IMPERIAL COLLEGE OF LONDON

MSc: Quantum Fields and Fundamental Forces

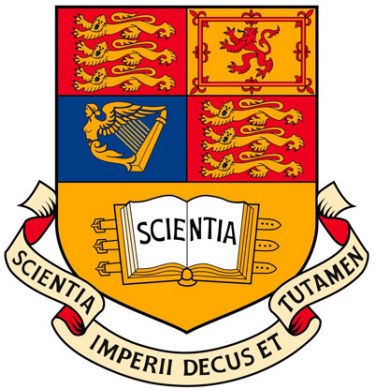
Master's Thesis

Instability of extreme Reissner-Nordström black holes

Marc Arène

Supervised by Dr Toby Wiseman

September 2014



Submitted in partial fulfilment of the requirements for the degree of Master of Science of Imperial College London

Abstract

Extreme black holes are of particular interest to us since they make the link between their non-extreme counterparts and naked singularities which are forbiden by the Weak Cosmic Censorship Conjecture. A subtle instability occuring on the event horizon of an extreme Reissner-Nordström black hole has recently been exhibited by Stephanos Aretakis. After exposing the physical motivations for the existence of such an instability, we intend to review the analyses carried out on that subject together with the context in which they find themselves, and finally confront our predictions with numerical simulations.

Acknowledgements

I would like to thank my supervisor, Dr Toby Wiseman, for taking the time to answer all my questions and carefully explain to me the subtleties hidden in this topic. Thanks to him, I have learnt a lot about black holes in general, but more importantly about how such a theoretical issue has its concrete importance in physics.

Contents

[1]For further details about the derivations done in this section and about the physical interpretations, the reader should refer to [19], section 12.5.

1. Introduction

Black holes constitute a central concept in the celebrated theory of general relativity. The only spherically symmetric solution of Einstein's vacuum equation is the Schwarzschild metric, leading to the one parameter family of Schwarzschild black holes, defined by their mass M. Reissner and Nordström found a similar but more complicated (however still spherically symmetric) solution in the case of a charged mass; hence leading to the two parameter (mass M and charge Q) family of Reissner-Nordström black holes. Finally the Kerr family describes uncharged rotating black holes, thus characterized by two parameters: mass M and spin parameter a[1]. Note however that their rotation breaks the spherical symmetry present in the two other families.

A major difference between Schwarzschild black holes and the two other families should be noticed. For Reissner-Nordström and Kerr, the mass parameter M is bounded by below by respectively Q and a. Indeed having either $M < Q$ or $M < |a|$ leads to naked singularities which are excluded by the Weak Cosmic Censorship Conjecture. However the case of equality remains theoretically acceptable, and such black holes are called **extreme**.

The physical relevance of the black hole notion rests in the expectation that black holes are stable objects. Schwarzschild black holes, non-extreme Reissner-Nordström and very recently (see [8]) non-extreme Kerr have been proved to be stable against scalar perturbations. The stability of the extreme cases was then assumed but not demonstrated. However, it is only quite recently that an instability has been highlighted for extreme Reissner-Nordström black holes.

In order to understand the context in which the study of this instability appears, we will start by reviewing the history of the main developments that have been achieved so as to prove the stability of the different families of black holes; then, in order to link our futur analysis to a more concrete domain, we will outline what is known about astrophysical black holes; finally to complete this introduction we should clarify the motivations for this study.

[1]Related to the angular momentum J by $a = J/M$

1.1 History

After the Schwarzschild solution was found in 1916, the question of its stability against perturbations was settled and partially solved by T. Regge and J.A. Wheeler in 1957 (see [15]). They started to prove the linear stability of the Scharwzschild metric using in particular classical mode analysis which relies on decompositions into spherical harmonics. Thereafter the full analysis was completed in 1970 by L.A. Edelstein and C.V. Vishveshwara in [16]. R. Wald alone in 1978 in [17] and then with B. Kay in 1986 in [18] provided a final point so as to proove the complete linear stability of the Schwarzschild metric. Their method did not use classical mode analysis and could be applicable to the Reissner-Nordström's metric stability problem; however it wouldn't be able to help in the Kerr case because superradiance occurs in the ergoregion[2]. A simpler and more robust proof was derived using the well-known **redshift effect** and hence bringing a physical argument supporting stability. A major issue remained to be proved: the Kerr case. It is only a few months ago, in February of this year, that M. Dafermos and I. Rodniansky made the first claim in [8] for a complete proof of the stability of the Kerr metric.

All the proofs and developments mentioned before only addressed the non-extreme cases. It is of course sufficient when dealing with the Schwarzschild metric as there is no extremality possible. However, for the Reissner-Nordström and Kerr backgrounds, the stability in the extreme case was surprisingly assumed for many years. It is indeed only in 2011 that light was shed on a subtle instability, occuring in an extreme Reissner-Nordström background, by Stephanos Aretakis in [3] and [4]. His work has settled the foundations of this new issue.

1.2 Black holes in astrophysics

At first black holes were thought to be only a theoretical invention, a weird idea, and Albert Einstein himself put their existence into question. We have known for long now from observation that they do exist in the universe and are not just a mathematical construction. All astrophysical black holes in the universe belong to the Kerr family and as a consequence are described by their two parameters M and a. The absence of Reissner-Nordström family of black holes comes from the electro-neutrality of matter in the universe, and if ever some charged particules were to fall in, opposite charges would very soon be swallowed in as a consequence of Coulomb's law hence making the total charge Q of the black hole null.

[2]Superradiance is the amplification of wavemodes in the ergoregion, mainly due to the fact that the stationary timelike Killing vector $\frac{\partial}{\partial_t}$ becomes spacelike in this region. See [33].

1.2.1 Stellar-mass and supermassive black holes

From observation, we can distinguish two kinds of astrophysical black holes: the stellar-mass and the supermassive ones.[3]

Stellar-mass black holes are often born from the collapse of sufficiently massive stars and their mass can add up to several tens that of the sun (defined by $M_\odot$): $5M_\odot < M_{stellar} < 20M_\odot$[4]. The number of stellar black holes in our galaxy is thought to be around 300 millions [31]. Stellar-mass black holes can be gravitationally coupled with a star named the *companion*. The two of them then form what is called a *binary system*. The black hole and its companion orbit around each other and the star leaves a tail of matter behind her which continuously feeds the black hole (see fig.1.1). Note that the orbit period of the binary system is generally very short considering such massive objects, ranging from a month to several hours only. This weak orbit period is of interest as it allows astronomers to observe phenomena over a short time when those same phenomena stretch over hundreds of years for supermassive black holes.

Contrary to stellar-mass black holes, we infer that there exists only one supermassive black hole in a galaxy, located at its very center. Nonetheless the birth of such black holes remains an open field of research. Their mass reaches millions or billions of times that of the sun: $10^6 M_\odot < M_{supermassive} < 10^9 M_\odot$, hence earning their title of supermassive. However, because the density of a black hole is inversely proportional to the square of their mass[5], supermassive black holes are not very compact object. Some of them even have the same density as water. As a result the tidal forces experienced by someone sitting on the horizon of a supermassive black hole would be comparable to those we experience daily on Earth.

1.2.2 Observations and measurements

Observations

Because there is no radiation coming out of a black hole[6], a direct observation with our telescopes is quite complicated. In order to get around this problem, astronomers use the gravitational effects that black holes generate on their local environment. Due to their incredible mass (and density for stellar-mass black holes), surrounding objects are unrelentingly attracted by black holes; those objects become subject to relativistic effects and display various behaviours which can be exploited to deduce the presence of a very compact object nearby. However black holes are not the only compact object in the universe. Neutron stars, which are the result of the collapse of a star when it is just

[3] The existence of a third family: *intermediate-mass black holes* is still subject to debate, however at least one black hole at this time: ESO 243-49/HLX-1 has convincingly demonstrated in [30] to have a mass of $\sim 500M_\odot$.

[4] The upper limit is quite arbitrary here but can be justified by the most massive stellar black hole observed so far which has a mass of about $15M_\odot$. On the contrary, the existence of a lower limit has been observationally proved in [28].

[5] Considering we can define the volume of the black hole as the volume of the "horizon sphere".

[6] Excluding obviously Hawking radiation, totally negligible for our concerns here.

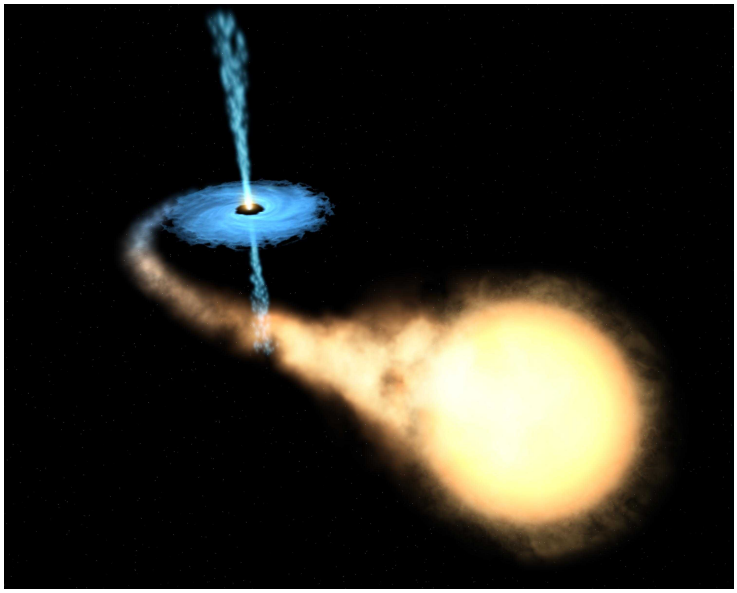


Figure 1.1: Artist representation of a stellar-mass black hole part of a binary system. The companion star is continuoulsy feeding the accretion disk while the black hole emits a relativistic jet.

not massive enough to produce a black hole, are dark and show a very high density as well. Hence they often come in competition with stellar-mass black holes when a compact object has been found. This does not mean we can't distinguish them. Neutron star display various proprieties (like thermnuclear flashes, a very high magnetic field) which can be exploited in that purpose[7]. Note that in [28], authors have demonstrated observationally the peculiar existence of a gap between the maximum value of a neutron star's mass: $\approx 3M_{\odot}$, and the minimum value of that of a black hole: $\approx 5M_{\odot}$ when actually we would expect a smooth transition.

The observation of stellar-mass black holes can be indirectly made when they are part of a binary system. Indeed, even though we can't observe the astrophysical object orbiting with the star, only a very compact object fits the description of the system.

Both families of astrophysical black holes are surrounded by a disk of matter spinning around them called the *disk of accretion*. It is made of any kind of matter (cloud of gas, dust, stars...) that has drifted to close to the black hole and thus will fall into it at some point. It is submitted to extreme conditions: before crossing the event horizon, matter at the inner edge of the disk has a velocity close to that of light, it gets very dense as well so its temperature can reach millions of degrees. As a consequence it will radiate in the X-ray spectrum. Astronomers exploit this specific radiation to deduce the existence of a compact object.

