D^+(S)} - I^-(D^+(S))$. If causality is violated to the future of S then this must occur outside $D^+(S)$, so $H^+(S)$ is non-empty. A standard result states that $H^+(S)$ is achronal and closed, and that every $p \in H^+(S)$ lies on a null geodesic contained in $H^+(S)$ which is past inextendible without a past endpoint in M [10].

Consider following a generator of $H^+(S)$ to the past. Since x is monotonic it must either diverge or approach a finite limit along this generator. If x diverges then the generator originates from infinity in $\mathbb{R}^2$. Consider the case that x approaches a finite limit in the past. From the fact that the generator is null w.r.t. g and hence non-timelike w.r.t. m we have $|dt/dx| \leq 1$, which implies (via integration) that t also approaches a finite limit. Hence the generator has an endpoint p in $\mathbb{R}^2$. But it cannot have an endpoint in M so $p \notin M$. Since we are assuming that (M,u) is inextendible, p must correspond either to a singularity of the

[6] In 2d let P and Q be non-zero vectors such that $g_{\mu\nu}P^\mu Q^\nu = 0$. If P is timelike (spacelike) w.r.t. g then Q must be spacelike (timelike) w.r.t. g. If P is null w.r.t. g then Q must also be null, and parallel to P.

spacetime (M,g), or to a "point at infinity" in (M,g). In the latter case, g would have to blow up at p, which is singular behaviour from the point of view of the Minkowski spacetime.

This proves that generators of $H^+(S)$ must emanate either from infinity in Minkowski spacetime or from a point of $\mathbb{R}^2$ that is singular w.r.t. (M,g) or "at infinity" w.r.t. (M,g). None of these possibilities corresponds to what is usually regarded as the condition for creation of a time machine in a bounded region of space, namely a "compactly generated" Cauchy horizon [171] (one whose generators remain in a compact region of (M,g) when extended to the past). If the generator does not emanate from infinity in $\mathbb{R}^2$ then it remains in a compact region of $\mathbb{R}^2$ but not a compact region of M: in M it "emerges from a singularity" or "from infinity".

To violate causality in a smooth way, the generators of $H^+(S)$ would have to emanate from infinity. This can happen even for the linear wave equation if S extends to left and/or right past null infinity in 2d Minkowski spacetime. In this case, $H^+(S)$ exists because information can enter the spacetime from past null infinity without crossing S. This is rather uninteresting (unrelated to any violation of causality) so consider instead the case of S extending to (left and right) spatial infinity in 2d Minkowski spacetime. For such S there is no Cauchy horizon for the linear wave equation so now consider such S for a nonlinear equation of the form (5.1). Assume that the initial data (u,du) is compactly supported on S. Under time evolution, the u field can propagate out to future null infinity. In $2d$, even for the linear wave equation solutions do not decay at null infinity, so u does not necessarily decay near future null infinity. This implies that g may not approach m near future null infinity. So perhaps causality violation could originate at infinity with a Cauchy horizon forming at left and/or right future null infinity and propagate into the interior of the spacetime along null geodesics of g which are spacelike w.r.t. m. It would be interesting to find an example for which this behaviour occurs.

5.3.3 Born-Infeld scalar

In two dimensional Minkowski spacetime, consider a scalar field with equation of motion obtained from the Born-Infeld action

$$S = -\frac{1}{c}\int d^2x\sqrt{1+cm^{\mu\nu}\partial_\mu\Phi\partial_\nu\Phi} \tag{5.7}$$

where c is a constant. By rescaling the coordinates we can set $c=\pm 1$. The case $c=1$ is the standard Born-Infeld theory. This theory is referred to as "exceptional" because, unlike in most nonlinear theories, a wavepacket in this theory propagates without distortion and never forms a shock [174].

The equation of motion is

$$g^{\mu\nu}\partial_\mu\partial_\nu\Phi = 0 \tag{5.8}$$

where

$$g^{\mu\nu} = m^{\mu\nu} - \frac{cm^{\mu\rho}m^{\nu\sigma}\partial_\rho\Phi\partial_\sigma\Phi}{\left(1+cm^{\lambda\tau}\partial_\lambda\Phi\partial_\tau\Phi\right)} \tag{5.9}$$

The inverse of $g^{\mu\nu}$ is

$$g_{\mu\nu} = m_{\mu\nu} + c\partial_\mu\Phi\partial_\nu\Phi \tag{5.10}$$

A calculation gives

$$\det g_{\mu\nu} = -\left(1+cm^{\rho\sigma}\partial_\rho\Phi\partial_\sigma\Phi\right) \tag{5.11}$$

Hence g is a Lorentzian metric (i.e. the equation of motion is hyperbolic) if, and only if,

$$1+cm^{\rho\sigma}\partial_\rho\Phi\partial_\sigma\Phi > 0 \tag{5.12}$$

In the language of section 5.2.1, a hyperbolic solution must satisfy this inequality.

Consider a vector V^μ. Note that

$$m_{\mu\nu}V^\mu V^\nu = g_{\mu\nu}V^\mu V^\nu - c\left(V\cdot\partial\Phi\right)^2 \tag{5.13}$$

If $c=1$ then the final term is non-positive. Hence if V is causal w.r.t. $g_{\mu\nu}$ then V is causal w.r.t. $m_{\mu\nu}$, i.e., the null cone of g lies on or inside that of m. However, for $c=-1$, the null cone of m lies on or inside that of g. Hence the $c=+1$ theory is subluminal and the $c=-1$ is superluminal according to the definitions of section 5.2.1.

The two theories are related by the transformation $(c,m,g)\to(-c,-m,-g)$ with Φ fixed. This is the map described in section 5.3.1.

5.3.4 Relation to Nambu-Goto string

It is well-known that the $c=1$ theory is a gauge-fixed version of an infinite Nambu-Goto string whose target space is $2+1$ dimensional Minkowski spacetime. The same is true for $c=-1$ except that the target space now has $+--$ signature, i.e., two time dimensions. The action of such a string is

$$S_{NG} = -\int d^2x\sqrt{-\det g} \tag{5.14}$$

where

$$g_{\mu\nu} = G_{AB}\partial_\mu X^A\partial_\nu X^B \tag{5.15}$$

with $G_{AB} = \text{diag}(-1,1,c)$ ($c = \pm 1$), x^μ are worldsheet coordinates, and $X^A(x)$ are the embedding coordinates of the string. It is assumed that the worldsheet of the string is timelike, i.e., that $g_{\mu\nu}$ has Lorentzian signature. Fixing the gauge as

$$x^0 = X^0 \qquad x^1 = X^1 \tag{5.16}$$

and defining $\Phi(x) = X^2(x)$, the action reduces to that of the Born-Infeld scalar described above, and the worldsheet metric $g_{\mu\nu}$ is the same as the effective metric given by equation (5.10). Note that the $c = \pm 1$ theories are mapped to each other under the transformation $(G,g) \to (-G,-g)$. From the worldsheet point of view, this corresponds to interchanging the definitions of timelike and spacelike, as discussed above.

Although the Born-Infeld scalar can be obtained from the Nambu-Goto string, we will not regard them as equivalent theories. We will view the BI scalar as a theory defined in a global 2-dimensional Minkowski spacetime. No such spacetime is present for the Nambu-Goto string. Of course any solution of the BI scalar theory can be "uplifted" to give some solution for the Nambu-Goto string. However, the converse is not true because not all solutions of the Nambu-Goto string can be written in the gauge (5.16). In particular, string profiles which "fold back" on themselves as in Fig. 5.1 are excluded by this gauge choice. From the BI perspective, such configurations will look singular. Of course such singularities can be eliminated by returning to the Nambu-Goto picture. However, we will not do this: the point is that the BI scalar is our guide to possible behaviour of nonlinear scalar field theories in 2d Minkowski spacetime, and most such theories do not have any analogue of the Nambu-Goto string interpretation.

5.3.5 Non-uniqueness

We can use the Nambu-Goto string to explain heuristically why there is a problem with the *subluminal* Born-Infeld scalar theory. (The superluminal case is harder to discuss heuristically because in this case the Nambu-Goto target space has two time directions.) Consider a left moving and a right moving wavepacket propagating along the string. As we will review below, if the wavepackets are sufficiently strong, when they intersect then the string can fold back on itself as described above. This is shown in Fig. 5.1. When this happens, the field Φ "wants to become multi-valued". But this is not possible in the BI theory because Φ is a scalar field in 2d Minkowski spacetime so Φ must be single-valued.

Clearly we have to "choose a branch" of the solution Φ at each point of 2d Minkowsi spacetime. We want to do this so that the solution is as smooth as possible. There are two obvious ways of doing this. We could start from the left of the string and extend until we

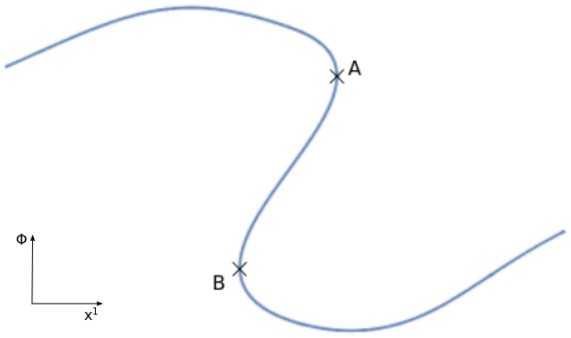


Fig. 5.1 An example of the string folding back on itself. The gradient is infinite at points A and B.

reach the point A of infinite gradient as shown in Fig 5.1. But beyond this point we have to jump to the other branch, so the solution is discontinuous as shown in Fig. 5.2. If the discontinuity is approached from the left then the gradient of Φ diverges as we approach A. However, if approached from the right the gradient remains bounded up to the discontinuity at A. Following out this procedure for the full spacetime produces a globally defined solution of the Born-Infeld theory. After some time, the wavepackets on the Nambu-Goto string separate and the resulting Born-Infeld solution becomes continuous again.

Now note that instead of starting on the left and extending to point A we could have started on the right and extended to point B. Now the discontinuity would occur at B instead of A. So now the solution appears as shown in Fig. 5.3. Approaching the discontinuity from the right, the gradient of Φ diverges at B. However approaching from the left, the gradient remains bounded up to the discontinuity at B. As above, this procedure gives a globally defined solution of the Born-Infeld theory. This is clearly a *different* solution from the solution discussed in the previous paragraph.

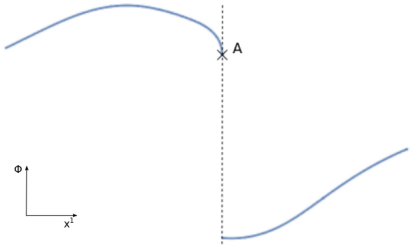


Fig. 5.2 Solution with discontinuity at A.

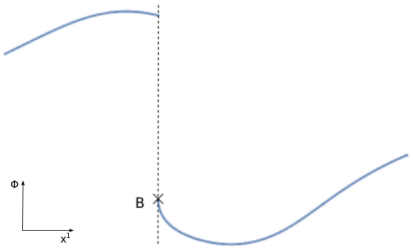


Fig. 5.3 Different solution with discontinuity at B.

Starting from initial data prescribed on some line S in the far past, the above constructions produce two *different* solutions which agree with the data on S. Now non-uniqueness is to be expected because the solution Φ is singular (at A or B), so the corresponding spacetimes

(M,g) will not be globally hyperbolic. Therefore lack of uniqueness is to be expected beyond the Cauchy horizon. However, we will show, in the subluminal case, that the lack of uniqueness *occurs before a Cauchy horizon forms*. In other words, the two solutions disagree in a region which belongs to $D^+(S)$ for both solutions. This implies that the two solutions cannot arise from the same MGHD of the data on S. Therefore MGHDs are not unique.

Clearly there are other ways we could construct Born-Infeld solutions from the Nambu-Goto solution: we do not have to take the discontinuity to occur at either point A or at point B, we could take it to occur at any point between A and B. This leads to an infinite set of possible solutions, and an infinite set of distinct MGHDs.

The above discussion was for the subluminal ($c = 1$) theory. We will show below that this problem does not occur for the superluminal theory. This is because, in the superluminal theory, from the 2d Born-Infeld perspective, A and B are *timelike* separated with B (say) occuring to the future of A. This implies that B lies to the future of the infinite gradient singularity at A hence B cannot belong to $D^+(S)$ if S is a surface to the past of A. Therefore there is a unique choice of branch in the superluminal theory. In this theory there is a unique MGHD.

5.3.6 General Solution

The $c = 1$ (subluminal) BI scalar theory was solved by Barbashov and Chernikov [168, 169]. We will follow the notation of Whitham [175], who gives a nice summary of their work. Because the superluminal and subluminal theories are related as discussed above, it is easy to write down the general hyperbolic[7] solution for both cases. Write the Minkowski metric as

$$m = -c(dx^0)^2 + c(dx^1)^2 \tag{5.17}$$

and define null coordinates

$$\xi = x^1 - x^0 \qquad \eta = x^1 + x^0 \tag{5.18}$$

The solution is written in terms of a mapping $\Psi : \mathbb{R}^2 \to \mathbb{R}^2$ given by

$$\Psi : (\rho,\sigma) \mapsto (\xi(\rho,\sigma), \eta(\rho,\sigma)) \tag{5.19}$$

[7]This solution was obtained using the method of characteristics which only works when the equation is hyperbolic so only hyperbolic solutions are obtained using this method.

where

$$\xi(\rho,\sigma) = \rho - \int_{-\infty}^{\sigma} \Phi_2'(x)^2 dx \tag{5.20}$$

and

$$\eta(\rho,\sigma) = \sigma + \int_{\rho}^{\infty} \Phi_1'(x)^2 dx \tag{5.21}$$

with $\Phi_1(\rho)$ and $\Phi_2(\sigma)$ smooth functions such that Φ_1' and Φ_2' decay at infinity fast enough to ensure that the integrals converge.[8] These two functions can be viewed as specifying the profiles of left moving and right moving wavepackets.

Assuming that Ψ is invertible we can write $\rho = \rho(\xi,\eta)$ and $\sigma = \sigma(\xi,\eta)$ and the solution is given by

$$\Phi(\xi,\eta) = \Phi_1(\rho(\xi,\eta)) + \Phi_2(\sigma(\xi,\eta)) \tag{5.22}$$

We can state the above result as a theorem [168, 169, 175]:

Theorem 5.3.1. *Let $\Phi_1(\rho)$ and $\Phi_2(\sigma)$ be smooth functions defined for all $(\rho,\sigma) \in \mathbb{R}^2$. Let V be a connected open subset of $\mathbb{R}^2$. If the map $\Psi : V \to U \subset \mathbb{R}^2$ defined by (5.20) and (5.21) is a diffeomorphism then (5.22) defines a smooth solution $\Phi : U \to \mathbb{R}$ of the Born-Infeld scalar equation of motion.*

Clearly it will be important to determine whether or not Ψ is a diffeomorphism.

Lemma 5.3.1. *A necessary (although not sufficient) condition for $\Psi : V \to U$ to be a diffeomorphism is that either $\Phi_1'(\rho)^2\Phi_2'(\sigma)^2 < 1$ throughout V or $\Phi_1'(\rho)^2\Phi_2'(\sigma)^2 > 1$ throughout V.*

Proof. The Jacobian of the map Ψ is

$$\det \frac{\partial(\xi,\eta)}{\partial(\rho,\sigma)} = 1 - \Phi_1'(\rho)^2\Phi_2'(\sigma)^2 \tag{5.23}$$

hence a necessary condition for Ψ to define a diffeomorphism is that the RHS cannot vanish at any point of V. Since V is connected the result follows immediately. □

A point on the boundary ∂V at which $\Phi_1'\Phi_2' = 1$ corresponds to a singularity:

Lemma 5.3.2. *Assume that $\Psi : V \to U$ is a diffeomorphism such that $\Phi_1'(\rho)\Phi_2'(\sigma) \to 1$ as $(\rho,\sigma) \to (\rho_0,\sigma_0)$ for some $(\rho_0,\sigma_0) \in \partial V$. Let $\gamma : (0,1) \to V$ be a smooth curve with $\gamma(t) \to (\rho_0,\sigma_0)$ as $t \to 1$. Then the gradient of the solution Φ at the point $\Psi(\gamma(t))$ diverges as $t \to 1$.*

[8] The latter assumption could be relaxed by replacing the infinite limits of the integrals by finite constants.

Proof. A calculation gives

$$\partial_\xi \Phi = \frac{\Phi_1'(\rho)}{1-\Phi_1'(\rho)\Phi_2'(\sigma)} \qquad \partial_\eta \Phi = \frac{\Phi_2'(\sigma)}{1-\Phi_1'(\rho)\Phi_2'(\sigma)} \tag{5.24}$$

The result follows immediately. □

It can be shown similarly that points of ∂V where $\Phi_1'\Phi_2' = -1$ correspond to a divergence in the second derivative of Φ although we will not need this result below.

We will be mainly interested in causal properties of the metric g defined by (5.10). If $\Psi : V \to U$ is a diffeomorphism then we can introduce (ρ,σ) as coordinates on V. The metric g defined by (5.10) takes a simple form in these coordinates:

Lemma 5.3.3. *Consider a Born-Infeld solution constructed as in Theorem 5.3.1. In coordinates (ρ,σ), the metric (5.10) is*

$$g = c\left(1+\Phi_1'(\rho)\Phi_2'(\sigma)\right)^2 d\rho d\sigma \tag{5.25}$$

Proof. Direct calculation using (5.20), (5.21) and (5.22). □

Note that the vector fields $\partial/\partial\rho$ and $\partial/\partial\sigma$ are null w.r.t. g. Let's determine whether they are future or past directed. Recall (section 5.2.1) that the time-orientation for g is determined by a choice of time orientation for Minkowski spacetime.

Lemma 5.3.4. *Consider a Born-Infeld solution constructed as in Theorem 5.3.1. In the subluminal case, $\partial/\partial\rho$ is past-directed and $\partial/\partial\sigma$ is future-directed w.r.t. g. In the superluminal case, if $\Phi_1'(\rho)^2\Phi_2'(\sigma)^2 < 1$ then $\partial/\partial\rho$ and $\partial/\partial\sigma$ are both future directed whereas if $\Phi_1'(\rho)^2\Phi_2'(\sigma)^2 > 1$ then they are both past-directed. In either case the spacetime (U,g) is stably causal.*

Proof. In the subluminal case ($c=1$) we know (section 5.2.1) that x^0 is a global time function for the spacetime (U,g) so this spacetime is stably causal. From (5.20) and (5.21) one finds $\partial x^0/\partial\rho < 0$ and $\partial x^0/\partial\sigma > 0$ and the result follows.

In the superluminal case ($c=-1$), $\partial/\partial x^1$ is timelike w.r.t. m so (section 5.2.1) we choose $\partial/\partial x^1$ as a time-orientation on (V,g). A calculation gives

$$\frac{\partial}{\partial x^1} = \frac{1}{1-\Phi_1'(\rho)^2\Phi_2'(\sigma)^2}\left[(1+\Phi_2'(\sigma)^2)\frac{\partial}{\partial\rho} + (1+\Phi_1'(\rho)^2)\frac{\partial}{\partial\sigma}\right] \tag{5.26}$$

The inner products (w.r.t. g) of $\partial/\partial x^1$ with $\partial/\partial\rho$ and $\partial/\partial\sigma$ can be calculated using (5.25). Clearly these inner products have the opposite sign to $1-\Phi_1'(\rho)^2\Phi_2'(\sigma)^2$ and so $\partial/\partial\rho$ and $\partial/\partial\sigma$ are both future directed if this quantity is positive and past directed if it is negative. If $\Phi_1'(\rho)^2\Phi_2'(\sigma)^2<1$ then let $X=\partial/\partial\rho+\partial/\partial\sigma$, which is future-directed and timelike w.r.t. g. We then have $g_{\mu\nu}X^\nu\propto-[d(\rho+\sigma)]_\mu$ hence $\rho+\sigma$ is a global time function for (U,g) and so (U,g) is stably causal. Similarly if $\Phi_1'(\rho)^2\Phi_2'(\sigma)^2>1$ then $-(\rho+\sigma)$ is a global time function for (U,g). □

In the superluminal case, this proves that solutions constructed using Theorem 5.3.1 cannot exhibit any violation of causality. However, we note that there may be solutions of (5.8) that cannot be obtained using Theorem 5.3.1. Such solutions would requires multiple charts V_α, each with corresponding coordinates $(\rho_\alpha,\sigma_\alpha)$ and diffeomorphisms Ψ_α. In any given chart the solution will take the form described above. With multiple charts, it may not be possible to construct a global time function for the superluminal theory.

We are interested in globally hyperbolic developments of initial data. It is very easy to determine whether or not a solution constructed using Theorem 5.3.1 is globally hyperbolic:

Lemma 5.3.5. *Consider a Born-Infeld solution constructed as in Theorem 5.3.1. Then (U,g) is globally hyperbolic with Cauchy surface S if, and only if, $(V,\hat{m})$ is globally hyperbolic with Cauchy surface $\Sigma=\Psi^{-1}(S)$, where $\hat{m}=cd\rho d\sigma$.*

Proof. This is an immediate consequence of (5.25) which shows that g and $\hat{m}$ define causally equivalent metrics on V. (Here we are not bothering to distinguish the metric g on U and the metric on V defined by pull-back of g w.r.t. Ψ.) □

Thus global hyperbolicity can be checked using the flat metric $\hat{m}$ on V. More generally, the causal properties of (U,g) are the same as those of the flat spacetime $(V,\hat{m})$.

We will show that, given initial data on a surface S, there exist multiple distinct maximal globally hyperbolic developments in the subluminal case ($c=1$) but there is a unique MGHD in the superluminal ($c=-1$) case. This difference can be traced to the following property:

Lemma 5.3.6. *Let p,q be distinct points such that $\Psi(p)=\Psi(q)$. Then the straight line connecting p,q in the (ρ,σ) plane is spacelike w.r.t. $\hat{m}$ in the subluminal case and timelike in the superluminal case.*

Proof. Let p and q have coordinates (ρ_2,σ_2) and (ρ_1,σ_1) respectively. From equations (5.20) and (5.21) we have

$$\delta\rho\equiv\rho_2-\rho_1=\int_{\sigma_1}^{\sigma_2}\Phi_2'(x)^2dx,\qquad\delta\sigma\equiv\sigma_2-\sigma_1=\int_{\rho_1}^{\rho_2}\Phi_1'(x)^2dx. \tag{5.27}$$

From the first equation we see that $\delta\sigma = 0$ implies $\delta\rho = 0$ and the second equation gives the converse. Hence $\delta\rho = 0$ if, and only if, $\delta\sigma = 0$, i.e., $p = q$. Since we are assuming $p \neq q$ we must have $\delta\rho \neq 0$ and $\delta\sigma \neq 0$. The first equation then implies that $\delta\rho$ has the same sign as $\delta\sigma$ so

$$\delta\rho\,\delta\sigma > 0. \tag{5.28}$$

The result follows from the definition of $\hat{m}$ in Lemma 5.3.5. □

Theorem 5.3.1 defines a solution in a subset U of Minkowski spacetime. The following theorem [169] guarantees a *global* solution:

Theorem 5.3.2. *Let Φ_1 and Φ_2 be smooth functions on the real line such that the integrals in (5.20) and (5.21) converge for $\rho \to -\infty$ and $\sigma \to \infty$. Assume that $\Phi_1'(\rho)^2\Phi_2'(\sigma)^2 < 1$ for all (ρ,σ). Then the map $\Psi : \mathbb{R}^2 \to \mathbb{R}^2$ defined by (5.20), (5.21) is a diffeomorphism and so the Born-Infeld solution of Theorem 5.3.1 is a globally defined smooth solution.*

Proof. Following [169], use (5.21) to write

$$\sigma = \sigma_\eta(\rho) \equiv \eta - \int_\rho^\infty \Phi_1'(x)^2 dx \tag{5.29}$$

and then substitute into (5.20) to obtain

$$\xi = F(\rho;\eta) \equiv \rho - \int_{-\infty}^{\sigma_\eta(\rho)} \Phi_2'(x)^2 dx \tag{5.30}$$

We want to use this equation to determine ρ as a function of ξ,η. A calculation gives

$$\left(\frac{\partial F}{\partial \rho}\right)_\eta = 1 - \Phi_1'(\rho)^2\Phi_2'(\sigma_\eta(\rho))^2 \tag{5.31}$$

So $\Phi_1'(\rho)^2\Phi_2'(\sigma)^2 < 1$ implies that F is a strictly increasing function of ρ and hence there exists at most one solution ρ of (5.30) for any (ξ,η). Given a solution for ρ, (5.29) determines σ uniquely. This proves that the map Ψ is injective.

We now show that there exists exactly one solution of (5.30). Our assumptions on Φ_1 imply that $\sigma_\eta(\rho) \to \eta$ as $\rho \to \infty$ and $\sigma_\eta(\rho) \to \eta - C$ as $\rho \to -\infty$ where $C = \int_{-\infty}^{\infty} \Phi_1'(x)^2 dx$. Our assumptions on Φ_1 imply that $\Phi_1'(\rho) \to 0$ as $\rho \to \pm\infty$. So now from (5.31) we see that $(\partial F/\partial\rho)_\eta \to 1$ as $\rho \to \pm\infty$. So, at fixed η, F is strictly increasing and has gradient 1 for $\rho \to \pm\infty$. This implies that, at fixed η, the map $\rho \to F(\rho;\eta)$ is a bijection from $\mathbb{R}$ to itself. Hence there exists exactly one solution of (5.30) for given (ξ,η). Hence Ψ is a bijection. That Ψ is a diffeomorphism now follows from the fact that the RHS of equation (5.23) is everywhere non-zero. □

Lemma 5.3.7. *The solution of Theorem 5.3.2 is globally hyperbolic.*

Proof. This follows immediately from Lemma 5.3.5 because $(V,\hat{m}) = (\mathbb{R}^2,\hat{m})$ so the causal structure w.r.t. g is the same as 2d Minkowski spacetime. In the subluminal case, surfaces of constant x^0 are Cauchy because x^0 is a global time function. In the superluminal case, a surface of constant $\rho+\sigma$ is Cauchy since the proof of Lemma 5.3.4 shows that $\rho+\sigma$ is a global time function. □

As discussed above, we need Ψ to be a diffeomorphism for equations (5.20), (5.21), (5.22) to define a solution of the Born-Infeld scalar. However, we note that these equations define a solution of the Nambu-Goto string irrespective of whether or not Ψ is a diffeomorphism. To see this, take (ρ,σ) as worldsheet coordinates and replace the LHS of (5.20) and (5.21) by $X^1 - X^0$ and $X^1 + X^0$ respectively. Together with $X^2 = \Phi = \Phi_1(\rho) + \Phi_2(\sigma)$ this specifies a globally well defined embedding of the string worldsheet into $\mathbb{R}^3$. The worldsheet metric is (5.25). The solution describes a superposition of left moving and right moving wavepackets described by $\Phi_1(\rho)$ and $\Phi_2(\sigma)$, each travelling at the speed of light with respect to g. The worldsheet metric degenerates at points where $\Phi_1'(\rho)\Phi_2'(\sigma) = -1$. These correspond to "cusp" singularities at which the string worldsheet becomes null. The string is smooth at points where $\Phi_1'(\rho)\Phi_2'(\sigma) = +1$, which correspond to points of infinite gradient like A or B in Fig. 5.1.

5.3.7 Example of non-uniqueness in subluminal case

We start by recording that the subluminal Born-Infeld scalar equation of motion (5.8) written out in coordinates x^μ reduces to:

$$-\left(1+(\partial_{x^1}\Phi)^2\right)\partial_{x^0}^2\Phi + 2\partial_{x^0}\Phi\partial_{x^1}\Phi\cdot\partial_{x^0}\partial_{x^1}\Phi + \left(1-(\partial_{x^0}\Phi)^2\right)\partial_{x^1}^2\Phi = 0\,. \tag{5.32}$$

In this section we will demonstrate the existence of two different maximal globally hyperbolic developments (MGHDs) arising from the same initial data for the above equation. We will do this with an example involving a specific choice of the functions Φ_1 and Φ_2, and construct solutions using Theorem 5.3.1.