Nevertheless, this observation is indirect. For a long time astronomers have been looking for a direct observation of the event horizon of a black hole. A true evidence for this would be the observation of the progressive destruction of a cloud of gas being pulled in by the

[7] See section 12.2 of [24] for a further discussion.

black hole. It is actually now happenning with Sgr A*, the supermassive black hole at the center of the Milky Way, where a cloud of gas only three times the mass of the Earth is shinning brightly while falling into the black hole at a speed of (now) $8 \cdot 10^6$ km/h. ESO telescopes are monitoring this opportunity for the first direct evidence of an event horizon [29].

Measurement of the spin a

Measuring the mass of a black hole can be very easy when it is part of a binary system. One only needs to observe the orbit of the companion star and apply Kepler's law. However some uncertainties remain for single black holes [24].

The measurement of the spin a is never straightforward as it may be for the mass. Nevertheless the disk of accretion is spinning around the black hole, therefore it becomes a fundamental tool for the determination of a. Note that the disk can spin retrogradely, i.e. in the reverse direction as that of the black hole, in which case we set by convention a negative value for a. A prograde rotation has $a > 0$. However there exists different types of disks and each type behaves differently. The importance of the disk of accretion has led to the development of a new theory: *the black hole accretion disk theory* [24] which has given birth to several models which have successfully simulated disks: thin, slim, thick, ADAF. The most commonly used method to measure the spin is the *spectral fitting of the X-ray continuum*. However it has so far only been applied to stellar-mass black holes in X-ray binaries. In order to determine the parameter a using this method, one separately needs to know the mass M of the black hole, the inclination i of the disk and the distance D to the companion star. Fig.(1.2) shows a distribution of the spin a with respect to the mass of six black holes. One will notice that some of them rotate very quickly and thus are very close to being extreme.

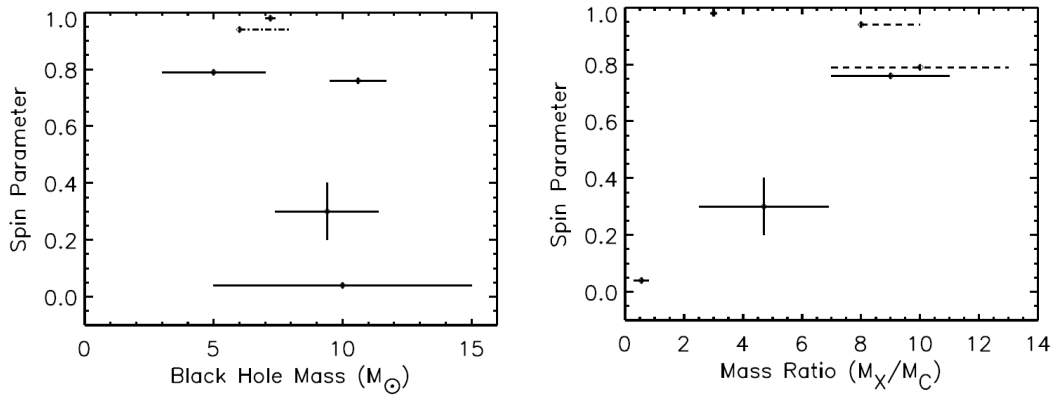


Figure 1.2: Plot of a for six different stellar black holes part of a X-ray binary system using the spectral fitting of the X-ray continuum method. On the right hand side plot, M_X is the mass of the black hole and M_C that of its companion. See [27] for details. Of course $a < M$, however we see that some black hole rotate very quickly meaning they are *almost-extreme*.

1.3 Motivations

1.3.1 Why extreme

Extreme black holes form a very special case in the sens that pure extremality can never be reached by a black hole (see further discussion in section 3.2.2). Indeed the mass M will always be a tiny bit bigger than the spin a and strictly speaking the black hole is non-extreme. Hence we could talk of a black hole's degree of extremality, it being 98% extreme if $a = 0.98M$. The reason why we persist with the study of extreme black holes is because we hope to highlight some significant behaviour(s) which, being shared by *almost-extreme* black holes, will allow us to learn more about these peculiar objects and maybe enable us to distinguish them in our observations.

For instance relativistic jets might be connected to a high spinning of the black hole. A jet is a long and thin steam of matter dramatically ejected from a compact region very close to the black hole, usually in a direction perpendicular to the plane described by the accretion disk (see fig.1.1). The correlation between extremality and these jets is still at debate [32], however the study from [27] brings modest supports towards that connection as the most powerfull jets have been linked to the fastest rotating black holes. Understanding the instability of extreme black holes on a theoretical point of view might bring further arguments supporting or refuting this connection.

1.3.2 Why Reissner-Nordström

As mentionned previously, astrophysical black holes have a null charge Q and they all belong to the Kerr family (when $a = 0$ it reduces to the Schwarzschild family). Our focus on the Reissner-Nordström case might thereafter seem surprising. The reason for this, beyond the intellectual curiosity we all share, is a matter of simplicity. Contrary to Kerr, the Reissner-Nordström solution of Einstein's equation is spherically symmetric which makes the analysis much simpler. We therefore aim to catch important features of this instability which would certainly help for the analysis of the Kerr case. Note however that the previous motivation put forward concerning the possible connection between jets and a high spin parameter relies on the rotation of the black hole and this rotation is precisely what breaks the spherical symmetry of the system.

We will start in section 2 by reviewing the Reissner-Nordström spacetime and define some notions which we will need for our analysis. Section 3 will begin by a necessary inspection of the concepts of stability in general relativity; then, after an outline of the physical motivations for an instability in the extreme case, we will proove its existence through two different approaches and confront our conclusions with numerical simulations. However details about the latter and especially about the choices for initial data will only be given in section 4, which will as well give further numerical support to our analysis.

2. Review of Reissner-Nordström black holes

In this chapter we will review the main aspects of Reissner-Nordström black holes and take that opportunity to define our notations. Black holes metrics appear as solutions to Einstein's equation. The latter is a set of nonlinear equations governing the whole theory of general relativity. It relates curvature of space, described by a metric g solution of the equation, to the distribution of matter an energy, described by the stress-energy tensor $T_{\mu\nu}$:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} = 8\pi T_{\mu\nu}, \tag{2.1}$$

where the Ricci tensor $R_{\mu\nu}$ and the Ricci scalar R depend on g. In vacuum, where $T_{\mu\nu} = 0$, it reads:

$$R_{\mu\nu} = 0. \tag{2.2}$$

The Schwarzschild metric is the only spherically symmetric solution of eq.(2.2). However we already asserted that the solution found by Reissner and Nordström shares this spherical symmetry. This is possible because the Reissner-Nordström metric is not exactly a solution of eq.(2.2). Indeed, if we consider a charged mass, it implies the presence of an electromagnetic field which means the space considered is not in pure vacuum anymore. Nevertheless we refer to it as *electrovacuum*. In the end, the Reissner-Nordström metric is the unique spherical solution to Einstein-Maxwell equations (or equivalently to Einstein's electrovacuum equation):

$$R_{\mu\nu} = 8\pi T_{\mu\nu} \tag{2.3}$$

$$\nabla^{\mu}F_{\mu\nu} = 0 \tag{2.4}$$

$$\nabla_{[\mu}F_{\nu\rho]} = 0 \tag{2.5}$$

where $F_{\mu\nu}$ is the electromagnetic tensor field related to $T_{\mu\nu}$ by
$T_{\mu\nu} = \frac{1}{4\pi}\left(F_{\mu\rho}F_{\nu}^{\rho} - \frac{1}{4}g_{\mu\nu}F_{\alpha\beta}F^{\alpha\beta}\right)$ [19].

2.1 From the metric

Therefore we consider a black hole with a total mass M and a total electric charge Q. Let's examine the non-extreme case $Q < M$ and the extreme one where $Q = M$ separately.

$\underline{Q < M}$:

The solution of Einstein's equation in this case finds the following metric:

$$ds^2 = -(1 - \frac{2M}{r} + \frac{Q^2}{r^2})dt^2 + (1 - \frac{2M}{r} + \frac{Q^2}{r^2})^{-1}dr^2 + r^2 d\Omega^2, \tag{2.6}$$

where $d\Omega$ is the line element on a round unit two sphere. This can be rewritten as:

$$ds^2 = -\frac{(r - r_+)(r - r_-)}{r^2}dt^2 + \left(\frac{(r - r_+)(r - r_-)}{r^2}\right)^{-1} dr^2 + r^2 d\Omega^2, \tag{2.7}$$

where $r_\pm = M \pm \sqrt{M^2 - Q^2}$. r_+ corresponds to the event horizon of the black hole and r_- is the inner horizon. Here $r_- < M < r_+$.

$\underline{Q = M}$:

One immediatly notices that $r_+ = r_- = M$ and the metric becomes:

$$ds^2 = -(1 - \frac{M}{r})^2 dt^2 + (1 - \frac{M}{r})^{-2}dr^2 + r^2 d\Omega^2. \tag{2.8}$$

Both cases:

By defining $F(r) = \frac{(r-r_+)(r-r_-)}{r^2}$ we can generically write the metric as:

$$ds^2 = -F(r)dt^2 + F(r)^{-1}dr^2 + r^2 d\Omega^2. \tag{2.9}$$

So far we have considered the Reissner-Nordström metric in local coordinates (t, r, θ, ϕ). The metric behaves badly in those coordinate as the event horizon $r = r_+$ appears as a singularity. To avoid this coordinate singularity, we usually describe black holes in *Eddington-Finkelstein* coordinates. As a consequence, those coordinates are much more suited for the study of black holes. However once we switch to those coordinates an important difference appears between the extreme and non-extreme cases.

The Eddington-Finkelstein coordinate system relies on the so called *tortoise* coordinate $r_*(r)$ which in both cases is defined such that:

$$dr_*(r) = F(r)^{-1}dr. \tag{2.10}$$

That way, if we go to the ingoing Eddington-Finkelstein coordinates (v, r) with $v = t + r_*$, we get the following metric:

$$ds^2 = -F(r)dv^2 + 2dvdr + r^2 d\Omega^2. \tag{2.11}$$

The interesting feature to notice is the expression of r_* in both cases:

non-extreme: $Q < M$:[1]

$$r_*(r) = r + \frac{1}{2\kappa_+}\ln|\frac{r - r_+}{r_+}| + \frac{1}{2\kappa_-}\ln|\frac{r - r_-}{r_-}|, \tag{2.12}$$

[1] κ_+ and κ_- are two constants which depend only on r_+ and r_-.

extreme: $Q = M$:

$$r_*(r) = r - M + 2M \ln |\frac{r-M}{M}| - \frac{M^2}{r-M}. \tag{2.13}$$

In the non-extreme case, r_* is logarithmic in r whereas in the extreme case it is inverse linear. This way, when in the (t, r, θ, ϕ) expressions of the metrics it appeared we could continuously go from one case to the other, we see that in the Eddington-Finkelstein system, which is the right one to consider as the metrics are well defined, there is a discontinuity between the non-extreme case and the extreme one.

This discontinuity is very symptomatic of the instability we are interesded in as it is sign of a *gap* which exists between non-extremality and extremality.

2.2 Penrose diagrams

This *gap* can also be observed by comparing the Penrose diagrams of the two cases.

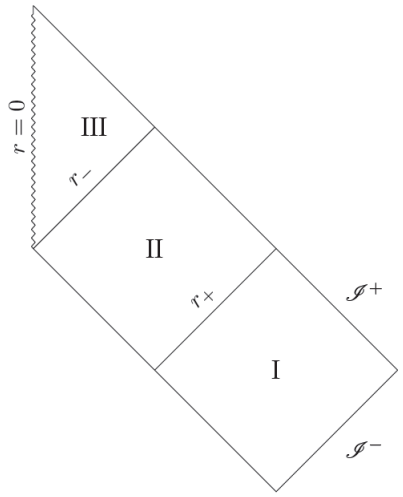


Figure 2.1: **non-extreme**

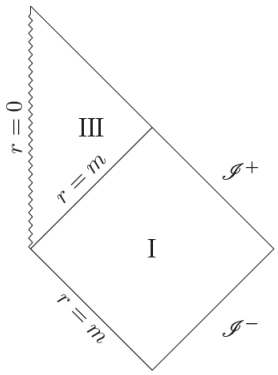


Figure 2.2: **extreme**

Figure 2.3: Carter-Penrose diagram of a Reissner-Nordström black hole in the non-extreme case and extreme cases.

We can see that the shrink from the diagram in the non-extreme case to the one in the extreme case isn't continuous as $M \to Q$. While one would have thought of continuously bringing the borders of region II in fig.(2.1) defined by $r = r_+$ and $r = r_-$ closer one to another as $M \to Q$, this is actually not possible. Indeed because of the very way Penrose diagrams are built (first going to Kruskal coordinates and then compactifying), regions I and II have to remain squared. They will do so as long as $Q < M$ and suddenly shrink when $M = Q$.

2.3 The surface gravity[2]

The surface gravity plays a very important role in the physics of black holes in general and, as we will see, a very subtle but no less important one in the instability we are interested in. The three families of black holes: Schwarzschild, Reissner-Nordström and Kerr are stationary, that is to stay they contain an asymptotically timelike Killing vector field. Furthermore this implies that their horizon is a Killing horizon, ie a null hypersurface which normal vector is a Killing one. For Schwarzschild and Reissner-Nordström, their spherical symmetry impose them, via Birkoff's theorem[3], to be static and $\xi = \dfrac{\partial}{\partial v}$ is the Killing vector normal to the horizon. Now if we consider the integral curves[4] of ξ generating the event horizon; these are not affinely parametrised in general, but there will always exist a vector field l, proportional to ξ, which will be. Hence defining $l^\mu = f\xi^\mu$, we get:

$$l \cdot \nabla l = 0, \tag{2.14}$$

leading to

$$\xi \cdot \nabla \xi = -\frac{\xi \cdot \nabla f}{f}\xi, \tag{2.15}$$

and we define

$$\kappa = -\frac{\xi \cdot \nabla f}{f} \tag{2.16}$$

as **the surface gravity of the Killing horizon**. We see here that κ measures the deviation from affine parametrisation, that is to say the failure of the Killing parameter v to agree with the affine parameter λ along the null generators of the event horizon. It is easy to check that $f = e^{-\kappa v}$ solves eq.(2.16). Thus we obtain

$$l^\mu = e^{-\kappa v}\xi^\mu, \tag{2.17}$$

which shows the following relation between the Killing parameter v and the affine one λ:

$$\frac{d\lambda}{dv} \propto e^{\kappa v}. \tag{2.18}$$

If $\kappa \neq 0$, we obtain

$$\lambda \propto \frac{e^{\kappa v}}{\kappa}, \tag{2.19}$$

however if $\kappa = 0$ we get an affine relation between λ and v consistently meaning that in the case of a vanishing surface gravity, the null generators along ξ of the horizon are always affinely parametrised. A related phenomenon, controlled by κ as well, plays a very important role in the instability we are aiming to examine; it is the redshifted effect which will be discussed in details in section 3.2.3.

In the case of static black holes, it can be shown that a physical interpretation of κ is to be the limit value of the force exerted at infinity to hold a unit mass in place on the

[2]For further details about the derivations done in this section and about the physical interpretations, the reader should refer to [19], section 12.5.

[3]More precisely its extension to Einstein-Maxwell equations for Reissner-Nordström black holes.

[4]Which, because the horizon is a null hypersurface, are (null) geodesics.

horizon[5], hence earning its denomination *surface gravity of the black hole*. Even if this interpretation does not hold anymore for rotating black holes, we continue to refer to κ as the surface gravity. It can also be proved (see [19]) that κ is constant on the horizon. Indeed, looking at the Reissner-Nordström case, one derives the following expression:

$$\kappa = \frac{r_+ - r_-}{2r_+^2}. \tag{2.20}$$

As a result $\kappa = 0$ for an extreme Reissner-Nordström black hole. This will play a role we will put to light in the next parts of this dissertation. **The vanishing property of the surface gravity can actually be taken as a definition of extremality**. Although we won't proove it here, this definition applies as well for the Kerr family of black holes.

[5]Note however that the locally exerted force diverges on the horizon.

3. Different ways of tackling the instability

3.1 About stability and perturbations

Before going into details about the instability we are willing to analyse, we should discuss a very important point: the concept of stability, closely linked to the perturbation of a system. If we want to be able to grasp the ins and outs of our analysis, this review is necessary. We can distinguish two kinds of stability, and hence of instability: the nonlinear and the linear one.

Recall from the previous chapter the expression of Einstein's vacuum equation:

$$R_{\mu\nu}[g] = 0, \tag{3.1}$$

where $R_{\mu\nu}$ is the Ricci tensor. The true problem of stability of a metric in general relativity, that is to say its **nonlinear stability**, can only be understood after formulating its Cauchy problem, ie that of its dynamics (see [6] for further discussion and references). It turns out to be very complicated to solve but it is the ultimate goal we would be aiming for. So far proving nonlinear stability of a metric has only been achieved in the case of Minkowski space in a monumental monograph by Christodoulou and Klainermann in 1993 [13]. However, it is possible to simplify this problem by linearizing Einstein's equation and hence tackling the question of **linear stability**.

Let ${}^0g_{\mu\nu}$ be an exact solution of eq.(3.1); analysing its stability begins by perturbing ${}^0g$ so we can examine how the perturbation behaves. Considering only small perturbations around this solution, we can write:

$$g_{\mu\nu}(\epsilon) = {}^0g_{\mu\nu} + \epsilon\gamma_{\mu\nu} + O(\epsilon^2), \tag{3.2}$$

where ϵ measures the deviation from ${}^0g$ and $\gamma_{\mu\nu}$ is the perturbation itself, called the *gravitational perturbation*. For small ϵ, $g(\epsilon)$ should be an approximate solution of eq.(3.1)[1]:

$$R_{\mu\nu}[g(\epsilon)] = 0, \tag{3.3}$$

which leads to

$$\underbrace{R_{\mu\nu}[{}^0g]}_{=0} + \epsilon \frac{dR_{\mu\nu}[g(\epsilon)]}{d\epsilon}|_{\epsilon=0} + O(\epsilon^2) = 0. \tag{3.4}$$

[1] See further discussion about *linearization stability*, i.e. the existence of a one-parameter family of exact solutions corresponding to a solution of the linearized equation, in [19] section 7.5.

allowing us to write the linearized Einstein's vacuum equation:

$$\dot{R}_{\mu\nu}[g(\epsilon)] = 0, \tag{3.5}$$

where the dotted notation refers to the derivative with respect to ϵ. This equation is linear in $\gamma_{\mu\nu}$ and can be rewritten as (see [19]):

$$-\frac{1}{2}\nabla_\mu\nabla_\nu\gamma - \frac{1}{2}\nabla^\rho\nabla_\rho\gamma_{\mu\nu} + \nabla^\rho\nabla_{(\nu}\gamma_{\mu)\rho} = 0, \tag{3.6}$$

where $\gamma = {}^0g^{\mu\nu}\gamma_{\mu\nu}$ and all the derivative operators are with respect to ${}^0g_{\mu\nu}$. As shown in [19], one can always choose a gauge (called the *transverse traceless gauge*) where

$$\nabla^\nu\left(\gamma_{\mu\nu} - \frac{1}{2}g_{\mu\nu}\gamma\right) = 0. \tag{3.7}$$

In this gauge, tracing eq.(3.6) yields:

$$\nabla^\rho\nabla_\rho\gamma = 0, \tag{3.8}$$

and after some derivations on eq.(3.6) we obtain the following simple expression for the linearized Einstein's equation:

$$\nabla^\rho\nabla_\rho\gamma_{\mu\nu} - 2R^\rho{}_{\mu\nu}{}^\sigma\gamma_{\rho\sigma} = 0, \tag{3.9}$$

where $R^\rho{}_{\mu\nu}{}^\sigma$ is the Riemann tensor of the metric ${}^0g$. Solving the previous equation entirely answers the question of linear stability of ${}^0g$. However eq.(3.9) involves a very complicated system of partial differential equations which makes its resolution very difficult, except for flat space where the Riemann tensor vanishes. Nevertheless one will note that both eq.(3.8) and eq.(3.9) have a *Klein-Gordon term*. As a consequence, looking at the dynamics of a scalar field ϕ by solving its wave equation

$$\Box_g\phi = 0 \tag{3.10}$$

in a spacetime defined by g is seen as a prerequisite for linear stability of the metric. Such a field is seen as a scalar perturbation of g; showing that the field is linearly stable by **demonstrating that ϕ and all its derivatives are bounded is a proof of linear stability of the metric**. ϕ should be of the most general form (massive, no special symmetry) for the proof to be complete, however it is often easier to start with a field showing some interesting symmetry which will simplify the analysis. Solving eq.(3.10) for a massless and spherically symmetric scalar field has been Aretakis' approach in [3] and [4] and is the one we will review in this section.