To construct a solution of (5.32) we choose functions

$$\Phi_1(x) = \Phi_2(x) = \phi(x) \equiv \int_{-\infty}^{x} ae^{-t^2}dt \tag{5.33}$$

where $a > 1$ is a constant. This gives $\Phi_1'(x) = \Phi_2'(x) = \phi'(x) := ae^{-x^2}$. Hence $\Phi_1'(\rho)^2\Phi_2'(\sigma)^2 = a^2e^{-r^2}$ where $r = \sqrt{\rho^2+\sigma^2}$. Let $r_0 = \sqrt{2\ln(a)}$. In the (ρ,σ) plane we have

$$\begin{aligned} &\Phi_1'(\rho)\Phi_2'(\sigma) < 1 \quad && \text{outside the circle of radius } r_0 \\ &\Phi_1'(\rho)\Phi_2'(\sigma) > 1 \quad && \text{inside the circle of radius } r_0 \\ &\Phi_1'(\rho)\Phi_2'(\sigma) = 1 \quad && \text{on the circle of radius } r_0\ . \end{aligned} \tag{5.34}$$

Theorem 5.3.2 does *not* apply, and we do not have a global solution. Indeed the map Ψ defined by this choice of Φ_1 and Φ_2 is not injective on $\mathbb{R}^2$. In section 5.3.7.1 we will determine numerically the region in which injectivity fails and explain heuristically how this leads to non-uniqueness of MGHDs.

5.3.7.1 Numerical demonstration of non-uniqueness of MGHDs

Step 1. We start by showing that, for the example (5.33), Ψ is non-injective on $\mathbb{R}^2$ but its restriction to a subset V' of $\mathbb{R}^2$ *is* injective and so we obtain a solution of (5.32) via Theorem 5.3.1.

The region in which injectivity of Ψ fails can be determined numerically[9] and is shown in Fig. 5.4: three open regions D, E and F of the (ρ,σ) plane map to the same region X of Minkowski spacetime. Here D is the disc $r < r_0$. The region $X \equiv \Psi(D)$ is shown in Fig. 5.5. The inverse image of any point in X consists of three points, one in each of D, E and F.[10] However, the map Ψ *is* injective on $V' \equiv \mathbb{R}^2\backslash\overline{D\cup E\cup F}$ and (5.34) implies that the condition of Lemma 5.3.1 is satisfied on V' so Ψ defines a diffeomorphism from V' to $U' \equiv \Psi(V') = \mathbb{R}^2\backslash\overline{X}$. Hence Theorem 5.3.1 defines a solution $\Phi : U' \to \mathbb{R}^2$ of (5.32).

Step 2. Next we will show that the solution $\Phi : U' \to \mathbb{R}^2$ is not a GHD but, by restricting its domain, we can construct a GHD.

Lemma 5.3.5 establishes that (U', g) is globally hyperbolic if, and only if, $(V', \hat{m})$ is globally hyperbolic. Introduce coordinates (y^0, y^1) in the (ρ,σ) plane such that

$$\rho = y^1 - y^0 \qquad \sigma = y^1 + y^0. \tag{5.35}$$

In these coordinates we have

$$\hat{m} = -(dy^0)^2 + (dy^1)^2 \tag{5.36}$$

[9] These plots were determined using the *FindRoot* function in Mathematica to numerically construct an inverse function. A different starting point for the numerics was used for each region.

[10] In the Nambu-Goto string interpretation, X is is the region of spacetime in which the string worldsheet folds back on itself as in Fig. 5.1.

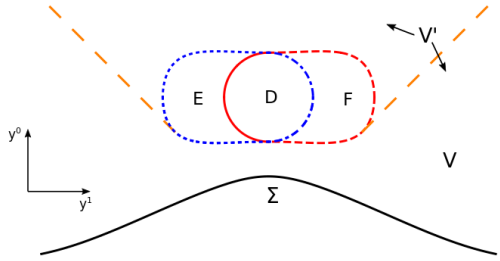


Fig. 5.4 Plot of the (ρ,σ) plane in coordinates (y^0,y^1) defined by (5.35). The open sets D,E,F have the same image under Ψ. The dotted blue (dashed red) curve has the same image as the dot-dashed blue (solid red) curve. Ψ is injective on V', the complement of $\overline{D\cup E\cup F}$. The orange large dashed lines are the future Cauchy horizon for the initial surface Σ in the flat spacetime $(V',\hat{m})$. V is the region of V' lying to the past of this Cauchy horizon.

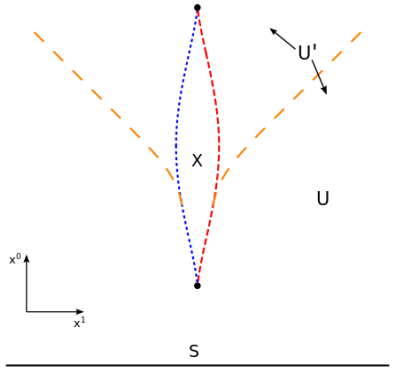


Fig. 5.5 Minkowski spacetime with coordinates (x^0,x^1). The region X is the image of D (or E or F) under Ψ. The region U' is the complement of $\overline{X}$. The two black dots are points at which the gradient of the solution $\Phi:U'\to\mathbb{R}$ diverges (by Lemma 5.3.2). The orange large dashed lines are the future Cauchy horizon of S in the spacetime (U',g). U is the region of U' lying to the past of this Cauchy horizon. $\Phi:U\to\mathbb{R}$ is a GHD of the initial data on S.

and Lemma 5.3.4 implies that $\partial/\partial y^0$ is future-directed. The causal properties of $\hat{m}$ (and hence g) in the (ρ,σ) plane are easy to read off from Fig. 5.4. In particular it is clear that the region V' is not globally hyperbolic w.r.t. $\hat{m}$ so U' is not globally hyperbolic w.r.t. g. Consider an initial surface S defined by $x^0=-T$, as shown in Fig. 5.5. Let U be the domain of dependence of S in (U',g). Then by restricting Φ to U we obtain a GHD $\Phi:U\to\mathbb{R}$ of the initial data on S. Appealing to Lemma 5.3.5, $U=\Psi(V)$ where V is the domain of dependence of $\Sigma\equiv\Psi^{-1}(S)$ in $(V',\hat{m})$. Viewed as a subset of V', V is bounded by the future Cauchy horizon shown in Fig. 5.4, which maps to a corresponding future Cauchy horizon in Fig. 5.5.

Step 3. Now we will show that the GHD $\Phi:U\to\mathbb{R}$ is not maximal and it can be smoothly extended to give a GHD $\Phi_a:U_a\to\mathbb{R}$ that contains part of region X. We will show that this extended GHD is smooth on the "left" boundary of X but singular on the "right" boundary of X.

We enlarge the GHD $\Phi:U\to\mathbb{R}$ by pushing the left large dashed orange line of Fig. 5.4 into region E until it is tangent to the boundary of D. Specifically, consider the region V_a defined in Fig. 5.6. Since V_a contains no points of D or F, the map Ψ is still injective on this enlarged region and still satisfies (5.23), hence Ψ is a diffeomorphism and so Theorem 5.3.1 defines a solution $\Phi_a:U_a\to\mathbb{R}$ where $U_a=\Psi(V_a)$. Furthermore, $(V_a,\hat{m})$ is globally

hyperbolic with Cauchy surface Σ and so (U_a, g) is globally hyperbolic with Cauchy surface S. Hence Φ_a is a GHD of the initial data on S. The region U_a is shown in Fig. 5.7: it extends across the left boundary of X all the way to the right boundary of X. This right boundary is not part of U_a, indeed the solution Φ_a is discontinuous across this boundary.[11]

Consider a curve γ_1 approaching the right boundary of X from the left (i.e. from within X) as in Fig. 5.7. Then $\Psi^{-1}(\gamma_1)$ is a curve approaching the solid red curve of Fig. 5.6 from within E. Since $\Phi_1'\Phi_2' = 1$ on this red curve, Lemma 5.3.2 implies that the gradient of Φ_a diverges along γ_1 as one approaches the boundary. Thus the gradient of Φ_a diverges along the right boundary of X when approached from the left. On the other hand, if γ_2 is a curve approaching this boundary from the right (i.e. from outside X) as in Fig. 5.7 then $\Psi^{-1}(\gamma_2)$ approaches the dotted red curve of Fig. 5.6, which is in the region where $\Phi_1'\Phi_2' < 1$ so the gradient of Φ_a remains bounded along γ_2. Hence the gradient of Φ_a is bounded as one approaches the right boundary of X from outside X.

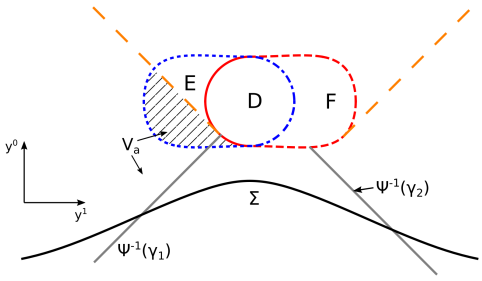


Fig. 5.6 The large dashed orange line on the left is a line of constant σ which is tangent to the boundary of D at their point of contact. The region V_a is the union of V with the region to the past (w.r.t. $\hat{m}$) of this line and the shaded section of E. The future bounday of V_a consists of the pair of large dashed orange null lines together with the (spacelike) sections of the solid and dashed red curves that connect them.

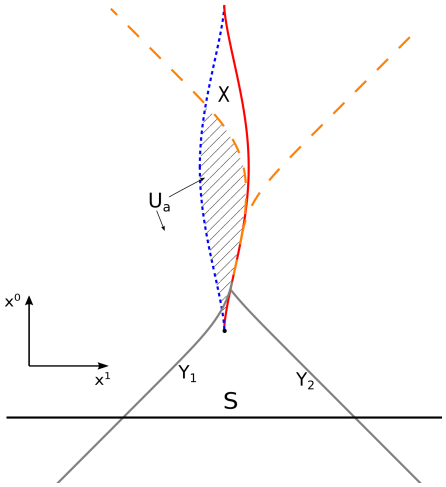


Fig. 5.7 The region $U_a = \Psi(V_a)$ contains part of the left boundary of X and extends up to the right boundary of X, where the gradient of the solution Φ_a diverges. The future (w.r.t. g) boundary of U_a consists of the large dashed orange curves (null w.r.t. g) and a section of the right boundary of X (spacelike w.r.t. g) starting at the black dot.

Step 4. Finally we show that there is a *different* way of extending $\Phi : U \to \mathbb{R}$ to give a GHD and that this implies non-uniqueness of MGHDs.

[11] In the Nambu-Goto string interpretation, the string worldsheet on a surface of constant x^0 intersecting X resembles Fig. 5.2 with point A on the right boundary of X.

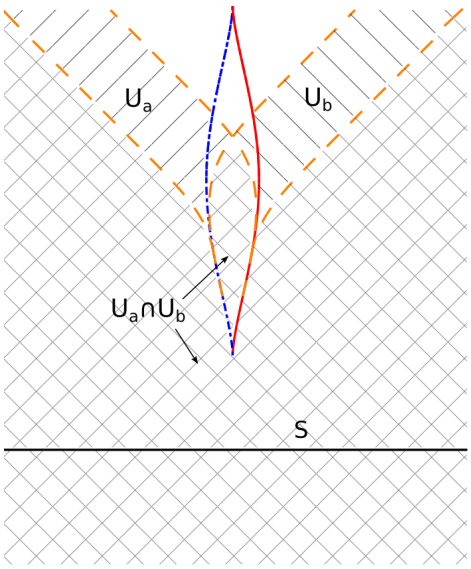


Fig. 5.8 The regions U_a and U_b are given by the right/left hatching respectively. The intersection of these regions is disconnected, with one component lying inside X and the other component (containing S) outside X.

We construct this new extension of $\Phi : U \to \mathbb{R}$ as follows. Define V_b to be the reflection of V_a under $y^1 \to -y^1$. So V_b is an extension of V into region F. Everything we've said about V_a is true also of V_b and so this defines another GHD $\Phi_b : U_b \to \mathbb{R}$ where $U_b = \Psi(V_b)$. In this case, U_b extends across the *right* boundary of X all the way the the left boundary of X, where the gradient of Φ_b diverges when approaching from the right.[12]

We now have two different GHDs of the same intial data on S, $\Phi_a : U_a \to \mathbb{R}$ and $\Phi_b : U_b \to \mathbb{R}$. These two solutions agree in U but they differ in X because Φ_a has divergent gradient on the right boundary of X whereas Φ_b has divergent gradient on the left boundary of X. Thus the corresponding maximal GHDs must differ in X. *This demonstrates the non-uniqueness of maximal GHDs for* (5.32).

We will now discuss this result and highlight properties of our example that are relevant to the general results of Section 5.4.

Consider the intersection $U_a \cap U_b$ shown in Fig. 5.8. Note that this is *disconnected*, consisting of two connected components. One component contains S but no points of X and the other component is a subset of X. The two solutions agree on the former component but they disagree on the latter component. In Section 5.4 we will see that this disconnectedness is a necessary condition for two GHDs to differ in some region.

[12] In the Nambu-Goto string interpretation, this corresponds to Fig. 5.2 with point B on the left boundary of X.

Another point to emphasize is that the boundary of U_a consists of a section (along the right boundary of X, between the lower black dot and the orange curves of Fig. 5.7), which can be approached from both sides (either the left or the right) within U_a. In other words U_a *lies on both sides of its boundary*. (The same is true for U_b.) In section 5.4 we will see that this property is a necessary condition for non-uniqueness of MGHDs.

We have shown that there exist two distinct MGHDs arising from the same data on S. In fact one can show that there are infinitely many such MGHDs (cf section 5.3.5). The different MGHDs all agree in the region U but they differ in X. In section 5.4 we will define the *maximal unique globally hyperbolic development* (MUGHD) $\Phi_{\max} : R \to \mathbb{R}$ of the initial data on S as follows. R is the largest open subset of Minkowski spacetime on which the solution is *uniquely* determined by the data on S. Such a development is necessarily globally hyperbolic with Cauchy surface S. For the above example, we have $R = U$ and $\Phi_{\max} = \Phi$. As we have seen, the solution $\Phi : U \to \mathbb{R}$ can be extended, whilst maintaining global hyperbolicity, but not in a unique way. From Figs 5.4 and 5.5 we see that the future boundary of R consists of a singular point (the lower black dot in Fig. 5.5) from which emanate a pair of spacelike (w.r.t. g) curves which connect to a pair of null (w.r.t. g) curves. The solution can be smoothly, but not uniquely, extended across these spacelike and null curves.