As mentioned in the introduction, early prooves for linear stability of the Schwarzschild metric naively used mode analysis and decompostion into spherical harmonics of the scalar field perturbation. In substance, the scalar field ϕ was decomposed as follow:

$$\phi = \sum_{m,l} f_{m,l}(v,r)Y_{m,l}(\theta,\varphi) \tag{3.11}$$

where $Y_{m,l}$ are the standard spherical harmonics. Then the functions $f_{m,l}$ were assumed to be separable in variables $f_{m,l}(v,r) = e^{\lambda_{m,l}v}g_{m,l}(r)$. This separation of variable would

lead to an ordinary differential equation on $g_{m,l}$. If for a given mode $Re(\lambda_{m,l}) > 0$, then the mode would be unstable. These firsts attemps to prove linear stability were incomplete. Firstly one cannot conclude from the stability of each mode that an infinite sum of them remains stable. Secondly and more importantly, it is unclear whether ϕ can be written as a sum of functions separable in variables [7]. It is a very strong and probably wrong assumption. Indeed if a mode was to be unstable then the divergence would be exponential; however, as it will be shown thereafter, the divergence we find for a extreme Reissner-Nordström background is linear in v and not exponential. As a result, a more modern approach was used. The idea is to focus on *energy-type quantities*. We still decompose our field into spherical harmonics but drop the separation of variables. Hence we are left with a nonlinear partial differential equation on $f_{m,l}$ which we aim to solve.

However it should be kept in mind that the scalar field approach can be regarded as a toy model for linearized gravitational perturbations, ie for the attempt to solve eq.(3.9) without any approximation. For the latter, the idea is to use the tetrad formalism and Teukolsky equation so as to reflect important aspect of the spacetime. In [20], J. Lucietti and H.S. Reall have considered such perturbations applied to an extreme Kerr black hole. On the first hand they have generalised Aretakis's result of a linear instability under a massless scalar field perturbation to linearized gravitational perturbations; and on the other hand they have shown that not only extreme Reissner-Norström black holes demonstrate such an instability but that any extreme black hole does[2].

3.2 Physical motivations for an instability

A last important step before revealing this instability with the help of mathematics, analysis and numerical results, is to understand some physical reasons which lead to believe there might be or should be some kind of unstable behaviour in the case of an extreme Reissner-Nordström black hole.

3.2.1 Extremality: an intermediary case

As it was said before, the case of a non-extreme Reissner-Nordström black hole where $Q < M$ has been proved to be stable. However, the case $M < Q$ is excluded by the weak cosmic censorship conjecture which doesn't allow the presence of naked singularities. This can be understood equivalently by saying that naked singularities are dynamically unstable [10]. Considering that the extreme case: $M = Q$, lies between the to latter ones; we can therefore expect it to have both stable and unstable behaviours.

3.2.2 Black holes mechanics and thermodynamics

A key property of an extreme black hole is to have a vanishing surface gravity κ. However, the third law of black holes mechanics states that $\kappa = 0$ cannot be achieved by finitely

[2]This has been proved for scalar perturbations only and not for linearized gravitational ones.

many operations. This statement is the equivalent of the third law of thermodynamic about temperature T; and both make sens together once Stephen Hawking related the temperature of a black hole to its surface gravity with the famous formula:

$$T = \frac{\hbar}{2\pi k_B}\kappa, \tag{3.12}$$

where $\hbar$ and k_B are respectively the (reduced) Planck and Boltzmann constants.

From that perspective we see how extreme black holes become singular. This means there is no way (finite way) to obtain an extreme black hole from a non-extreme one and therefore that a pure extreme black hole would have had to be born extreme. This assertion is not exactly true however and needs a better definition of extremality, using for instance trapped surfaces as W. Israel in [34]. However the appearance of an instability at the boundary of extremality now seems less suprising.

3.2.3 The redshift effect

There are two main redshift effects occuring with black holes.

The first one corresponds to the following situation where an observer A crosses the event horizon while an other observer B does not. Photons sent by A to B will be exponentially shifted to the red over time and B will loose track of A at some point. However this is essentially only a consequence of the fact that it takes a finite proper time for A to cross the event horizon while it takes a infinite proper time for B not to cross it.

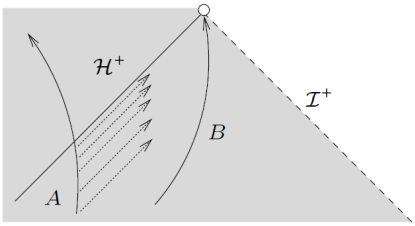


Figure 3.1: Redshift effect between an observer A who crosses the event horizon and B who does not [6].

The second one, which is the one we are interested in, corresponds to the situation where A and B cross the horizon but at different times. In that case, photons sent from A to B will undergo a redshift proportional to $e^{-\kappa v}$ where κ is the surface gravity of the black hole and v the usual ingoing Eddington-Finkelstein time coordinate. Hence we see that for a non-extreme black hole, i.e. $\kappa \neq 0$, these radiations at the horizon will decay whereas for an extreme black hole, i.e. $\kappa = 0$, they will not.

The redshift effect has been shown in [6] to be sufficient to prove the linear stability of a wide range of non-extreme black holes. Because this effect disappears in the cases of extremality, it becomes natural to think of a possible instability occuring then.

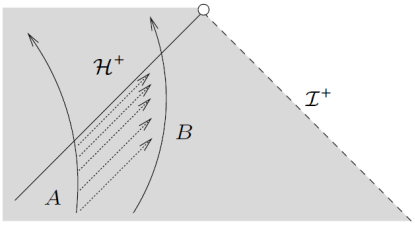


Figure 3.2: Redshift effect between an observer A who crosses the event horizon before B does [6].

3.3 Aretakis' approach

In his approach, Aretakis considers a scalar field perturbation, which is a simplified model for the complete proof of linear stability in comparison with linearized gravitational perturbations. Furthermore, a fundamental aspect of his approach is to analyse the dynamics of the scalar field in a **fixed extreme Reissner-Nordström background**. Our purpose is to examine its evolution as well as that of its derivatives. One can immediately see the limit of this approach. As part of the scalar field will enter the black hole, ie some energy will fall in, the mass M will increase by a tiny bit and hence become superior to the charge Q which means that the black hole will no longer be extreme. However we keep working in a fixed extreme Reissner-Nordström background. This effect will be taken into account in the following section. Although this backreaction on the metric is not dealt with here, Aretakis' approach highlights the main features of an instability occuring on the event horizon.

As Aretakis we will only consider a **massless and spherically symmetric scalar field:** ϕ. Working in the ingoing Eddington-Finkelstein coordinates (v, r), hence $\phi = \phi(v, r)$, we start from the equation of motion, then we derive the existence of a conserved quantity on the horizon: $[\partial_r(r\phi)]_{r=M}$ and finally show that $[\partial_r^2\phi]_{r=M}$ actually blows up with v; hence the instability.

Remember the fixed extreme Reissner-Nordström metric reads:

$$ds^2 = -F(r)dv^2 + 2dvdr + r^2 d\Omega^2, \tag{3.13}$$

with

$$F(r) = \left(1 - \frac{M}{r}\right)^2. \tag{3.14}$$

The equation of motion for a massless scalar field is simply:

$$\nabla^2\phi = 0. \tag{3.15}$$

After some derivations using the spherical symmetry of ϕ we get the following equation:

$$2r\partial_v\partial_r\phi + 2\partial_v\phi + rF(r)\partial_r^2\phi + \big(rF'(r) + 2F(r)\big)\,\partial_r\phi = 0. \tag{3.16}$$

It is very useful to define a quantity for $r\phi$: $\underline{\Phi \equiv r\phi}$; so we can simplify the previous equation:

$$2\partial_v\partial_r\Phi + F(r)\partial_r^2\Phi + F'(r)\left(\partial_r\Phi - \frac{\Phi}{r}\right) = 0. \tag{3.17}$$

Now let's evaluate this equation on the horizon: $r = r_+ = M$. Because F and F' vanish there, we simply get:

$$[\partial_v \partial_r \Phi]_{r=M} = 0, \tag{3.18}$$

thus we can write:[3]

$$[\partial_r \Phi]_{r=M} = constant \equiv H_0 \times M, \tag{3.19}$$

and coming back to $\partial_r \phi$ we obtain:

$$[\partial_r \phi]_{r=M} = H_0 - \frac{\phi(r=M)}{M}. \tag{3.20}$$

Here comes a key result established by Aretakis in [3]: using Morawetz and X estimates, he proved that ϕ, and thus Φ, vanishes on the horizon at late times (see [6] as well for a review). Therefore we can write the following:

$$[\partial_r \phi]_{r=M} \xrightarrow{v\to\infty} H_0. \tag{3.21}$$

For generic initial data, H_0 will be non 0 and $[\partial_r \phi]_{r=M}$ will not vanish over time. This behaviour, or **equivalently the conservation of $\partial_r \Phi$ on the horizon**, is at the core of the instability. As we will see later with numerical simulations on fig.(4.7)(a) and as it has been proved by Aretakis in [3], $\partial_r \Phi$ vanishes at late times outside the horizon, but is still constant on it. Hence the step between $[\partial_r \Phi]_{r=M}$ and $[\partial_r \Phi]_{r=M+\delta}$, $0 < \delta \ll M$, becomes sharper and sharper over time and we expect $\partial_r^2 \Phi$, thus $\partial_r^2 \phi$ as well, to blow up on the horizon.

Now to study the behaviour of $\partial_r^2 \phi$, we derive eq.(3.17) with respect to r, leading to:

$$2\partial_v \partial_r^2 \Phi + F(r)\partial_r^3 \Phi + 2F'(r)\partial_r^2 \Phi + \left(F''(r) - r^{-1}F'(r)\right)\left(\partial_r \Phi - r^{-1}\Phi\right) = 0. \tag{3.22}$$

Evaluating again on the horizon, F and F' vanish, however F'' does not and results in $F''(r=M) = \dfrac{2}{M^2}$, thus:

$$[\partial_v \partial_r^2 \Phi]_{r=M} = -\frac{1}{M^2}\left([\partial_r \Phi]_{r=M} - \frac{\Phi_{r=M}}{M}\right). \tag{3.23}$$

Now using eq.(3.19) and the fact that ϕ, and hence Φ, vanish at late time on the horizon, we obtain:

$$[\partial_v \partial_r^2 \Phi]_{r=M} \xrightarrow{v\to\infty} -\frac{H_0}{M}, \tag{3.24}$$

leading to the behaviour of $[\partial_r^2 \phi]_{r=M}$:

$$[\partial_r^2 \phi]_{r=M} \overset{v\to\infty}{\sim} -\frac{H_0}{M^2} v. \tag{3.25}$$

As expected $\partial_r^2 \phi$ diverges on the horizon over time. We have now shown that the divergence is linear. This prooves the instability. However the case of $H_0 = 0$ (which can be settled by a specific set of initial conditions) should still be dealt with. Indeed in that situation, $\partial_r \phi$ vanishes at late times on the horizon and we will have to look for another conserved

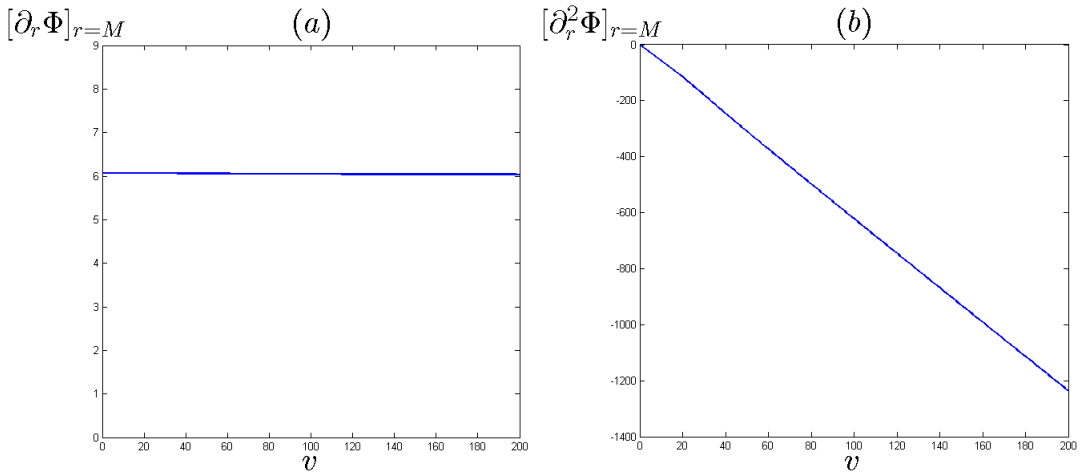


Figure 3.3: **Outgoing wave**: Functions $[\partial_r\Phi](v, r = M)$ and $[\partial_r^2\Phi](v, r = M)$ for an outgoing wave close to the horizon in a fixed extreme Reissner-Nordström background. $(\sigma, \mu) = (0.1, -0.1)$, see section 4 for details about the set up of the initial conditions. Graph (a) confirms that $[\partial_r\Phi]_{r=M}$ is constant on the event horizon which leads to the linear divergence of the second derivative as shown on (b).

quantity different from zero. This investigation will be carried out numerically in section 4.2.1.

Our numerical simulations reproduce the behaviours of eq.(3.19) and eq.(3.24). Information about the algorithm used and the set up of the initial data is given in section 4. However note that for our simulations in the fixed extreme background M is always set equal to 1; hence we can directly deduce from fig.(3.3)(a) that in the case of our simulated outgoing wave: $H_0 \approx 6$. A simple linear fit on the graph from fig.(3.3)(b) gives $y = -6.2x + 1.9$ which confirms our analysis.

As a conclusion, Aretakis' approach highlights an instability when working in a fixed extreme background. This instability relies on a quantity which is conserved on the horizon while decaying outside: $\partial_r\phi$, which leads to the linear divergence of $\partial_r^2\phi$. Let's now consider a more general approach by including the effect of the field on the metric itself: the backreaction.

3.4 Including backreaction on the metric

As mentionned before, this approach is more accurate in the sens it takes into account the effect of the presence of the field on the metric. By definition the black hole goes out of pure extremality when some energy from the field falls in (and remember that non-extreme black holes have been proved to be linearly stable), as a consequence we can expect the instability highlighted in the previous section to be present over a certain amount of time only and then to dissapear as the black hole settles down to a non-extreme solution of

[3]Definition of H_0 as given by Aretakis in [4], propostion 3.0.1.

Einstein-Maxwell's equations. The point here is to decide how to model the effect of the field on the metric. The total charge Q should remain constant thus this is equivalent to decide how to model the variation of the mass M. Our scalar field is viewed as a perturbation of the extreme Reissner-Nordström black hole. Therefore if we define ϵ **as the amplitude of the scalar field** ϕ then our model should be in agreement with the following constraint:

$$M \overset{\epsilon \to 0}{\longrightarrow} Q, \tag{3.26}$$

that is to say we should tend to extremality as we switch the perturbation off. ϵ being the amplitude of our field, a generic set of initial data would perturb the mass by a quantity of order ϵ: $M = Q + O(\epsilon)$. However one could set the field such that its perturbation has a minimum impact on the mass. Now remembering the formula of the stress-energy tensor of a massless scalar field:

$$T_{\mu\nu} = \nabla_\mu \phi \nabla_\nu \phi - \frac{1}{2} g_{\mu\nu} \nabla^\lambda \phi \nabla_\lambda \phi, \tag{3.27}$$

we see it is quadratic in ϕ; thus the field will at least produce a variation of the mass of order ϵ^2. As a consequence we will consider two different choices for M, a second order and first order perturbation of the metric:

$$M = Q + O(\epsilon^2) \tag{3.28}$$
$$M = Q + O(\epsilon) \tag{3.29}$$

with

$$\epsilon \ll Q. \tag{3.30}$$

As we pointed out, the second order perturbation models a milder effect of ϕ on the mass than a first order does. Hence in that case the black hole stays closer to extremality than what it does with an $O(\epsilon)$ perturbation. So we should expect the instability to be stronger with the second order than with the first order.

This can be understood as well if we take a look at the influence on the position of the event horizon, and also at the effect on the surface gravity. Recall that r_+ is given by $r_+ = M + \sqrt{M^2 - Q^2}$ and κ by $\kappa = \frac{r_+ - r_-}{2r_+^2}$ hence we get:

$$\begin{array}{lllll} M = Q + O(\epsilon^2) & \Rightarrow & r_+ = M + O(\epsilon) & \Rightarrow & \kappa = O(\epsilon) \\ M = Q + O(\epsilon) & \Rightarrow & r_+ = M + O(\epsilon^{1/2}) & \Rightarrow & \kappa = O(\epsilon^{1/2}). \end{array} \tag{3.31}$$

As a consequence the shift in the position of the event horizon is less important in the first case than in the second one. Likewise for the surface gravity which controls the redshift effect.

We are not in the extreme case anymore, thus $r_+ \neq r_- \neq M$. Remember as well that the function $F(r)$, which appears in the metric and then in the equation of motion, gets its more general expression: $F(r) = \frac{(r-r_+)(r-r_-)}{r^2}$. Starting again from the equation of motion eq.(3.15), this leads to the same equation eq.(3.17). When we evaluate it on the horizon at $r = r_+$, F vanishes, but contrary to the extreme case F' does not. Here is the core difference due to the fact that now $r_+ \neq r_-$. Therefore we can't state the conservation of $[\partial_r \Phi]_{r=r_+}$ anymore.

From those equations K. Murata, H.S. Reall and N. Tanahashi in [2] have found an interesting way of exposing the role played by the surface gravity κ. Recall that:

$$\begin{array}{lll} \kappa & = & \dfrac{r_+ - r_-}{2r_+^2} \\ \text{and} \quad F'(r) & = & \dfrac{r_+(r-r_-)}{r^3} + \dfrac{r_-(r-r_+)}{r^3}, \\ \text{therefore} \quad F'(r_+) & = & 2\kappa. \end{array} \tag{3.32}$$

Hence when we evaluate equation eq.(3.17) on the horizon we can write:

$$\partial_v[\partial_r\Phi]_{r=r_+} + \kappa\left([\partial_r\Phi]_{r=r_+} - \frac{\Phi_{r=r_+}}{r_+}\right) = 0. \tag{3.33}$$

The trick now is to multiply both sides of the equation by $e^{\kappa v}$. Then the factor κ appears to come from a derivation of the exponential with respect to v and we can simplify equation eq.(3.33) as:[4]

$$\partial_v\left(e^{\kappa v}[\partial_r\Phi]_{r=r_+}\right) = \kappa e^{\kappa v}\frac{\Phi_{r=r_+}}{r_+}. \tag{3.34}$$

In the extreme case: $\kappa = 0$ and we directly recover Aretakis' result about the conservation of $[\partial_r\Phi]_{r=r_+}$. Even better, in [2] the authors have derived a general solution to the previous equation:

$$[\partial_r\Phi](v, r_+) = e^{-\kappa v}[\partial_r\Phi](0, r_+) + \kappa I_1(v), \tag{3.35}$$

where $I_1(v)$ is an integral which depends on κ and $\Phi_{r=r_+}$. The first term is the one that leads to the conservation of $\partial_r\Phi$ on the event horizon in the limit $\kappa \to 0$. Note that Aretakis' approach keeps the metric extreme, i.e. does not perturb it. Hence his approach is worse, in terms of power of the instability, than any choice we might make concerning the order of the perturbation. As a consequence for our derivations, once we find the term(s) which lead us to Aretakis' results when taking $\kappa \to 0$, we know we can consider the other terms as corrections (to a certain order in $\kappa\epsilon$) at least as long as $e^{-\kappa v}[\partial_r\Phi](0, r_+) \approx [\partial_r\Phi](0, r_+)$, ie $v \ll 1/\kappa$. Note that this condition on v is always satisfied when $\kappa = 0$. For late times: $v \to \infty$, we can't state much about those other terms. At best they will remain negligible, at worst they will overcome *Aretakis' term* but we know they won't diverge. Indeed it would not make sens for them to do so at late times because they would appear as a new reason for an instability when we actually know from former proofs that non-extreme black holes are stable.

Thus we write:[5]

$$[\partial_r\Phi](v, r_+) = e^{-\kappa v}[\partial_r\Phi](0, r_+) + O(\kappa\epsilon). \tag{3.36}$$

$[\partial_r\Phi](v, r_+)$ is not constant anymore. It undergoes an exponential decay with a characteristic time $1/\kappa$. Our numerical simulations have reproduced this decay very well as shown on fig.(3.4) and fig.(3.5). Comparing those two, we clearly see that the decay is much faster in the case of an $O(\epsilon)$ perturbation than in that of a $O(\epsilon^2)$ which stays closer to extremality.

[4] It looks like in [2], the authors have forgotten the factor $\dfrac{1}{r_+}$ in equation eq.(3.34)

[5] Refer to [2] for the justifications about the order of correction

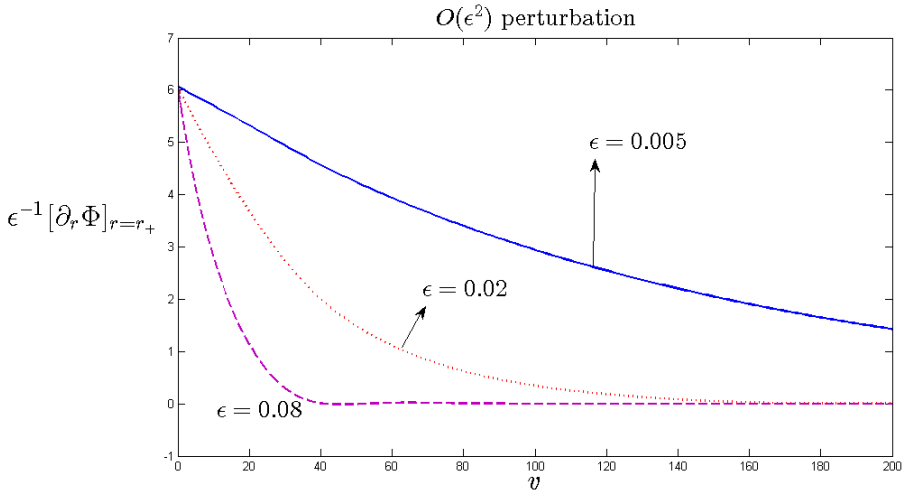


Figure 3.4: Function $\epsilon^{-1}[\partial_r\Phi](v, r = r_+)$ for an outgoing wave close to the horizon in the case of a $O(\epsilon^2)$ perturbation of the metric. $(\sigma, \mu) = (0.1, -0.1)$, see section 4 for details about the algorithm used. Plotting the *normalised* derivative by dividing it by ϵ allows us to make a better comparison of its behaviours when ϵ varies. This is why in all three cases here, the starting point is around 6 as in the extreme case. We see that the bigger ϵ the faster the decay, consistently with the fact the bigger ϵ the further away the black hole is from extremality, hence the milder the instability.