The extendibility across the spacelike curves is a new kind of breakdown of predictability. Fig. 5.5 suggests that we should view these spacelike curves (the early time sections of the red and blue dotted curves) as a “consequence" of the formation of a singularity (the black dot). This interpretation is suggested if one uses x^0 as a time function (e.g. in a numerical simulation). However, since these curves are spacelike, they are not in causal contact with the singularity. Furthermore, it is just as legitimate to use y^0 as a time function. From this point of view, Fig. 5.4 shows that the spacelike curves form before (i.e. at earlier y^0) the singular point. So it is incorrect to ascribe the breakdown of predictability to the formation of the singularity.

This behaviour is worrying. Given a development of the data on S, there is no general way of determining, *from the solution itself*, which region of it belongs to the MUGHD. To determine this region one has to construct all GHDs with the same initial data! This is much worse than the failure of predictability associated with the formation of a Cauchy horizon because the location of a Cauchy horizon within a development can be determined from the solution itself.

How would the non-uniqueness of MGHDs manifest itself in, say, a numerical simulation? The answer is that the solution will depend not just on the initial data but also on the choice of time function. To see this, consider the globally hyperbolic development $\Phi_a : U_a \to \mathbb{R}$. Since S is a Cauchy surface we can choose a global time function for U_a such that S is a

surface of constant time. We can do the same for $\Phi_b : U_b \to \mathbb{R}$. Of course these two time functions are different but either could be used for a numerical evolution starting from the data on S. For points in the MUGHD U, the results of these two numerical evolutions will agree. However, for points in X, the results will disagree. In practice one would not know *a priori* which points belong to the MUGHD, i.e., one would not know in what region the results of the numerical evolution are independent of the choice of time function.[13]

Note that, for any solution, the domain of dependence of S defined using the Minkowski metric m is a subset of the domain of dependence of S defined using g. Hence a solution which is globally hyperbolic w.r.t. g is also globally hyperbolic w.r.t. m. We could therefore ask about uniqueness of MGHDs defined w.r.t. m instead of w.r.t. g. We'll refer to these as m-MGHDs. For the above example, there is indeed a unique m-MGHD: it is bounded to the future by two future-directed null (w.r.t. m) lines emanating from the lower black dot in Fig. 5.5. It was shown in [163] that any subluminal equation always admits a unique m-MGHD, which is a subset of the MUGHD. However, if the speed of propagation w.r.t. g is much less than the speed of propagation w.r.t. m then the m-MGHD will not be a very useful concept because it will not contain a large part of the MUGHD.

We have used the Born-Infeld scalar as an example exhibiting non-uniqueness of MGHDs. This example is rather artificial because there is a "more fundamental" underlying theory, namely the Nambu-Goto string, for which there is no problem with predictability. However, our point is that if this pathological feature can occur for a particular scalar field theory then it is to be expected to occur also for other scalar field theories for which there is no analogue of the Nambu-Goto string interpretation.

This is a heuristic discussion of our example of non-uniqueness; it is be made rigorous in the series of theorems presented in [163].

5.3.8 Uniqueness for superluminal case

Non-uniqueness of MGHDs is *not* a problem in the superluminal ($c = -1$) case. This is true for an arbitrary superluminal equation, as given in a theorem in section 5.4 below. In this section we will discuss briefly the interpretation of the example (5.33) in the superluminal case.

[13] Since we are dealing with a subluminal theory, one could just declare that x^0 is a preferred time function and ignore the above problems. However this is unsatisfactory: if one uses x^0 as the time function (with S a surface of constant x^0 at sufficiently early time) then from Fig. 5.5 the evolution must stop at the line of constant x^0 passing through the singularity corresponding to the (lower) black dot so one obtains only part of the MUGHD.

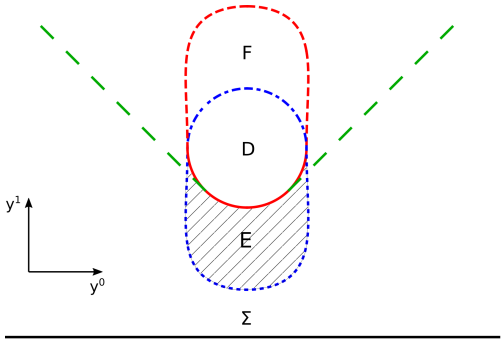


Fig. 5.9 In the superluminal case we orient the plot so that the time function y^1 is the vertical axis. The large dashed green lines are lines of constant ρ or σ that are tangent to the circle at their point of contact. The MGHD of the data on S is defined by choosing V to be the region bounded to the future by the pair of large dashed green lines together with the section of the solid red curve joining them. This includes the hatched section of E but not the two small regions of E between the large dashed green lines and D.

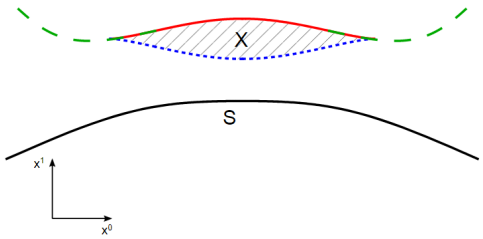


Fig. 5.10 Plot of $U = \Psi(V)$ in Minkowski spacetime, oriented so that x^1 is the vertical axis. The MGHD is the region bounded to the future by the spacelike (w.r.t. g) solid red curve and the pair of null (w.r.t. g) large dashed green curves. This includes most of the region X. The gradient of Φ diverges on the solid red curve. The solution can be smoothly extended across the large dashed green curves, but not as a GHD of the data on S.

In the superluminal case, recall that the Minkowski metric (5.17) is

$$m = -(dx^1)^2 + (dx^0)^2 \tag{5.37}$$

and we choose time orientation $\partial/\partial x^1$. Defining coordinates (y^0, y^1) in the (ρ, σ) plane as in (5.35) gives, for the flat metric of Lemma 5.3.5

$$\hat{m} = -(dy^1)^2 + (dy^0)^2 \tag{5.38}$$

Now consider the example (5.33). We want to construct a solution using Theorem 5.3.1 so assume that $\Psi : V \to U$ is a diffeomorphism. Lemma 5.3.1 implies that either $V \subset D$ or V lies outside D. We consider the latter case, so $\Phi_1'(\rho)^2\Phi_2'(\sigma)^2 < 1$ in V. The proof of Lemma 5.3.4 reveals that y^1 is a global time function for (U, g) (or $(V, \hat{m})$) so we take our initial surface $S = \Psi(\Sigma)$ where Σ is a line $y^1 = -Y$ where Y is large enough so that Σ lies to the past of $\overline{D \cup E \cup F}$ as shown in Fig. 5.9.[14]

[14] Equivalently we could define S to be a surface $x^0 = -T$ where T is chosen large enough to make S spacelike w.r.t. g. However, this would gives plots with a lot of white space between S (or Σ) and the region of interest.

The unique MGHD is obtained by taking V to be the region defined in Fig. 5.9. The future boundary of V is the union of a spacelike curve (a segment of the boundary of D) along which the gradient of Φ diverges (by Lemma 5.3.2), and a pair of null curves across which the solution is smoothly extendible (but not as a GHD).[15] The corresponding picture in Minkowski spacetime is shown in Fig. 5.10.

The reason that there is a unique MGHD in the superluminal case but not in the subluminal case was identified in Lemma 5.3.6. In the subluminal case, different GHDs can be constructed by including points from E or from F, or from both. But in the superluminal case, Lemma 5.3.6 implies that F lies to the future of D so from any point of F there is a past directed timelike curve that ends on the boundary of D and hence does not cross Σ. So no point of F can belong to the domain of dependence of Σ.

5.3.9 Higher dimensions

It is easy to see that the pathological behaviour in the subluminal case is not restricted to two spacetime dimensions. The Born-Infeld scalar field theory in $(d+1)$-dimensional Minkowski spacetime is defined by generalizing the action (5.7) to $d+1$ dimensions. The two dimensional theory can be obtained trivially from the $d+1$ dimensional theory by assuming that Φ does not depend on $d-1$ of the spatial coordinates. Hence our 2d solutions can be interpreted as solutions in $d+1$ dimensions with translational invariance in $d-1$ directions. Such solutions do not decay at infinity. However, given initial data for such a solution, one could modify the data outside a ball of radius R so that it becomes compactly supported. In the subluminal case, the resulting solution would be unchanged in the region inside the ingoing Minkowski lightcone emanating from the surface of this ball. Hence if R is chosen large enough then the evolution of the solution inside the ball will behave as discussed above for long enough to see non-uniqueness of MGHDs.

In the higher-dimensional superluminal case, there is a unique MGHD: we will see below that *any* superluminal equation always admits a unique MGHD.

[15] Note that the extendibility across the null sections of the boundary implies that the analogue of the strong cosmic censorship conjecture is false for the superluminal equation. But the behaviour is much better than in the subluminal case for which the object one needs to define to formulate this conjecture (the MGHD) is not even unique!

5.4 Uniqueness properties of the initial value problem for quasilinear wave equations

5.4.1 Introduction

In this section we consider a quasilinear wave equation of the form

$$g^{\mu\nu}(u,du)\partial_\mu\partial_\nu u = F(u,du)\ , \tag{5.39}$$

where $u:\mathbb{R}^{d+1}\supseteq U\to\mathbb{R}$, g is a smooth Lorentz metric valued function,[16] F is smooth with $F(0,0)=0$, and the coordinates used for defining (5.39) are the canonical coordinates x^μ on $\mathbb{R}^{d+1}$.

Let $S\subseteq\mathbb{R}^{d+1}$ be a connected hypersurface of $\mathbb{R}^{d+1}$. *Initial data for* (5.39) *on* S consists of a smooth real valued function $f_0:S\to R$ and a smooth one form α_0 (with values in $T^*\mathbb{R}^{d+1}$) along S such that $X(f_0)=\alpha_0(X)$ holds for all vectors X tangent to S and such that the hypersurface S is spacelike with respect to the Lorentzian metric $g(f_0,\alpha_0)$. A *globally hyperbolic development* (GHD) of initial data (f_0,α_0) on a hypersurface S for (5.39) consists of a smooth solution $u:U\to\mathbb{R}$ of (5.39) ($U\subseteq\mathbb{R}^{d+1}$ being open) with $S\subseteq U$ and $u|_S=f_0$, $du|_S=\alpha_0$, and such that U is globally hyperbolic with respect to the Lorentzian metric $g(u,du)$ with Cauchy hypersurface S.

As we will show/recall in the following, the initial value problem for the equation (5.39) with initial data given on a hypersurface S is *locally well-posed*. Here, we mean by this that the following two properties hold:

1. there exists a globally hyperbolic development $u:U\to\mathbb{R}$ of the initial data
2. given two globally hyperbolic developments $u_1:U_1\to\mathbb{R}$ and $u_2:U_2\to\mathbb{R}$ of the same initial data, then there exists a *common globally hyperbolic development* (CGHD), that is, a globally hyperbolic development $v:V\to\mathbb{R}$ of the initial data with $V\subseteq U_1\cap U_2$ and $v=u_1|_V=u_2|_V$.

Note that the second property is only a weak version of what one might understand under 'local uniqueness', since it allows for the existence of a third globally hyperbolic development

[16] All the results presented in this section generalise literally unchanged to the setting of Section 5.2, where one does not assume that g in (5.39) is a Lorentz metric valued function, but one restricts consideration to *hyperbolic solutions*, i.e., solutions of (5.39) for which g *is* Lorentz metric valued. The only slight modification necessary is for the proof of the local existence result, Theorem 5.4.2. Here one can for example cut off the principal symbol of the quasilinear equation in the fashion of Remark **??** to create a quasilinear wave equation and then apply the local existence result for quasilinear wave equations to show local existence of hyperbolic solutions for quasilinear equations with hyperbolic initial data.

$u_3 : U_3 \to \mathbb{R}$ of the same initial data such that there exists an $x \in V \cap U_3$ with $u_3(x) \neq u_1(x) = u_2(x)$.

The aim of this section is to investigate the uniqueness properties for solutions of quasilinear wave equations. In Section 5.4.2 we provide theorems related to the second property of the local well-posedness statement from above and then give the main theorem of this section: two globally hyperbolic developments of the same initial data agree on the intersection of their domains *if this intersection is connected*. Section 5.4.3 then specialises to quasilinear wave equations (5.39) with the property that

$$\text{there exists a vector field } T \text{ on } \mathbb{R}^{d+1} \text{ such that } T \text{ is timelike with respect to } g^{\mu\nu}(u,du) \text{ for all } u, du. \tag{5.40}$$

In particular superluminal equations have this property. We show that for such equations the intersection of the domains of two globally hyperbolic developments of the same initial data is always connected – and we thus obtain that any two globally hyperbolic developments agree on the intersection of their domains.

The next section deals with existence questions: it gives the first property of the above local well-posedness statement, establishes the existence of a unique maximal globally hyperbolic development for quasilinear wave equations with the property (5.40), and considers subluminal equations and shows the existence of a maximal region on which solutions are unique and which is globally hyperbolic (i.e. a MUGHD).

The final section, Section 5.4.5, presents a uniqueness criterion for *general* quasilinear wave equations of a very different flavour. It states that if there exists a maximal globally hyperbolic development with the property that its domain of definition always lies to just one side of its boundary, then this maximal globally hyperbolic development is the unique one. In particular this implies uniqueness of the MGHD constructed in Ref. [170].

5.4.2 Uniqueness results for general quasilinear wave equations

Before we start talking about global uniqueness, we need to make sure that this is not entirely unreasonable by first establishing *local* uniqueness.

Proposition 5.4.1 (Local uniqueness)**.** *Let $u_1 : U_1 \to \mathbb{R}$ and $u_2 : U_2 \to \mathbb{R}$ be two globally hyperbolic developments for* (5.39) *of the same initial data prescribed on a hypersurface $S \subseteq \mathbb{R}^{d+1}$. Then there exists a common globally hyperbolic development $v : V \to \mathbb{R}$ (for some subset $V \subset U_1 \cap U_2$).*

Proof. See [163] □

One can now ask whether *global uniqueness* holds, which is the property that if $u_1 : U_1 \to \mathbb{R}$ and $u_2 : U_2 \to \mathbb{R}$ are two globally hyperbolic developments of the same initial data, then u_1 and u_2 agree on $U_1 \cap U_2$. Note that 'global' refers to the property that 'the two solutions agree in *all of* $U_1 \cap U_2$' – in contrast to the local result provided by Proposition 5.4.1, which only guarantees uniqueness in some smaller subset of $U_1 \cap U_2$.