ϵ	0.005	0.02	0.08
κ	0.0070	0.0267	0.0904
κ_{fit}	0.0072	0.0278	0.0829
H_0^{fit}	6.108	6.258	6.248

Table 3.1: Exponential fit of the curves from fig.(3.4) using the expression: $y = H_0^{fit} e^{-\kappa_{fit} v}$. Our fits confirm our previous analysis as the numerical values found for κ are in good agreement with the theoretical ones.

As a consequence of this exponential decay, we can reasonably expect the instability of the second derivative to be present over an amount of time $O(1/\kappa)$ and then to vanish, as it would finally be overwhelmed by the real non-extreme nature of the black hole.

In order to discover the true behaviour of the instability, we obviously have to take a look at the second derivative. Therefore we evaluate once again equation eq.(3.22) on the horizon, then by using the same trick as before to reveal the surface gravity we obtain the following:

$$\partial_v \left(e^{2\kappa v} [\partial_r^2 \Phi]_{r=r_+} \right) = -k e^{2\kappa v} \left([\partial_r \Phi]_{r=r_+} - \frac{\Phi_{r=r_+}}{r_+} \right), \tag{3.37}$$

with

$$k = \frac{5r_- - 3r_+}{2r_+^3}. \tag{3.38}$$

Naturally, going back to the extreme case by taking $\kappa = 0$, we recover Aretakis' result in equation eq.(3.23). With the use of eq.(3.35) we can integrate the previous equation and

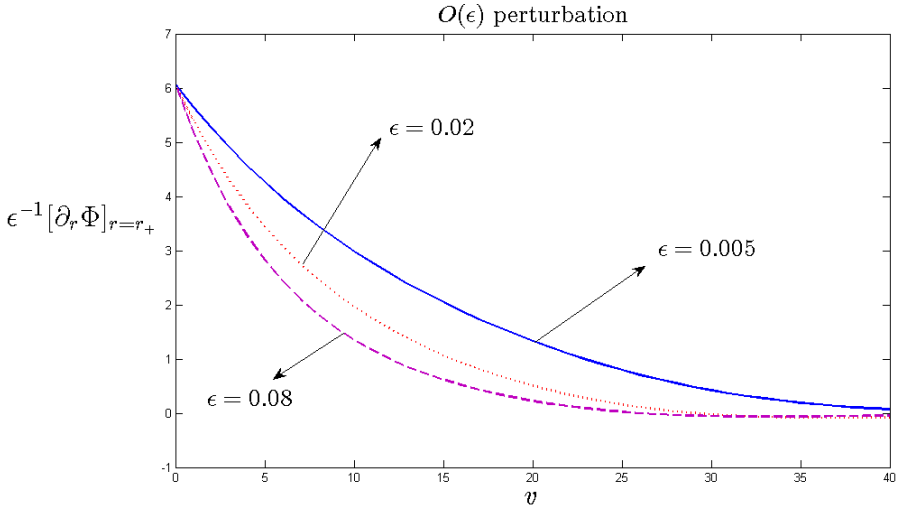


Figure 3.5: The same as fig.(3.4) but this time in the case of a $O(\epsilon)$ perturbation of the metric. We observe that for the same values of ϵ the decay is much faster than for an $O(\epsilon^2)$ perturbation, consistently with the fact that the $O(\epsilon)$ case models a perturbation which stays further from extremality.

ϵ	0.005	0.02	0.08
κ	0.0820	0.1348	0.1843
κ_{fit}	0.0807	0.1251	0.1636
H_0^{fit}	6.269	6.189	6.071

Table 3.2: Exponential fit of the curves from fig.(3.5) using the expression: $y = H_0^{fit} e^{-\kappa_{fit} v}$

get a general expression for $[\partial_r^2\Phi](v, r_+)$. However once again we will only keep the terms we are interested in and consider the others as corrections, at least for early times.

$$[\partial_r^2\Phi](v, r_+) = -\frac{k}{\kappa}\left(e^{-\kappa v} - e^{-2\kappa v}\right)[\partial_r\Phi](0, r_+) + O(\kappa\epsilon). \tag{3.39}$$

Considering only v such that $v \ll 1/\kappa$, we can Taylor expand the exponentials and recover the linear divergence of the second derivative from equation eq.(3.24). Nevertheless, when we look at late times by taking $v \to \infty$, *Aretakis' term* undergoes an exponential decay: the instability ends up vanishing. The contributions of the other terms we have neglected so far are not negligible anymore for a late time analysis. However, eventhough their expression might be complicated (the reader should refer to [2] for more details) we know they will decay for late times. Nonetheless it remains interesting to know when the turning point of the first term, ie of Aretakis' instability, occurs. To do this one only needs to solve

$$\partial_v\left(-\frac{k}{\kappa}\left(e^{-\kappa v} - e^{-2\kappa v}\right)[\partial_r\Phi](0, r_+)\right) = 0, \tag{3.40}$$

which yields $v = ln(2)/\kappa$. This is in agreement with our expectation that the real non-extreme nature of the black hole would express itself after a time of order $1/\kappa$ and wipe out the instability in the end. However $v = ln(2)/\kappa$ does not correspond to early times

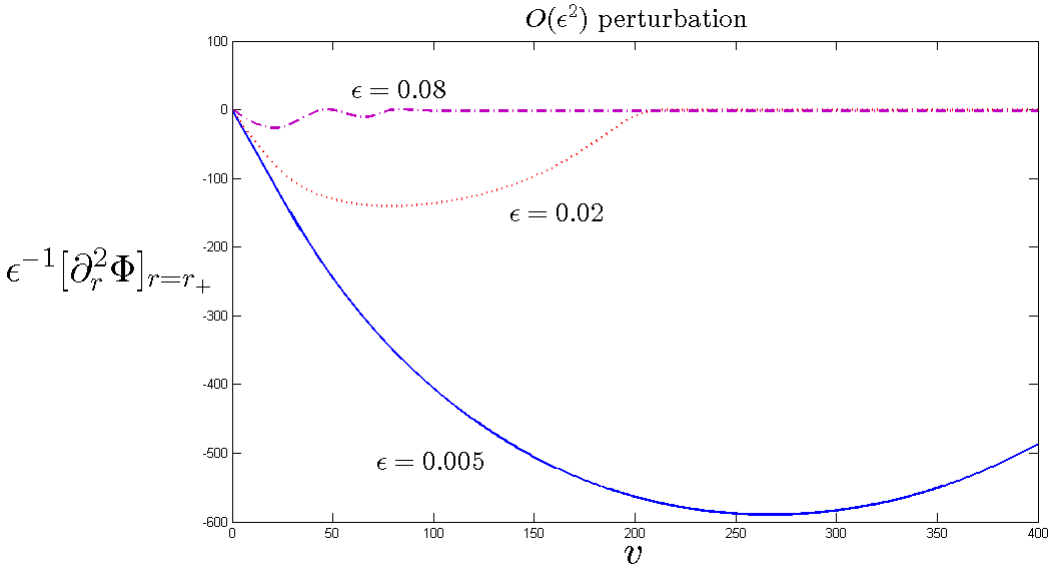


Figure 3.6: Function $[\partial_r^2\Phi](v, r = r_+)$ for an outgoing wave close to the horizon in the case of a $O(\epsilon^2)$ perturbation of the metric. $(\sigma, \mu) = (0.1, -0.1)$, see section 4 for details about the algorithm used. The first part of the curves reproduce Aretakis' linear instability but then undergo a progressive decay starting at a time approximately $O(1/\kappa)$. The wavy part of the curve $\epsilon = 0.08$ between $v = 50$ and $v = 80$ is due to a numerical instability I have been unable to solve.

ϵ	0.005	0.02	0.08
$ln(2)/\kappa$	100	26	8
Abscissa of the inflexion on the curves	270	80	20

Table 3.3: Comparison between the expected times of the inflexion of $[\partial_r^2\Phi](v, r = r_+)$ with the times numerically observed for different values of ϵ. The inflexion occurs later than expected.

when we know Aretakis' term prevails, hence the other terms might contribute at that moment.

On fig.(3.6) the linear divergence for $v \ll 1/\kappa$ is nicely reproduced by our numerical simulations, and then the second derivative starts decaying from $v = O(1/\kappa)$. We would expect the maximum of each curve to be reached at exactly $v = ln(2)/\kappa$, however this is not the case as one can see on tab.(3.3). The inflexion of each curve starts later than expected. This is probably a consequence of the fact that Aretakis' term does not prevail anymore at that stage. Therefore the other terms we have neglected for our analysis should be taken into account. Note as well that the decays on fig.(3.6) are not exponential but rather polynomial. This cannot be explained by considering only the term reproducing the linear instability, but as suggested in [2] this could be an effect of some scattering outside the black hole which would replace the exponential decay by a power law tail.

What about the difference between an second and a first order perturbation on the metric?

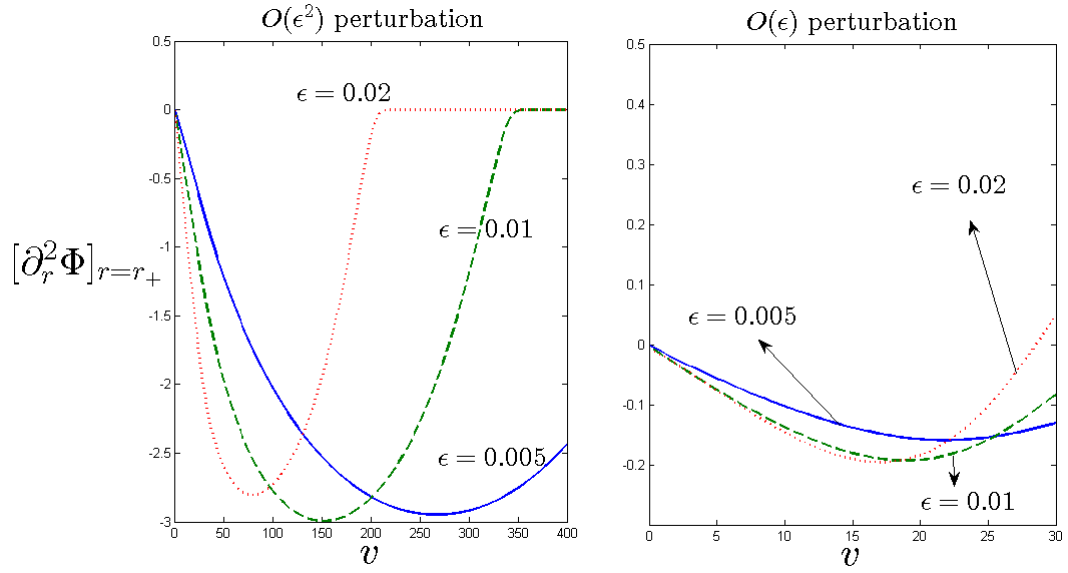


Figure 3.7: Function $[\partial_r^2\Phi](v, r_+)$ for different values of ϵ in both the $O(\epsilon^2)$ and the $O(\epsilon)$ case of the perturbation of the metric. We see that contrary to a second order perturbation, for a first order one the maximum value of $[\partial_r^2\Phi](v, r_+)$ vanishes as $\epsilon \to 0$. Hence there is no instability on the second derivative. Note we have not plotted the curve for $\epsilon = 0.08$ because the errors due a numerical instability are too important -especially in the $O(\epsilon)$ case- for us to draw any conclusions.

Remember the instability in the $O(\epsilon)$ case should be milder than in the $O(\epsilon^2)$ case. As we just prooved, at worst the instability on the second derivative will vanish after a time $O(1/\kappa)$. We can first notice this mildness from eq.(3.31) as $\epsilon^{1/2} < \epsilon$ for small ϵ. Thus the decay of the second derivative for a first order perturbation will occur sooner than for a second order perturbation on the metric. However to see the real difference between these two models, we should take a look at what happens to the maximum of $[\partial_r^2\Phi](v, r_+)$ in the limit $\epsilon \to 0$. Hence we fix $v = v_{max} = ln(2)/\kappa$. Recall that ϵ was originally taken to be the amplitude of the field, thus $[\partial_r\Phi](0, r_+) = O(\epsilon)$. As a result of eq.(3.39) the maximum value of the first term: $[\partial_r^2\Phi](v_{max}, r_+)$, is $O(\epsilon/\kappa)$.

$$\begin{array}{lcll} O(\epsilon^2) \text{ perturbation} & \Rightarrow & \kappa = O(\epsilon) & \Rightarrow \quad [\partial_r^2\Phi](v_{max}, r_+) = O(1) \overset{\epsilon\to 0}{\nrightarrow} 0 \\ O(\epsilon) \text{ perturbation} & \Rightarrow & \kappa = O(\epsilon^{1/2}) & \Rightarrow \quad [\partial_r^2\Phi](v_{max}, r_+) = O(\epsilon^{1/2}) \xrightarrow{\epsilon\to 0} 0. \end{array} \tag{3.41}$$

We see here a major difference between the two orders of perturbation we have chosen. For the $O(\epsilon^2)$ case, the maximum value of $[\partial_r^2\Phi](v, r_+)$ survives the limit $\epsilon \to 0$ which is the proof of an instability. On the contrary we can't even talk of an unstable behaviour of $[\partial_r^2\Phi](v, r_+)$ in the case of an $O(\epsilon)$ perturbation as its maximum value tends to zero when we switch the perturbation off, ie $\epsilon \to 0$. This is shown on fig.(3.7).

Nevertheless it does not mean there is no instability when we perturb the mass by $O(\epsilon)$. To grasp the instability, we have to look at the third derivative on the horizon: $[\partial_r^3\Phi](v, r_+)$.

Using the same processes as before, one can show that it will behave as [2]:[6]

$$[\partial_r^3\Phi](v, r_+) = \frac{kk'}{2\kappa^2}\left(e^{-\kappa v} - 2e^{-2\kappa v} + e^{-3\kappa v}\right)[\partial_r\Phi](0, r+) + O(\epsilon\kappa). \tag{3.42}$$

As we see, the factor in $1/\kappa^2$ will now compensate the $O(\epsilon)$ of $[\partial_r\Phi](0, r_+)$ and the maximum value of the third derivative will survive the limit $\epsilon \to 0$, hence the instability.

As a conclusion, taking into account the backreaction of the field on the metric leads to a deeper understanding of the instability we are facing, firstly hightlighted by Aretakis in a fixed extreme background. His main argument for instability relied on the conservation of $\partial_r\Phi$ on the horizon. We have seen here that a more detailed analysis shows it undergoes a very slow exponential decay, with a caracteristic time inversely proportional to how much the extreme black hole is perturbed: $O(1/\kappa)$. The absence of the redshift effect, controlled by the surface gravity κ, for extreme black holes was one of the main physical reasons why we suspected the presence of an instability. We understand here the role it plays: the sum of outgoing radiations on the horizon have a diverging energy until the redshift effect begins to take effect starting from a time of order $1/\kappa$. In our model's framework $[\partial_r^2\Phi]_{r=r_+}$ diverges during a period of time $O(1/\kappa)$ and then starts decaying, caught up with the real non-extreme nature the black hole acquired by the very presence of the field. We note however some important differences concerning how we model the perturbation of the field on the metric. A weak effect of the field (modeled by an $O(\epsilon^2)$ variation of the mass) recovers Aretakis' analysis about the linear divergence of $[\partial_r^2\Phi]_{r=r_+}$ whereas for a strong effect (modeled by an $O(\epsilon)$ variation of M) one has to look a derivative further, at $\partial_r^3\Phi$, to observe the instability.

[6] $k' = \dfrac{13r_- - 7r_+}{2r_+^3}$

4. Numerical Simulations

In this section, I will describe the code I have used for my numerical simulations and give details about the choices made for the initial data of our Cauchy problem.

4.1 (U,v) coordinates

4.1.1 Motivations

In the previous section, we have entirely worked in the ingoing Eddington-Finkelstein (v, r) coordinate system. It was the right one to use in order to understand on a theoretical point of view the evolution of our scalar field near the black hole. However once we want to simulate this evolution, we have to solve a Cauchy problem; as a consequence we need to choose initial conditions for the scalar field. Remember that the instability we are dealing with is closely related to the redshift effect, as a result we are especially interested in the evolution of an outgoing wave close to the horizon which would be the equivalent of photons evolving on the horizon for instance. Simulating a purely outgoing wave in the (v, r) coordinate system would be very complicated as r is not a null coordinate. Hence we want to switch to the double null coordinate system (u, v). In this coordinate system, we can choose a spacelike hypersurface Σ on which we would specify the initial data for our scalar field ϕ in a way that it simulates either a purely outgoing or ingoing wave.

However in (u, v) coordinates, remember the metric reads:

$$ds^2 = -F\left(r(u,v)\right) du dv + r(u,v) d^2\Omega, \tag{4.1}$$

furthermore on the horizon: $F(r_+) = 0$, thus we see that the metric is singular. This is obviously a major problem since all our analysis is carried out on the horizon. Hence we should look for a different coordinate system. Let's define the coordinate U such that:

$$\frac{u}{2} \equiv -r_*(r_+ - U), \tag{4.2}$$

where r_* is the tortoise coordinate definined in equations eq.(2.12) or eq.(2.13) depending if we are in the extreme or non-extreme case respectively. Combining the previous equation with eq.(2.10) we obtain:

$$du = \frac{2}{F(r_+ - U)} dU, \tag{4.3}$$

which allows us to rewrite the metric in (U, v) coordinates as:

$$ds^2 = -\frac{2F(r)}{F(r_+ - U)} dU dv + r^2 d^2\Omega. \tag{4.4}$$

Using the definition of the Edington-Finkelstein coordinates u and v, and eq.(4.2); we can write:

$$r_*(r) = \frac{v}{2} + r_*(r_+ - U). \tag{4.5}$$

From this we see that $r = r(U, v)$ is analytic in U, v. In the (U, v) coordinate system, the horizon is located at $U = 0$ and $U < 0$ corresponds to the exterior of the back hole. Hence

$$r(U, v) \xrightarrow{U \to 0} r_+ \tag{4.6}$$

leading to

$$\frac{F(r)}{F(r_+ - U)} \xrightarrow{U \to 0} 1; \tag{4.7}$$

therefore the metric defined in (4.4) is analytic on the horizon.

As a consequence we will work in this coordinate system for our numerical simulations. U is a null coordinate, hence we can still define a purely outgoing or ingoing wave on a spacelike hypersurface Σ with:

$$\Sigma = \underbrace{\{U > U_0, v = v_0\}}_{\Sigma_1} \cup \underbrace{\{v > v_0, U = U_0\}}_{\Sigma_2}.$$

An outgoing wavepacket is such that[1]:

$$\Phi(U, v) = \begin{cases} \epsilon \times exp\left(-\dfrac{(U-\mu)^2}{2\sigma^2}\right) & \text{on } \Sigma_1 \\ 0 & \text{on } \Sigma_2 \end{cases} \tag{4.8}$$

An ingoing wavepacket is such that:

$$\Phi(U, v) = \begin{cases} 0 & \text{on } \Sigma_1 \\ \epsilon \times exp\left(-\dfrac{(v-\mu')^2}{2\sigma'^2}\right) & \text{on } \Sigma_2 \end{cases} \tag{4.9}$$

Due to the absence of the redshift effect, outgoing photons on the horizon were assumed from the beginning to be unstable. Thus we are generally more interested into the evolution of outgoing waves as they model the behaviour of outgoing radiations on the horizon. If someone wants to perturb an extreme Reissner-Nordström black hole by dropping some energy in, this would be best modeled by the action of an ingoing wave. Following numerical results show that there is also an instability for such ingoing waves, however milder than for their outgoing counterparts. Indeed we can already observe by comparing fig.(4.2) and

[1]The initial condition here is put on $\Phi = r\phi$ and not the field itself. This doesn't change this outgoing / ingoing feature of the wave, nor the analysis.

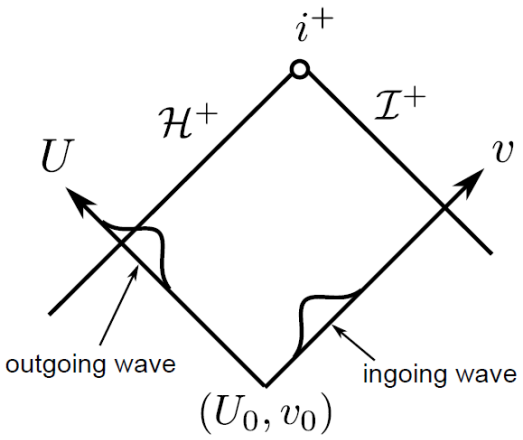


Figure 4.1: Schematic plot of an outgoing and an ingoing wave which we set as initial conditions for our data on a spacelike surface Σ. Figure from [1].