An idea for a proof of global uniqueness was sketched in Section 1.4.1 of [85]. However, this sketch has the flaw that it tacitly assumes that given two globally hyperbolic developments $u_1 : U_1 \to \mathbb{R}$ and $u_2 : U_2 \to \mathbb{R}$ of the same initial data, that $U_1 \cap U_2$ is then connected – which is in general not true, as illustrated by the example presented in Section 5.3.7 of this chapter.

As mentioned in section 1.4, for the Einstein equations one does not need to condition the global uniqueness statement, since one has the freedom to construct the underlying manifold – there is no fixed background. We will explain this in the following: given two globally hyperbolic developments u_1 and u_2 for the Einstein equations one constructs a bigger one in which both are contained (and thus proves global uniqueness) by glueing u_1 and u_2 together along the MCGHD of u_1 and u_2. However, in the case that u_1 and u_2 are two globally hyperbolic developments of a quasilinear wave equation on a fixed background such that $U_1 \cap U_2$ is disconnected, glueing them together along the MCGHD (which equals the connected component of $U_1 \cap U_2$ which contains the initial data hypersurface), would yield a solution which is no longer defined on a subset of $\mathbb{R}^{d+1}$, but instead on a manifold which projects down on $U_1 \cup U_2 \subseteq \mathbb{R}^{d+1}$ and contains the other connected components of $U_1 \cap U_2$ twice. Of course this is not allowed if we insist that solutions of (5.39) should be defined on a subset of $\mathbb{R}^{d+1}$. So the key difference between the Einstein equations and a quasilinear wave equation (5.39) is that for the former the underlying manifold is constructed along with the solution whereas for the latter, it is fixed *a priori*. This is the reason why one does not need to condition the global uniqueness statement for the Einstein equations.

Theorem 5.4.1. *Let $u_1 : U_1 \to \mathbb{R}$ and $u_2 : U_2 \to \mathbb{R}$ be two globally hyperbolic developments of* (5.39) *arising from the same initial data given on a connected hypersurface $S \subseteq \mathbb{R}^{d+1}$. Assume that $U_1 \cap U_2$ is connected. Then u_1 and u_2 agree on $U_1 \cap U_2$.*

Proof. See [163]. □

The following is an immediate consequence of the previous theorem. It shows that global uniqueness can only be violated for quasilinear wave equations in a specific way.

Corollary 5.4.1. *Let $u_1 : U_1 \to \mathbb{R}$ and $u_2 : U_2 \to \mathbb{R}$ be two globally hyperbolic developments of* (5.39) *arising from the same initial data on a connected hypersurface $S \subseteq \mathbb{R}^{d+1}$. If there exists an $x \in U_1 \cap U_2$ with $u_1(x) \neq u_2(x)$, then $U_1 \cap U_2$ is not connected.*

In particular, we recover that globally defined solutions are unique:

Corollary 5.4.2. *Let $u_1 : \mathbb{R}^{d+1} \to \mathbb{R}$ be a globally defined globally hyperbolic development of* (5.39) *arising from some initial data on a connected hypersurface $S \subseteq \mathbb{R}^{d+1}$. Let $u_2 : U_2 \to \mathbb{R}$ be another globally hyperbolic development of* (5.39) *of the same initial data. Then $u_1|_{U_2} = u_2$.*

At this point it is useful to refer back to the example in section 5.3.7.1 of non-uniqueness for the subluminal Born-Infeld scalar (a quasilinear wave equation of the form (5.39)). Recall that we constructed two globally hyperbolic developments, U_a and U_b, that disagree in some subset $X \subset \mathbb{R}^2$, *i.e.* $\Phi_a(x) \neq \Phi_b(x)$ for $x \in X$. From Corollary 5.4.1 we know that for this to be possible $U_a \cap U_b$ must be disconnected. This is most obvious from Fig. 5.8, which shows U_a and U_b: $U_a \cap U_b$ consists of two disconnected regions, one lying outside of X and including the initial surface S, the other lying inside X. Note that these regions are disconnected because points on ∂X are in *either* U_a or U_b but not both (Φ_a blows up on the red curve in Fig. 5.8, so it is not in U_a; similarly for the blue dot-dashed curve and U_b). As expected, the solutions agree in the connected component of $U_a \cap U_b$ that includes the initial surface, but differ in the other component.

In the next section we consider two globally hyperbolic developments $u_1 : U_1 \to \mathbb{R}$ and $u_2 : U_2 \to \mathbb{R}$ of the same initial data and discuss a criterion that ensure that $U_1 \cap U_2$ is connected. Here, the choice of the initial data hypersurface S plays an important role. This can already be seen from the special case of the linear wave equation in Minkowski space: consider a spacelike *but not achronal* hypersurface that winds up around the x^0-axis in $\mathbb{R} \times (B_2^d(0) \setminus B_1^d(0)) \subseteq \mathbb{R}^{d+1}$. Prescribing generic initial data on this hypersurface, the extent of the future development restricts the extent of the past development. Given two globally hyperbolic developments, their intersection is in general not connected and global uniqueness does not hold. However, it is easy to show (see also Section 5.4.3) that for spacelike initial data hypersurfaces which are moreover achronal, this pathology for the linear wave equation in Minkowski space cannot occur. This example shows that any result demonstrating connectedness of $U_1 \cap U_2$ for more general quasilinear equations will require some additional assumptions on the initial surface S analogous to the achronality assumption in Minkowski spacetime.

5.4.3 Uniqueness results for superluminal quasilinear wave equations

In the following we consider quasilinear wave equations (5.39) that enjoy property (5.40), i.e., that there exists a vector field T on $\mathbb{R}^{d+1}$ such that T is timelike with respect to $g^{\mu\nu}(u, du)$

for all u, du. In particular, superluminal equations enjoy this property, since one can take $T = \partial/\partial x^0$ where x^μ are inertial frame coordinates. We will show that for such equations the complication of $U_1 \cap U_2$ being disconnected cannot arise, as long as the initial data is prescribed on a hypersurface S with the property that every maximal integral curve of T intersects S at most once.[17]

Lemma 5.4.1. *Assume that there exists a vector field T on $\mathbb{R}^{d+1}$ such that T is timelike with respect to $g^{\mu\nu}(u, du)$ for all u, du, where g is as in* (5.39)*. Let $u_1 : U_1 \to \mathbb{R}$ and $u_2 : U_2 \to \mathbb{R}$ be two globally hyperbolic developments of* (5.39) *arising from the same initial data on a connected hypersurface S which has the property that every maximal integral curve of T intersects S at most once. Then $U_1 \cap U_2$ is connected.*

Proof. See [163]. □

Let us remark, that one can replace in the above lemma the assumption that S is a connected hypersurface such that every maximal integral curve of T intersects S at most once, with the assumption that S is a hypersurface that separates $\mathbb{R}^{d+1}$ into two components.

Corollary 5.4.3. *Assume that there exists a vector field T on $\mathbb{R}^{d+1}$ such that T is timelike with respect to $g^{\mu\nu}(u, du)$ for all u, du, where g is as in* (5.39) *and that initial data is posed on a connected hypersurface S which has the property that every maximal integral curve of T intersects S at most once.*

Given two globally hyperbolic developments $u_1 : U_1 \to \mathbb{R}$ and $u_2 : U_2 \to \mathbb{R}$, we then have $u_1(x) = u_2(x)$ for all $x \in U_1 \cap U_2$.

Proof. This follows directly from Lemma 5.4.1 and Theorem 5.4.1. □

Since this applies to superluminal quasilinear wave equations, we can relate it back to the example of the superluminal Born-Infeld scalar in section 5.3.8. In this example, we found that it was only possible to construct a single maximal globally hyperbolic development; all possible globally hyperbolic developments are subsets of this one, so the solutions must be the same in their intersections. The reason for this was given in Lemma 5.3.6; any other solution obtained using the mapping described in section 5.3.6 cannot be a globally hyperbolic development.

[17] For a superluminal equation with $T = \partial/\partial x^0$, any S which is a Cauchy surface for Minkowski spacetime has this property. Of course S also has to obey the assumptions discussed at the beginning of section 5.4.1 e.g. S has to be spacelike w.r.t. $g(u, du)$.

5.4.4 Existence results for general quasilinear wave equations

For our discussion of predictability the above theorems about uniqueness are the most important, but it is also useful to establish that the objects we have been studying actually exist, which we do in this section. We give three relevant results: the first is local existence, then existence of a unique maximal globally hyperbolic development for superluminal equations, and finally existence of a maximal uniqe globally hyperbolic development for subluminal equations.

First of all we provide the other half of the local well-posedness statement for quasilinear wave equations with data on *general* hypersurfaces: the local existence result.

Theorem 5.4.2 (Local existence)**.** *Given initial data for a quasilinear wave equation* (5.39)*, there exists a globally hyperbolic development.*

Moreover, this result is needed for the two existence results that follow.

The next theorem establishes that a unique maximal GHD *exists* for superluminal quasilinear wave equations.

Theorem 5.4.3. *Assume that there exists a vector field T on $\mathbb{R}^{d+1}$ such that T is timelike with respect to $g^{\mu\nu}(u,du)$ for all u,du, where g is as in* (5.39) *and that initial data is posed on a connected hypersurface S which has the property that every maximal integral curve of T intersects S at most once.*

Given such initial data, there then exists a unique maximal globally hyperbolic development *$u_{\max} : U_{\max} \to \mathbb{R}$, that is, a globally hyperbolic development $u_{\max} : U_{\max} \to \mathbb{R}$ with the property that for any other globally hyperbolic development $u : U \to \mathbb{R}$ of the same initial data we have $U \subseteq U_{\max}$ and $u_{\max}|_U = u$.*

Proof. See [163]. □

Remark 5.4.1. *We note that the construction of a unique maximal globally hyperbolic development is always possible provided the property of global uniqueness holds.*

Finally we consider subluminal quasilinear wave equations. As mentioned before, for such equations there does not generally exist a unique maximal globally hyperbolic development. The following theorem shows existence of a globally hyperbolic development on the domain of which the solution is uniquely defined and which is maximal among all GHDs that have this property.

First we establish some terminology: we consider a subluminal quasilinear wave equation of the form (5.39) and consider initial data prescribed on a connected hypersurface S that is *acausal* with respect to the Minkowski metric m, i.e., there does not exist a pair of points

on S that can be connected by a causal curve within the Minkowski spacetime. We call a GHD $u_1 : U_1 \to \mathbb{R}$ a *unique globally hyperbolic development* (UGHD) iff for all other GHDs $u_2 : U_2 \to \mathbb{R}$ we have $u_1 = u_2$ on $U_1 \cap U_2$.

Theorem 5.4.4. *Consider a subluminal quasilinear wave equation of the form* (5.39)*. Given initial data on a connected hypersurface S that is acausal with respect to the Minkowski metric there exists a UGHD $u : U \to \mathbb{R}$ with the property that the domain of any other UGHD is contained in U. The UGHD $u : U \to \mathbb{R}$ is called the* maximal unique globally hyperbolic development *(MUGHD).*

Proof. See [163]. □

It is again useful to refer back to the example in section 5.3.7.1. In that case, the maximal unique GHD is given by the subset $U \subset \mathbb{R}^2$ with solution Φ. It can be extended as a GHD but not in a unique way: we constructed two possible extensions, U_a and U_b, but there are many more.

We summarise that given a GHD for a *superluminal* equation, one knows that it is contained in the unique maximal GHD. For *subluminal* equations, there are in general GHDs which are not contained in the maximal UGHD.

Let us also remark that we expect that the analogue of Theorem 5.4.4 does not hold for more general quasilinear wave equations, i.e., ones which are neither subluminal nor superluminal. Indeed, even more strongly, we formulate the following

Conjecture 5.4.1. *There are quasilinear wave equations of the form* (5.39) *for which there exists initial data such that there does not exist* any *UGHD.*

This conjecture is based on the following scenario which we think might happen: there exists a quasilinear wave equation of the form (5.39) and initial data such that there exists an infinite family of GHDs the domains of which bend round back towards the initial data hypersurface S and approach it arbitrarily closely, as shown in Figure 5.11. This would imply that there is no neighbourhood of S on which the solution is uniquely defined. In particular, this would establish the sharpness of the local uniqueness statement of Proposition 5.4.1.

5.4.5 A uniqueness criterion for general quasilinear wave equations at the level of MGHDs

If we have succeeded in finding a maximal globally hyperbolic development for a general quasilinear equation it is useful to know whether or not it is unique; if it turns out that it is not unique we then know it doesn't make sense to consider it a physical object.

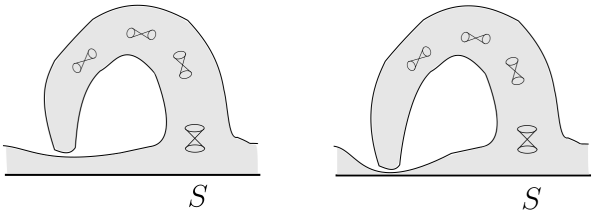


Fig. 5.11 A possible mechanism for a resolution of Conjecture 5.4.1. The figure shows the light cones of $g(u,du)$.

In this section we consider a general quasilinear wave equation of the form (5.39). Recall that a GHD $u_1 : U_1 \to \mathbb{R}$ of given initial data posed on a hypersurface S is called a *maximal globally hyperbolic development* (MGHD) iff there does not exist a GHD $u_2 : U_2 \to \mathbb{R}$ of the same initial data with $U_1 \subsetneq U_2$. Note that by Theorem 5.4.1 any such GHD $u_2 : U_2 \to \mathbb{R}$ would agree with u_1 on U_1, and thus it would correspond to an extension of $u_1 : U_1 \to \mathbb{R}$. In other words, a MGHD is a GHD that cannot be extended as a GHD.

The example from Section 5.3.7 shows that in general there can exist infinitely many MGHDs for given initial data. Consider now two such MGHDs $u_1 : U_1 \to \mathbb{R}$ and $u_2 : U_2 \to \mathbb{R}$ arising in the example of Section 5.3.7. Then $U_1 \cap U_2$ is disconnected. Let A denote the connected component containing S. Consider a point $x \in U_1 \cap U_2$ which does not lie in A. *The phenomenon of non-uniqueness, i.e., that $u_1(x)$ does not equal $u_2(x)$, arises, because the 'path of evolution' the second solution takes from A to reach x is blocked because the first solution is already defined in that very region.* In the example of Section 5.3.7, this behaviour arises because U_1 (say) lies "on both sides of its boundary". The following theorem makes this precise and shows that this is the only mechanism at the level of MGHDs that leads to non-uniqueness for general quasilinear wave equations. It states that given an MGHD with the property that its domain of definition always lies to *just one side* of its boundary, i.e., the domain of definition cannot block evolution elsewhere, then it is the unique MGHD.

Theorem 5.4.5. *Let $u_1 : U_1 \to \mathbb{R}$ be a MGHD of given initial data for a quasilinear wave equation of the form* (5.39) *and assume that*

$$\begin{aligned} &\textit{for every } p \in (\partial U_1 \setminus \partial S) \textit{ there exists a neighbourhood } V \textit{ of } p \textit{ together} \\ &\textit{with a chart } \psi : V \to (-\varepsilon,\varepsilon)^{d+1},\ \varepsilon > 0, \textit{ and a continuous function} \\ &f : (-\varepsilon,\varepsilon)^d \to (-\varepsilon,\varepsilon) \textit{ such that } \psi^{-1}(\mathrm{graph} f) = \partial U_1 \cap V, \textit{ all points} \\ &\textit{below } \mathrm{graph} f \textit{ in } (-\varepsilon,\varepsilon)^{d+1} \textit{ are mapped into } U_1 \textit{ and all points above} \\ &\mathrm{graph} f \textit{ in } (-\varepsilon,\varepsilon)^{d+1} \textit{ are mapped into } \mathbb{R}^{d+1} \setminus U_1. \end{aligned} \tag{5.41}$$

Then $u_1 : U_1 \to \mathbb{R}$ is the unique MGHD, i.e., any other GHD $u : U \to \mathbb{R}$ satisfies $U \subseteq U_1$ and thus also $u_1|_U = u$.

Note that in order to apply this theorem to a concrete example one has to first construct a/the whole MGHD and is only then able to infer *a posteriori* that the evolution was indeed unique.

Proof. See [163]. □

This is demonstrated in our example in section 5.3.7.1. For the subluminal Born-Infeld scalar, we can see from Fig. 5.7 that U_a lies on both sides of the section of its boundary that is the red curve (a subset of ∂X). Theorem 5.4.5 tells us that it is then possible that U_a is not the unique MGHD, which is indeed the case, as can be seen by the construction of the different MGHD U_b.

On the other hand, in the example of the superluminal Born-Infeld scalar in section 5.3.8, the MGHD U that we constructed there only lies on one side of its boundary, which can be seen from Fig. 5.10. Thus from Theorem 5.4.5 we know that U is the unique MGHD, in agreement with our example.

We conclude with presenting a simple criterion that ensures that condition (5.41) is satisfied. It is tailored to small data results.

Lemma 5.4.2. *Let $u_1 : U_1 \to \mathbb{R}$ be a GHD of given initial data posed on an open and connected subset S of $\{x^0 = 0\}$ for a quasilinear wave equation of the form* (5.39)*. Furthermore, assume that*

$$\begin{aligned} &\text{there exists a } \delta > 0 \text{ such that } \partial_0 + \textstyle\sum_{i=1}^d \delta_i \partial_i \text{ is timelike with respect} \\ &\text{to } g(u_1, du_1) \text{ for all } \delta_i \in \mathbb{R} \text{ with } \textstyle\sum_{i=1}^d |\delta_i| < \delta. \end{aligned} \tag{5.42}$$

Then the condition (5.41) *is satisfied.*

As an application of the lemma and of Theorem 5.4.5 let us mention the work [170] of Christodoulou in which he studies the formation of shocks for relativistic perfect fluids. In the irrotational case the equations of motion give rise to a subluminal wave equation. For sufficienly small initial data he explicitly constructs a MGHD and Theorem 13.1, conclusion iii) in [170] shows that the assumptions of the above lemma are met.

Chapter 6

Conclusion

In this thesis we have studied various applications and properties of wave equations on curved spacetimes. One reason they are of particular importance in GR is that questions such as whether or not a spacetime is stable, or whether the strong cosmic censorship conjecture is true, are very difficult to solve but can be significantly simplified by looking at linearised formulations. The behaviour of these linear perturbations can be determined by studying the scalar wave equation. Although we have only described two of its uses in GR, there are many more. For example, some intuition and indicative results as to whether or not it is possible to have a naked singularity violating the weak cosmic censorship conjecture [176] can be achieved by looking at linear perturbations that satisfy the wave equation (although again, this is a linear version of the much more complicated non-linear problem); another use is in the study of Hawking radiation (see e.g. [10]).

Despite the fact that the wave equation is often used as a simplification of more complicated problems, it is by no means easy to solve. In chapters 2 and 4 we have made significant use of the relation between null geodesics and solutions of the wave equation to find quasinormal mode solutions. This demonstrates the importance of null geodesics in determining properties of spacetimes (such as the stability, quasinormal mode spectrum etc.); whether or not these geodesics are stably or unstably trapped is of particular significance for this.

To study solutions of the wave equation, we found quasinormal modes in the eikonal limit using the geometric optics approximation in chapter 4 and a modified type of matched asymptotic expansion in chapter 2, but there are many other ways to determine them. For example, they can be found numerically (as in chapters 2 and 4), by taking different limits such as the near-extremal limit, or by using either the WKB approximation [177] or other expansion methods in the eikonal limit (see e.g. [178] for a review).

We have been interested mostly in quasinormal modes for two different reasons: for the microstate geometries we expected the quasinormal modes to be related to the stably trapped null geodesics; in Kerr-de Sitter the late-time behaviour of solutions depends entirely on the quasinormal frequencies. Other types of solutions, for example solutions that have finite energy on hypersurfaces of constant t, can be studied using different methods than those mentioned above. These include using energy estimates (see [105] for an example of this for quasimode solutions in the microstate geometries) or microlocal analysis (see [179] for an example of this in Kerr-de Sitter, which can also be used to find the quasinormal frequencies) among many others.

Both the heuristic argument and the study of the wave equation in chapter 2 lead to the expectation that the supersymmetric microstate geometries are unstable. This is due to the presecence of an evanescent ergosurface on which there are stably trapped null geodesics with zero energy. These give rise to extremely slowly decaying solutions of the wave equation which we expect to lead to an instability in the non-linear problem.

Since geodesics are so important for this analysis we study them in detail in chapter 3. We find that there are more trapped geodesics than the ones on the evanescent ergosurface used in chapter 2, but that these do not have zero energy. In addition, we found that it is possible to set up the scenario in the heuristic argument using a massive particle at infinity that follows a geodesic inwards towards the evanescent ergosurface. Despite the lack of a more usual ergoregion (in 6d) it is also still possible to construct a Penrose process in the microstate geometries, although when we compared the geodesics to those around a black hole we found that there were significant differences.

Geodesics are also of the utmost importance in chapter 4 where we study linear perturbations in Kerr-de Sitter. The quasinormal modes that are related to the null geodesics on the photon sphere decay slowly enough that we expect strong cosmic censorship to be respected, in contrast to some Reissner-Nordström-de Sitter black holes. The geometric optics approximation was crucial in the analytic approximation to the quasinormal frequency, and was verified numerically. The results for the linearised gravitational perturbations back up the findings for the scalar wave equation.

It would be interesting to see if these results for the scalar wave equation in Kerr-de Sitter could be made rigorous. I have been attempting to do this more recently by first of all studying the wave equation in the exterior to rigorously find the quasinormal modes (as poles of the resolvent) in the eikonal limit. The effective potentials in the o.d.e.s arising from separation of variables have local extrema at leading order, and it may be possible to construct the quasinormal modes using this property and results from semiclassical analysis (see *e.g.* [180, 181] for general results and [37] where this has been done for slowly rotating Kerr-de

Sitter). Unfortunately this is where the main difficulty arises, since the effective potentials depend on the quasinormal frequencies and are thus complex (albeit at next to leading order), and so such results cannot be applied straight away. If it is possible to construct these modes in the exterior then the WKB approximation can be used in the interior, with the error calculated and proven to be small, to find the reflection and transmission coefficients and show that these quasinormal mode solutions are not in H^1_{loc} at the Cauchy horizon, as expected from the work in chapter 4.

Both the stability problem in GR and the strong cosmic censorship conjecture are related to predictability; a classical physical theory should be able to predict everything starting from suitable initial data. In chapter 5 we study predictability in a slightly different sense to the first chapters. It is generally thought that superluminal wave equations are much worse in terms of predictability than subluminal equations, perhaps allowing for the construction of time machines. Although we do not answer the question of whether it is possible to build a time machine, we find that the superluminal form of the Born-Infeld scalar in two dimensions is in some sense more predictable than the subluminal case. More precisely, the maximal globally hyperbolic development arising from given initial data is unique in the superluminal case but is *not* unique in the subluminal case; this can be seen in the example we constructed. This also applies more generally, and could therefore lead to problems with some commonly used subluminal wave equations.

Overall, we have seen the importance of wave equations on curved spacetimes in providing some insight into problems relating to predictability; however, solving the full non-linear versions of these problems involves considerable difficulties and they are far from being resolved.

References

[1] A. Einstein, "Die Feldgleichungen der Gravitation", *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), Seite 844-847.*, 1915.

[2] C. M. Will, "The Confrontation between General Relativity and Experiment", *Living Rev. Rel.* **17** (2014) 4, arXiv:1403.7377.

[3] U. J. Le Verrier, "Theorie du mouvement de Mercure", *Annales de l'Observatoire de Paris* **5** (1859).

[4] E. A. S. Dyson Frank Watson and D. C., "A determination of the deflection of light by the sun's gravitational field, from observations made at the total eclipse of May 29, 1919", *Philosophical Transactions of the Royal Society of London. Series A* **220** (1920).

[5] D. M. Popper, "Red Shift in the Spectrum of 40 Eridani B.", **120** (1954) 316.

[6] K. Schwarzschild, "Über das Gravitationsfeld eines Massenpunktes nach der Einsteinschen Theorie", *Sitzungsberichte der Königlich Preußischen Akademie der Wissenschaften (Berlin), 1916, Seite 189-196*, 1916.

[7] S. Gillessen, F. Eisenhauer, S. Trippe, T. Alexander, R. Genzel, F. Martins, and T. Ott, "Monitoring stellar orbits around the Massive Black Hole in the Galactic Center", *Astrophys. J.* **692** (2009) 1075–1109, arXiv:0810.4674.

[8] A. Celotti, J. C. Miller, and D. W. Sciama, "Astrophysical evidence for the existence of black holes: Topical review", *Class. Quant. Grav.* **16** (1999) A3, arXiv:astro-ph/9912186.

[9] **LIGO Scientific Collaboration and Virgo Collaboration** Collaboration, "Observation of gravitational waves from a binary black hole merger", *Phys. Rev. Lett.* **116** Feb (2016) 061102.

[10] R. M. Wald, "General relativity", Chicago Univ. Press, Chicago, IL, 1984.

[11] J. Sbierski, "The C_0-inextendibility of the Schwarzschild spacetime and the spacelike diameter in Lorentzian geometry", *J. Diff. Geom.* **108** (2018), no. 2, 319–378, arXiv:1507.00601.

[12] J. Sbierski, "On the proof of the C^0-inextendibility of the Schwarzschild spacetime", *J. Phys. Conf. Ser.* **968** (2018), no. 1, 012012, arXiv:1711.11380.

[13] M. Dafermos and J. Luk, "The interior of dynamical vacuum black holes I: The C^0-stability of the Kerr Cauchy horizon", arXiv:1710.01722.

[14] A. D. Ionescu and S. Klainerman, "On the uniqueness of smooth, stationary black holes in vacuum", *Invent. Math.* **175** (2008) 35, arXiv:0711.0040.

[15] D. Christodoulou and S. Klainerman, "The global nonlinear stability of the minkowski space", Princeton University Press, 1993.

[16] B. F. Whiting, "Mode stability of the Kerr black hole", *Journal of Mathematical Physics* **30** (1989), no. 6, 1301–1305, https://doi.org/10.1063/1.528308.

[17] M. Dafermos and I. Rodnianski, "Decay for solutions of the wave equation on Kerr exterior spacetimes I-II: The cases |a| « M or axisymmetry", arXiv:1010.5132.

[18] M. Dafermos, I. Rodnianski, and Y. Shlapentokh-Rothman, "Decay for solutions of the wave equation on Kerr exterior spacetimes III: The full subextremal case |a| < M", *Ann. of Math.*, 2016 arXiv:1402.7034.

[19] M. Dafermos, I. Rodnianski, and Y. Shlapentokh-Rothman, "A scattering theory for the wave equation on Kerr black hole exteriors", arXiv:1412.8379.

[20] M. Dafermos, G. Holzegel, and I. Rodnianski, "The linear stability of the Schwarzschild solution to gravitational perturbations", arXiv:1601.06467.

[21] M. Dafermos, G. Holzegel, and I. Rodnianski, "Boundedness and decay for the Teukolsky equation on Kerr spacetimes I: the case $|a| \ll M$", arXiv:1711.07944.

[22] M. Dafermos, 2006, unpublished.

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

[24] G. Moschidis, "A proof of the instability of AdS for the Einstein–massless Vlasov system", arXiv:1812.04268.

[25] F. John, "Blow-up for quasi-linear wave equations in three space dimensions", *Communications on Pure and Applied Mathematics* **34** (1981), no. 1, 29–51, https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.3160340103.

[26] S. Klainerman, "The null condition and global existence to nonlinear wave equations", *Lectures in Appl. Math.* **23** (1986) 293–326.

[27] H. Lindblad and I. Rodnianski, "The Global stability of the Minkowski space-time in harmonic gauge", arXiv:math/0411109.

[28] H. Lindblad, "Global solutions of nonlinear wave equations", *Communications on Pure and Applied Mathematics* **45** (1992), no. 9, 1063–1096, https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.3160450902.

[29] J. Keir, "The weak null condition and global existence using the p-weighted energy method", arXiv:1808.09982.

[30] J. V. Ralston, "Solutions of the wave equation with localized energy", *Communications on Pure and Applied Mathematics* **22** (1969), no. 6, 807–823, https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpa.3160220605.

[31] J. Sbierski, "Characterisation of the Energy of Gaussian Beams on Lorentzian Manifolds - with Applications to Black Hole Spacetimes", *Anal. Part. Diff. Eq.* **8** (2015) 1379–1420, arXiv:1311.2477.

[32] G. Holzegel and J. Smulevici, "Quasimodes and a Lower Bound on the Uniform Energy Decay Rate for Kerr-AdS Spacetimes", arXiv:1303.5944.

[33] J. Keir, "Slowly decaying waves on spherically symmetric spacetimes and ultracompact neutron stars", *Class. Quant. Grav.* **33** (2016), no. 13, 135009, arXiv:1404.7036.

[34] G. Benomio, "The Stable Trapping Phenomenon for Black Strings and Black Rings and its Obstructions on the Decay of Linear Waves", arXiv:1809.07795.

[35] A. S. Barreto and M. Zworski, "Distribution of resonances for spherical black holes", *Math. Res. Lett*, 1997.

[36] J.-F. Bony and D. Häfner, "Decay and Non-Decay of the Local Energy for the Wave Equation on the De Sitter–Schwarzschild Metric", *Communications in Mathematical Physics* **282** Sep (2008) 697–719.

[37] S. Dyatlov, "Asymptotic distribution of quasi-normal modes for Kerr-de Sitter black holes", *Annales Henri Poincare* **13** (2012) 1101–1166, arXiv:1101.1260.

[38] S. Dyatlov, "Asymptotics of linear waves and resonances with applications to black holes", *Commun. Math. Phys.* **335** (2015), no. 3, 1445–1485, arXiv:1305.1723.

[39] P. Hintz and A. Vasy, "The global non-linear stability of the Kerr-de Sitter family of black holes", arXiv:1606.04014.

[40] P. Hintz, "Non-linear stability of the Kerr-Newman-de Sitter family of charged black holes", arXiv:1612.04489.

[41] Y. B. Zel'Dovich, "Amplification of Cylindrical Electromagnetic Waves Reflected from a Rotating Body", *Soviet Journal of Experimental and Theoretical Physics* **35** (1972) 1085.

[42] Y. B. Zel'Dovich, "Generation of Waves by a Rotating Body", *Soviet Journal of Experimental and Theoretical Physics Letters* **14** (1971) 180.

[43] W. H. Press and S. A. Teukolsky, "Floating Orbits, Superradiant Scattering and the Black-hole Bomb", **238** (1972) 211–212.

[44] A. A. Starobinskiĭ, "Amplification of waves during reflection from a rotating "black hole"", *Soviet Journal of Experimental and Theoretical Physics* **37** (1973) 28.

[45] R. Penrose and R. M. Floyd, "Extraction of Rotational Energy from a Black Hole", *Nature* **229** (1971) 177–179.

[46] Y. Choquet-Bruhat, P. T. Chrusciel, and J. Loizelet, "Global solutions of the Einstein-Maxwell equations in higher dimensions", *Class. Quant. Grav.* **23** (2006) 7383–7394, arXiv:gr-qc/0608108.

[47] C. M. Warnick, "On quasinormal modes of asymptotically anti-de Sitter black holes", *Commun. Math. Phys.* **333** (2015), no. 2, 959–1035, arXiv:1306.5760.

[48] V. Ferrari and B. Mashhoon, "New approach to the quasinormal modes of a black hole", *PRD* **30** (1984) 295–304.

[49] K. D. Kokkotas and B. G. Schmidt, "Quasinormal modes of stars and black holes", *Living Rev. Rel.* **2** (1999) 2, arXiv:gr-qc/9909058.

[50] S. Chandrasekhar, "The mathematical theory of black holes", Oxford Univ. Press, Oxford, 2002.

[51] C. W. Misner, K. S. Thorne, and J. A. Wheeler, "Gravitation", W. H. Freeman, San Francisco, first edition ed., 1973.

[52] H. Yang, D. A. Nichols, F. Zhang, A. Zimmerman, Z. Zhang, and Y. Chen, "Quasinormal-mode spectrum of Kerr black holes and its geometric interpretation", *Phys. Rev.* **D86** (2012) 104006, arXiv:1207.4253.

[53] C. J. Goebel, "Comments on the "vibrations" of a Black Hole.", *Astrophysical Journal* **172** (1972) L95.

[54] V. Ferrari and B. Mashhoon, "Oscillations of a black hole", *Phys. Rev. Lett.* **52** Apr (1984) 1361–1364.

[55] B. Mashhoon, "Stability of charged rotating black holes in the eikonal approximation", *Phys. Rev.* **D31** (1985), no. 2, 290–293.

[56] L. Bombelli and E. Calzetta, "Chaos around a black hole", *Classical and Quantum Gravity* **9** dec (1992) 2573–2599.

[57] N. J. Cornish and J. J. Levin, "Lyapunov timescales and black hole binaries", *Class. Quant. Grav.* **20** (2003) 1649–1660, arXiv:gr-qc/0304056.

[58] V. Cardoso, A. S. Miranda, E. Berti, H. Witek, and V. T. Zanchin, "Geodesic stability, Lyapunov exponents and quasinormal modes", *Phys. Rev.* **D79** (2009) 064016, arXiv:0812.1806.

[59] S. R. Dolan, "The Quasinormal Mode Spectrum of a Kerr Black Hole in the Eikonal Limit", *Phys. Rev.* **D82** (2010) 104003, arXiv:1007.5097.

[60] F. C. Eperon, H. S. Reall, and J. E. Santos, "Instability of supersymmetric microstate geometries", *JHEP* **10** (2016) 031, arXiv:1607.06828.

[61] G. Moschidis, "Logarithmic local energy decay for scalar waves on a general class of asymptotically flat spacetimes", arXiv:1509.08495.

[62] Y. Choquet-Bruhat and R. Geroch, "Global aspects of the cauchy problem in general relativity", *Communications in Mathematical Physics* **14** Dec (1969) 329–335.

[63] R. Penrose, "Structure of space-time", *Battelle Rencontres, 1967 Lectures in Mathematics and Physics*, 1968 222.

[64] R. Penrose, "Singularities of spacetime", 1978.

[65] M. Simpson and R. Penrose, "Internal instability in a reissner-nordström black hole", *International Journal of Theoretical Physics* **7** Apr (1973) 183–197.

[66] J. M. McNamara, "Instability of black hole inner horizons", *Proceedings of the Royal Society of London Series A* **358** (1978) 499–517.

[67] S. Chandrasekhar and J. B. Hartle, "On crossing the cauchy horizon of a reissner-nordstrom black-hole", *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **384** (1982), no. 1787, 301–315.

[68] E. Poisson and W. Israel, "Internal structure of black holes", *Phys. Rev. D* **41** Mar (1990) 1796–1809.

[69] D. Christodoulou, "On the global initial value problem and the issue of singularities", *Classical and Quantum Gravity* **16** nov (1999) A23–A35.

[70] M. Dafermos, "Black holes without spacelike singularities", *Commun. Math. Phys.* **332** (2014) 729–757, `arXiv:1201.1797`.

[71] J. M. McNamara, "Behaviour of scalar perturbations of a reissner-nordström black hole inside the event horizon", *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences* **364** (1978), no. 1716, 121–134.

[72] A. Ori, "Inner structure of a charged black hole: An exact mass-inflation solution", *Phys. Rev. Lett.* **67** (1991) 789–792.

[73] M. Dafermos, "The Interior of charged black holes and the problem of uniqueness in general relativity", *Commun. Pure Appl. Math.* **58** (2005) 0445–0504, `arXiv:gr-qc/0307013`.

[74] A. T. Franzen, "Boundedness of Massless Scalar Waves on Reissner-Nordström Interior Backgrounds", *Commun. Math. Phys.* **343** (2016), no. 2, 601–650, `arXiv:1407.7093`.

[75] J. Luk and S.-J. Oh, "Strong cosmic censorship in spherical symmetry for two-ended asymptotically flat initial data I. The interior of the black hole region", `arXiv:1702.05715`.

[76] D. Christodoulou, "The Formation of Black Holes in General Relativity", in "On recent developments in theoretical and experimental general relativity, astrophysics and relativistic field theories. Proceedings, 12th Marcel Grossmann Meeting on General Relativity, Paris, France, July 12-18, 2009. Vol. 1-3", pp. 24–34. 2008. `arXiv:0805.3880`.

[77] D. Civin, "Stability of charged rotating black holes for linear scalar perturbations", 2015.

[78] P. Hintz, "Boundedness and decay of scalar waves at the Cauchy horizon of the Kerr spacetime", arXiv:1512.08003.

[79] M. Dafermos and Y. Shlapentokh-Rothman, "Time-Translation Invariance of Scattering Maps and Blue-Shift Instabilities on Kerr Black Hole Spacetimes", *Commun. Math. Phys.* **350** (2017), no. 3, 985–1016, arXiv:1512.08260.

[80] J. Luk and J. Sbierski, "Instability results for the wave equation in the interior of Kerr black holes", arXiv:1512.08259.

[81] M. Dafermos and I. Rodnianski, "The Red-shift effect and radiation decay on black hole spacetimes", *Commun. Pure Appl. Math.* **62** (2009) 859–919, arXiv:gr-qc/0512119.

[82] F. Mellor and I. Moss, "Stability of black holes in de sitter space", *Phys. Rev. D* **41** Jan (1990) 403–409.

[83] P. R. Brady, I. G. Moss, and R. C. Myers, "Cosmic censorship: As strong as ever", *Phys. Rev. Lett.* **80** (1998) 3432–3435, arXiv:gr-qc/9801032.

[84] Y. Choquet-Bruhat, "Theoreme d'existence pour certains systemes derivees partielles non lineaires", *Acta Mat.* **88** (1952) 141–225.

[85] J. Sbierski, "On the Existence of a Maximal Cauchy Development for the Einstein Equations - a Dezornification", *Annales Henri Poincare* **17** (2016), no. 2, 301–329, arXiv:1309.7591.

[86] O. Lunin and S. D. Mathur, "AdS / CFT duality and the black hole information paradox", *Nucl. Phys.* **B623** (2002) 342–394, arXiv:hep-th/0109154.

[87] J. M. Maldacena and L. Maoz, "Desingularization by rotation", *JHEP* **12** (2002) 055, arXiv:hep-th/0012025.

[88] V. Balasubramanian, J. de Boer, E. Keski-Vakkuri, and S. F. Ross, "Supersymmetric conical defects: Towards a string theoretic description of black hole formation", *Phys. Rev.* **D64** (2001) 064011, arXiv:hep-th/0011217.

[89] O. Lunin, J. M. Maldacena, and L. Maoz, "Gravity solutions for the D1-D5 system with angular momentum", arXiv:hep-th/0212210.

[90] O. Lunin, "Adding momentum to D-1 - D-5 system", *JHEP* **04** (2004) 054, arXiv:hep-th/0404006.

[91] S. Giusto, S. D. Mathur, and A. Saxena, "Dual geometries for a set of 3-charge microstates", *Nucl. Phys.* **B701** (2004) 357–379, arXiv:hep-th/0405017.

[92] S. Giusto, S. D. Mathur, and A. Saxena, "3-charge geometries and their CFT duals", *Nucl. Phys.* **B710** (2005) 425–463, arXiv:hep-th/0406103.

[93] S. Giusto and S. D. Mathur, "Geometry of D1-D5-P bound states", *Nucl. Phys.* **B729** (2005) 203–220, arXiv:hep-th/0409067.

[94] I. Bena and N. P. Warner, "Bubbling supertubes and foaming black holes", *Phys. Rev.* **D74** (2006) 066001, arXiv:hep-th/0505166.

[95] P. Berglund, E. G. Gimon, and T. S. Levi, "Supergravity microstates for BPS black holes and black rings", *JHEP* **06** (2006) 007, arXiv:hep-th/0505167.

[96] G. W. Gibbons and N. P. Warner, "Global structure of five-dimensional fuzzballs", *Class. Quant. Grav.* **31** (2014) 025016, arXiv:1305.0957.

[97] V. Jejjala, O. Madden, S. F. Ross, and G. Titchener, "Non-supersymmetric smooth geometries and D1-D5-P bound states", *Phys. Rev.* **D71** (2005) 124030, arXiv:hep-th/0504181.

[98] V. Cardoso, O. J. C. Dias, J. L. Hovdebo, and R. C. Myers, "Instability of non-supersymmetric smooth geometries", *Phys. Rev.* **D73** (2006) 064031, arXiv:hep-th/0512277.

[99] J. C. Breckenridge, R. C. Myers, A. W. Peet, and C. Vafa, "D-branes and spinning black holes", *Phys. Lett.* **B391** (1997) 93–98, arXiv:hep-th/9602065.

[100] V. Cardoso, O. J. C. Dias, and R. C. Myers, "On the gravitational stability of D1-D5-P black holes", *Phys. Rev.* **D76** (2007) 105015, arXiv:0707.3406.

[101] O. J. C. 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.

[102] G. Festuccia and H. Liu, "A Bohr-Sommerfeld quantization formula for quasinormal frequencies of AdS black holes", *Adv. Sci. Lett.* **2** (2009) 221–235, arXiv:0811.1033.

[103] O. Gannot, "Quasinormal Modes for Schwarzschild-AdS Black Holes: Exponential Convergence to the Real Axis", *Commun. Math. Phys.* **330** (2014) 771–799, arXiv:1212.1907.

[104] V. Cardoso, L. C. B. Crispino, C. F. B. Macedo, H. Okawa, and P. Pani, "Light rings as observational evidence for event horizons: long-lived modes, ergoregions and nonlinear instabilities of ultracompact objects", *Phys. Rev.* **D90** (2014), no. 4, 044069, arXiv:1406.5510.

[105] J. Keir, "Wave propagation on microstate geometries", arXiv:1609.01733.

[106] S. D. Mathur, "The Fuzzball proposal for black holes: An Elementary review", *Fortsch. Phys.* **53** (2005) 793–827, arXiv:hep-th/0502050.

[107] S. Aretakis, "Stability and Instability of Extreme Reissner-Nordström Black Hole Spacetimes for Linear Scalar Perturbations I", *Commun. Math. Phys.* **307** (2011) 17–63, arXiv:1110.2007.

[108] S. Aretakis, "Stability and Instability of Extreme Reissner-Nordstrom Black Hole Spacetimes for Linear Scalar Perturbations II", *Annales Henri Poincare* **12** (2011) 1491–1538, arXiv:1110.2009.

[109] H. K. Kunduri and J. Lucietti, "Supersymmetric Black Holes with Lens-Space Topology", *Phys. Rev. Lett.* **113** (2014), no. 21, 211101, arXiv:1408.6083.

[110] S. Tomizawa and M. Nozawa, "Supersymmetric black lenses in five dimensions", *Phys. Rev.* **D94** (2016), no. 4, 044037, arXiv:1606.06643.

[111] H. K. Kunduri and J. Lucietti, "Black hole non-uniqueness via spacetime topology in five dimensions", *JHEP* **10** (2014) 082, arXiv:1407.8002.

[112] S. Giusto, L. Martucci, M. Petrini, and R. Russo, "6D microstate geometries from 10D structures", *Nucl. Phys.* **B876** (2013) 509–555, arXiv:1306.1745.

[113] I. Bena, S. Giusto, E. J. Martinec, R. Russo, M. Shigemori, D. Turton, and N. P. Warner, "Smooth horizonless geometries deep inside the black-hole regime", *Phys. Rev. Lett.* **117** (2016), no. 20, 201601, arXiv:1607.03908.

[114] G. W. Gibbons, D. Kastor, L. A. J. London, P. K. Townsend, and J. H. Traschen, "Supersymmetric selfgravitating solitons", *Nucl. Phys.* **B416** (1994) 850–880, arXiv:hep-th/9310118.

[115] B. E. Niehoff and H. S. Reall, "Evanescent ergosurfaces and ambipolar hyperkähler metrics", *JHEP* **04** (2016) 130, arXiv:1601.01898.

[116] J. B. Gutowski, D. Martelli, and H. S. Reall, "All Supersymmetric solutions of minimal supergravity in six- dimensions", *Class. Quant. Grav.* **20** (2003) 5049–5078, arXiv:hep-th/0306235.

[117] M. Dafermos and I. Rodnianski, "A New physical-space approach to decay for the wave equation with applications to black hole spacetimes", in "XVIth International Congress on Mathematical Physics, P. Exner (ed.), World Scientific, London, 2009, pp. 421-433", pp. 421–433. 2009. arXiv:0910.4957.

[118] M. Dafermos, G. Holzegel, and I. Rodnianski, "A scattering theory construction of dynamical vacuum black holes", arXiv:1306.5364.

[119] M. Dafermos and I. Rodnianski, "Lectures on black holes and linear waves", *Clay Math. Proc.* **17** (2013) 97–205, arXiv:0811.0354.

[120] S. Aretakis, "Horizon Instability of Extremal Black Holes", *Adv. Theor. Math. Phys.* **19** (2015) 507–530, arXiv:1206.6598.

[121] J. Lucietti and H. S. Reall, "Gravitational instability of an extreme Kerr black hole", *Phys. Rev.* **D86** (2012) 104030, arXiv:1208.1437.

[122] I. Bena and P. Kraus, "Three charge supertubes and black hole hair", *Phys. Rev.* **D70** (2004) 046003, arXiv:hep-th/0402144.

[123] I. Bena and N. P. Warner, "Black holes, black rings and their microstates", *Lect. Notes Phys.* **755** (2008) 1–92, arXiv:hep-th/0701216.

[124] I. Bena, E. J. Martinec, R. Walker, and N. P. Warner, "Early Scrambling and Capped BTZ Geometries", *JHEP* **04** (2019) 126, arXiv:1812.05110.

[125] R. Emparan and H. S. Reall, "Black Rings", *Class. Quant. Grav.* **23** (2006) R169, arXiv:hep-th/0608012.

[126] O. Lunin and S. D. Mathur, "The Slowly rotating near extremal D1 - D5 system as a 'hot tube'", *Nucl. Phys.* **B615** (2001) 285–312, arXiv:hep-th/0107113.

[127] V. Ferrari and B. Mashhoon, "New approach to the quasinormal modes of a black hole", *PRD* **30** (1984) 295–304.

[128] B. Chakrabarty, D. Turton, and A. Virmani, "Holographic description of non-supersymmetric orbifolded D1-D5-P solutions", *JHEP* **11** (2015) 063, arXiv:1508.01231.

[129] M. Abramowitz and I. A. Stegun, "Handbook of mathematical functions with formulas, graphs, and mathematical tables", Dover, New York, 1964.

[130] S. Giusto, O. Lunin, S. D. Mathur, and D. Turton, "D1-D5-P microstates at the cap", *JHEP* **02** (2013) 050, arXiv:1211.0306.

[131] V. Cardoso, O. J. C. Dias, G. S. Hartnett, L. Lehner, and J. E. Santos, "Holographic thermalization, quasinormal modes and superradiance in Kerr-AdS", *JHEP* **04** (2014) 183, arXiv:1312.5323.

[132] F. C. Eperon, "Geodesics in supersymmetric microstate geometries", *Class. Quant. Grav.* **34** (2017), no. 16, 165003, arXiv:1702.03975.

[133] S. Chandrasekhar, *The Mathematical Theory of Black Holes*, pp. 326–331. Springer Netherlands, Dordrecht, 1984.

[134] E. Teo, "Spherical Photon Orbits Around a Kerr Black Hole", *General Relativity and Gravitation* **35** (2003), no. 11, 1909–1926.

[135] V. Diemer, J. Kunz, C. Lämmerzahl, and S. Reimers, "Dynamics of test particles in the general five-dimensional Myers-Perry spacetime", *Phys. Rev.* **D89** (2014), no. 12, 124026, arXiv:1404.3865.

[136] V. Diemer and J. Kunz, "Supersymmetric rotating black hole spacetime tested by geodesics", *Phys. Rev.* **D89** (2014), no. 8, 084001, arXiv:1312.6540.

[137] Y. Chervonyi and O. Lunin, "(Non)-Integrability of Geodesics in D-brane Backgrounds", *JHEP* **02** (2014) 061, arXiv:1311.1521.

[138] R. C. Myers and M. J. Perry, "Black Holes in Higher Dimensional Space-Times", *Annals Phys.* **172** (1986) 304.

[139] J. M. Bardeen, W. H. Press, and S. A. Teukolsky, "Rotating Black Holes: Locally Nonrotating Frames, Energy Extraction, and Scalar Synchrotron Radiation", *Astrophysical Journal* **178** (1972) 347–370.

[140] H. Elvang, R. Emparan, D. Mateos, and H. S. Reall, "A Supersymmetric black ring", *Phys. Rev. Lett.* **93** (2004) 211302, arXiv:hep-th/0407065.

[141] F. R. Tangherlini, "Schwarzschild field inn dimensions and the dimensionality of space problem", *Il Nuovo Cimento (1955-1965)* **27** (1963), no. 3, 636–651.

[142] J. L. Friedman, "Ergosphere instability", *Comm. Math. Phys.* **63** (1978), no. 3, 243–255.

[143] O. Gannot, "Quasinormal Modes for Schwarzschild–AdS Black Holes: Exponential Convergence to the Real Axis", *Communications in Mathematical Physics* **330** (2014), no. 2, 771–799.

[144] D. Marolf, B. Michel, and A. Puhm, "A rough end for smooth microstate geometries", *JHEP* **05** (2017) 021, arXiv:1612.05235.

[145] O. J. C. Dias, F. C. Eperon, H. S. Reall, and J. E. Santos, "Strong cosmic censorship in de Sitter space", *Phys. Rev.* **D97** (2018), no. 10, 104060, arXiv:1801.09694.

[146] M. Simpson and R. Penrose, "Internal instability in a Reissner-Nordström black hole", *International Journal of Theoretical Physics* **7** Apr (1973) 183–197.

[147] C. M. Chambers, "The Cauchy horizon in black hole de sitter space-times", *Annals Israel Phys. Soc.* **13** (1997) 33, arXiv:gr-qc/9709025, [,33(1997)].

[148] V. Cardoso, J. L. Costa, K. Destounis, P. Hintz, and A. Jansen, "Quasinormal modes and Strong Cosmic Censorship", *Phys. Rev. Lett.* **120** (2018), no. 3, 031103, arXiv:1711.10502.

[149] P. Hintz and A. Vasy, "Analysis of linear waves near the Cauchy horizon of cosmological black holes", *J. Math. Phys.* **58** (2017), no. 8, 081509, arXiv:1512.08004.

[150] J. L. Costa, P. M. Girão, J. Natário, and J. D. Silva, "On the Occurrence of Mass Inflation for the Einstein–Maxwell-Scalar Field System with a Cosmological Constant and an Exponential Price Law", *Commun. Math. Phys.* **361** (2018), no. 1, 289–341, arXiv:1707.08975.

[151] S. Hod, "Strong cosmic censorship in charged black-hole spacetimes: As strong as ever", arXiv:1801.07261.

[152] O. J. C. Dias, H. S. Reall, and J. E. Santos, "Strong cosmic censorship for charged de Sitter black holes with a charged scalar field", *Class. Quant. Grav.* **36** (2019), no. 4, 045005, arXiv:1808.04832.

[153] O. J. C. Dias, H. S. Reall, and J. E. Santos, "Strong cosmic censorship: taking the rough with the smooth", *JHEP* **10** (2018) 001, arXiv:1808.02895.

[154] M. Dafermos and Y. Shlapentokh-Rothman, "Rough initial data and the strength of the blue-shift instability on cosmological black holes with $\Lambda > 0$", *Class. Quant. Grav.* **35** (2018), no. 19, 195010.

[155] R. Luna, M. Zilhão, V. Cardoso, J. L. Costa, and J. Natário, "Strong Cosmic Censorship: the nonlinear story", *Phys. Rev.* **D99** (2019), no. 6, 064014, arXiv:1810.00886.

[156] B. Carter, "Hamilton-jacobi and schrödinger separable solutions of einstein's equations", *Comm. Math. Phys.* **10** (1968), no. 4, 280–310.

[157] C. M. Chambers and I. G. Moss, "Stability of the Cauchy horizon in Kerr-de Sitter space-times", *Class. Quant. Grav.* **11** (1994) 1035–1054, arXiv:gr-qc/9404015.

[158] O. J. C. Dias, J. E. Santos, and B. Way, "Numerical Methods for Finding Stationary Gravitational Solutions", *Class. Quant. Grav.* **33** (2016), no. 13, 133001, arXiv:1510.02804.

[159] E. Newman and R. Penrose, "An Approach to Gravitational Radiation by a Method of Spin Coefficients", *Journal of Mathematical Physics* **3** (1962), no. 3, 566–578, https://doi.org/10.1063/1.1724257.

[160] S. A. Teukolsky, "Rotating black holes: Separable wave equations for gravitational and electromagnetic perturbations", *Phys. Rev. Lett.* **29** Oct (1972) 1114–1118.

[161] S. A. Teukolsky, "Perturbations of a rotating black hole. 1. Fundamental equations for gravitational electromagnetic and neutrino field perturbations", *Astrophys. J.* **185** (1973) 635–647.

[162] O. J. C. Dias and J. E. Santos, "Boundary Conditions for Kerr-AdS Perturbations", *JHEP* **10** (2013) 156, arXiv:1302.1580.

[163] F. C. Eperon, H. S. Reall, and J. J. Sbierski, "Predictability of subluminal and superluminal wave equations", arXiv:1802.03020.

[164] A. Adams, N. Arkani-Hamed, S. Dubovsky, A. Nicolis, and R. Rattazzi, "Causality, analyticity and an IR obstruction to UV completion", *JHEP* **10** (2006) 014, arXiv:hep-th/0602178.

[165] E. Babichev, V. Mukhanov, and A. Vikman, "k-Essence, superluminal propagation, causality and emergent geometry", *JHEP* **02** (2008) 101, arXiv:0708.0561.

[166] R. Geroch, "Faster Than Light?", *AMS/IP Stud. Adv. Math.* **49** (2011) 59–70, arXiv:1005.1614.

[167] G. Papallo and H. S. Reall, "Graviton time delay and a speed limit for small black holes in Einstein-Gauss-Bonnet theory", *JHEP* **11** (2015) 109, arXiv:1508.05303.

[168] B. M. Barbashov and N. A. Chernikov, "Solution and Quantization of a Nonlinear Two-dimensional Model for a Born-Infeld Type Field", *Soviet Journal of Experimental and Theoretical Physics* **23** (1966) 861.

[169] B. M. Barbashov and N. A. Chernikov, "Solution of the two plane wave scattering problem in a nonlinear scalar field theory of the Born–Infeld type", *Sov. Phys. JETP* **24** (1967), no. 2, 437–442, [Zh. Eksp. Teor. Fiz.51,no.2,658(1966)].

[170] D. Christodoulou, "The formation of shocks in 3-dimensional fluids", European Mathematical Society, 2007.

[171] S. W. Hawking, "Chronology protection conjecture", *Phys. Rev. D* **46** Jul (1992) 603–611.

[172] Y. Choquet-Bruhat and R. Geroch, "Global aspects of the Cauchy problem in general relativity", *Communications in Mathematical Physics* **14** Dec (1969) 329–335.

[173] J. Beem, P. Ehrlich, and K. Easley, "Global lorentzian geometry, second edition", Taylor & Francis, 1996.

[174] T. Taniuti, "On Wave Propagation in Non-Linear Fields", *Progress of Theoretical Physics Supplement* **9** 01 (1959) 69–128, http://oup.prod.sis.lan/ptps/article-pdf/doi/10.1143/PTPS.9.69/5277402/9-69.pdf.

[175] G. Whitham, "Linear and nonlinear waves", Wiley, 2011.

[176] R. Penrose, "Gravitational collapse: The role of general relativity", *Riv. Nuovo Cim.* **1** (1969) 252–276, [Gen. Rel. Grav.34,1141(2002)].

[177] S. Iyer and C. M. Will, "Black Hole Normal Modes: A WKB Approach. 1. Foundations and Application of a Higher Order WKB Analysis of Potential Barrier Scattering", *Phys. Rev.* **D35** (1987) 3621.

[178] E. Berti, V. Cardoso, and A. O. Starinets, "Quasinormal modes of black holes and black branes", *Class. Quant. Grav.* **26** (2009) 163001, arXiv:0905.2975.

[179] A. Vasy and S. Dyatlov, "Microlocal analysis of asymptotically hyperbolic and Kerr-de Sitter spaces", arXiv:1012.4391.

[180] B. Helffer, "Semi-classical analysis for the schrödinger operator and applications", Springer Berlin Heidelberg, 2006.

[181] M. Dimassi and J. Sjostrand, "The wkb-method for a potential minimum", Cambridge University Press, 1999.