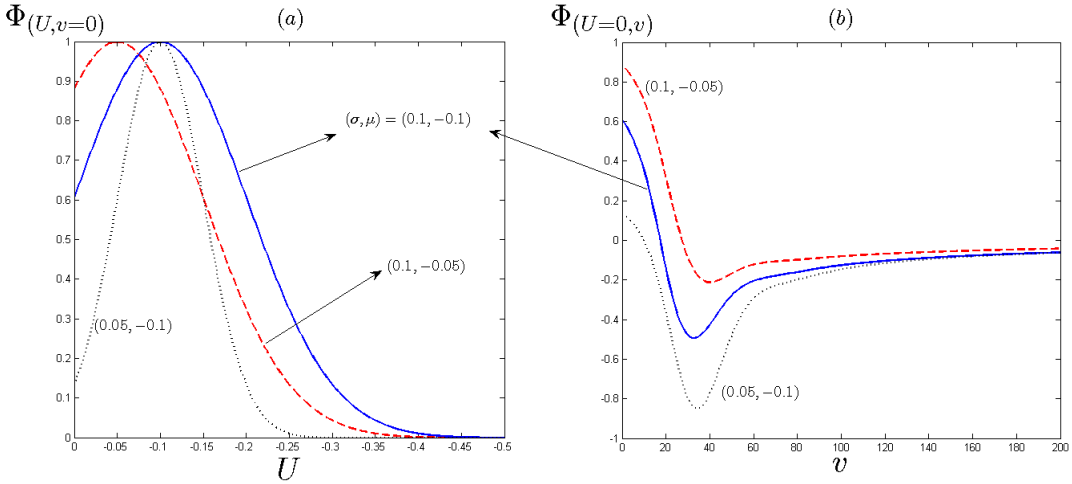


Figure 4.2: **Outgoing wave**: (a) is a plot of the profile of Φ on Σ_1 for different values of (σ, μ). (b) shows the evolution of Φ on the horizon in the extreme case.

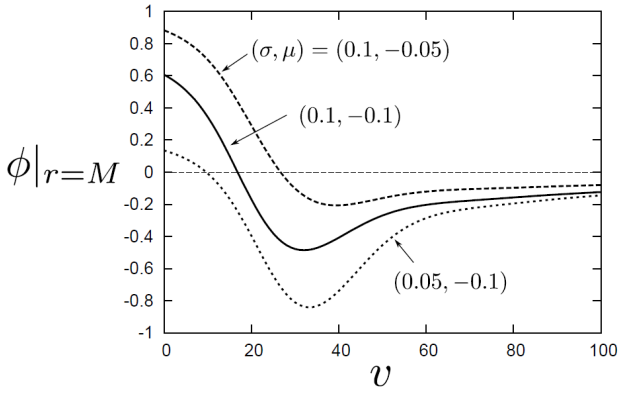


Figure 4.3: Graphic obtained by the authors of [1]. It supports my own numerical simulations when comparing it to fig.(4.2)(b). In [1], ϕ was defined as Φ is defined here.

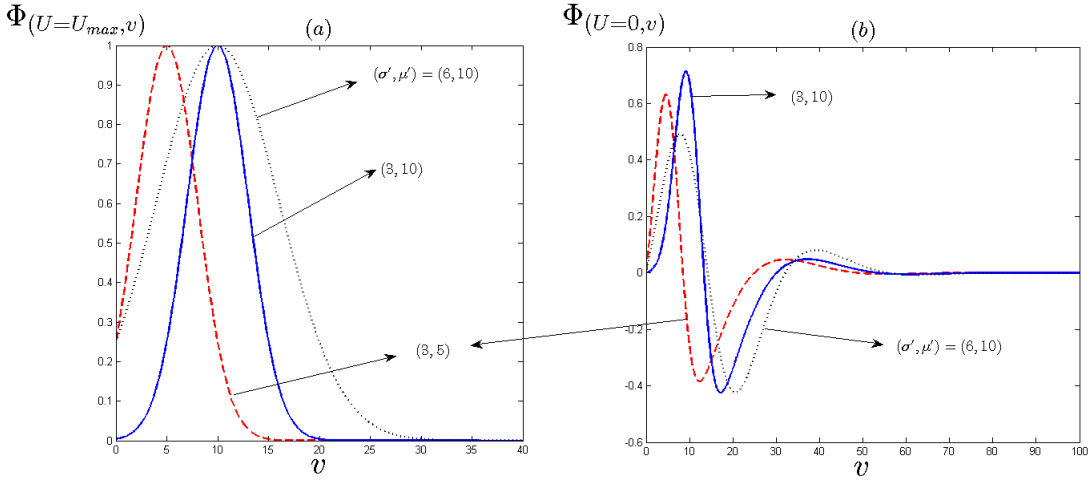


Figure 4.4: **Ingoing wave**: (a) is a plot of the profile of Φ on Σ_2 for different values of (σ', μ'). (b) shows the evolution of Φ on the horizon in the extreme case.

fig.(4.4) that for generic initial data, the scalar field on the horizon vanishes more rapidly for an ingoing wave than for an outgoing one.

In the extreme case $\epsilon = 0$, there is no backreaction of the scalar field on the metric hence it evolves in a fixed extreme Reissner-Nordström spacetime. However it would make no sens to keep ϵ as the amplitude of our wave in the initial data. The perturbation would be null initially, hence null all the time. Therefore the definitions from eq.(4.8) and eq.(4.9) change in the extreme case as we remove ϵ from being the amplitude and just keep a factor 1.

4.1.2 Equations

In the (U, v) coordinate system, the equation of motion eq.(3.15) (hence still in massless and spherically symmetric case) leads to the following:

$$-4\partial_v\partial_U\Phi = \frac{2F(r_+ - U)}{F(r)}\frac{F'(r)}{r}\Phi, \tag{4.10}$$

and we define

$$V(U,v) \equiv \frac{2F(r_+ - U)}{F(r)}\frac{F'(r)}{r} \tag{4.11}$$

as the potential related to ϕ. Here comes another motivation to switch to this coordinate system: eq.(4.10) is much simpler to solve numerically than eq.(3.17). Indeed it only involves first order derivatives with respect to the same variable, as a consequence we can use an explicit method to compute Φ in the future of the hypersurface Σ on which we have set our initial conditions. Recall again that in the extreme case: $F'(r_+) = F'(M) = 0$. As a result, eq.(4.10) directly leads to the conservation of $\partial_U\Phi$ on the horizon, located at $U = 0$. We once again recover Aretakis' result from eq.(3.19).

4.2 Outline of the code

All along our simulations we set $Q = 1$. As a result the horizon is always located at $r = M = Q = 1$ in the extreme case, around 1 in the non-extreme one. The two orders of perturbation of the extreme metric: $O(\epsilon^2)$ and $O(\epsilon)$ should be compared to 1 and we model them in the simple following way:

$$O(\epsilon^2) \quad \Rightarrow \quad M = 1 + \epsilon^2 \tag{4.12}$$

$$O(\epsilon) \quad \Rightarrow \quad M = 1 + \epsilon. \tag{4.13}$$

The point of our code is to solve eq.(4.10) starting from our initial data. Hence we work on a (U, v) grid delimited by $v_0 = 0$, $U_0 = -0.5$. This initial value for U is justified by the fact that $r(-0.5, 0) = 1.5$. Hence our waves are initially set far enough from the horizon. v_{max} will vary depending on the case we are analysing but U_{max} will be set to 0 where the horizon is. [2]

From eq.(4.10) we see that we need to compute $r(U, v)$ everywhere in order to be able to calculate $V(U, v)$ and then solve our Cauchy problem. So as to do this, we differentiate equation eq.(4.5) and use eq.(2.10) to get:

$$F(r)^{-1}dr = \frac{dv}{2} - F(r_+ - U)^{-1}dU. \tag{4.14}$$

Thus at constant U:

$$\frac{\partial r}{\partial v}_{|U} = \frac{F(r)}{2}. \tag{4.15}$$

To be able to integrate this equation, we need to know $r(U, v)$ initially. This is actually the case: evaluating eq.(4.5) at $v = 0$ we simply obtain: $r(U, 0) = r_+ - U$. Hence we can integrate eq.(4.15) on each $U = constant$ line and then get $r(U, v)$ everywhere as shown on fig.(4.5). Thereafter we can easily compute $V(U, v)$ on the whole domain and then solve eq.(4.10).

Note that eq.(4.15) does not depend on the tortoise coordinate r_*. Thus we can integrate it in the same way for both the extreme and non-extreme case, which is quite remarkable. As a result, the code used to simulate the backreaction on the metric is exactly the same as the one we use for the fixed extreme background but setting $\epsilon \neq 0$ and defining it as the amplitude of Φ.

As mentioned earlier, the method carried out to solve eq.(4.10) numerically is explicit. After discretising our (U, v) grid on both axis and setting the initial wave data on Σ, we compute the numerical value of Φ one point after the other thanks to an algorithm derived from eq.(4.10). The latter is very similar to the one used in [1] and is described here in appendix A. Some results of our explicit method are plotted on fig.(4.6).

[2] Actually almost 0 only because even though equation eq.(4.10) is by construction well defined on the horizon, numerically it poses a problem to compute $\frac{1}{F(r)}$ there as $F(r)$ vanishes at $r = r_+$. Hence we set $U_{max} = -10^{-6}$.

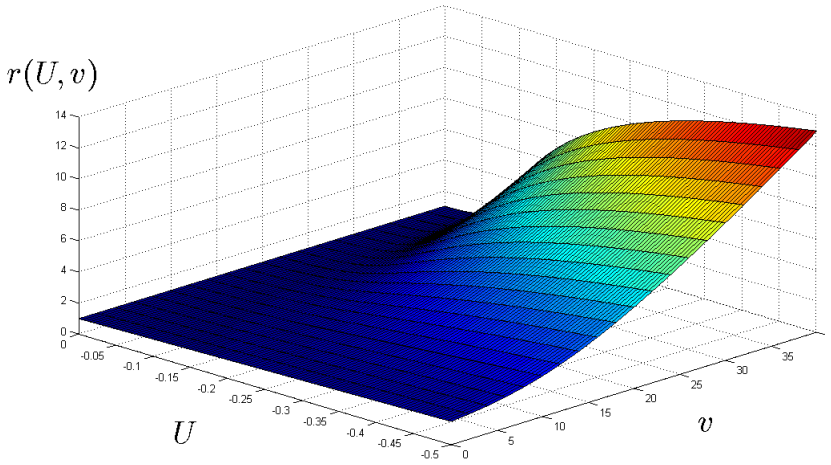


Figure 4.5: Evolution of $r(U, v)$ on the grid in the extreme case: $r_+ = M = 1$. We see that on the horizon at $U = 0$, $r(0, v) = r_+ = 1$. As well, initially $r(U_0, v_0) = 1.5$, however r increases very rapidly with v outside the event horizon.

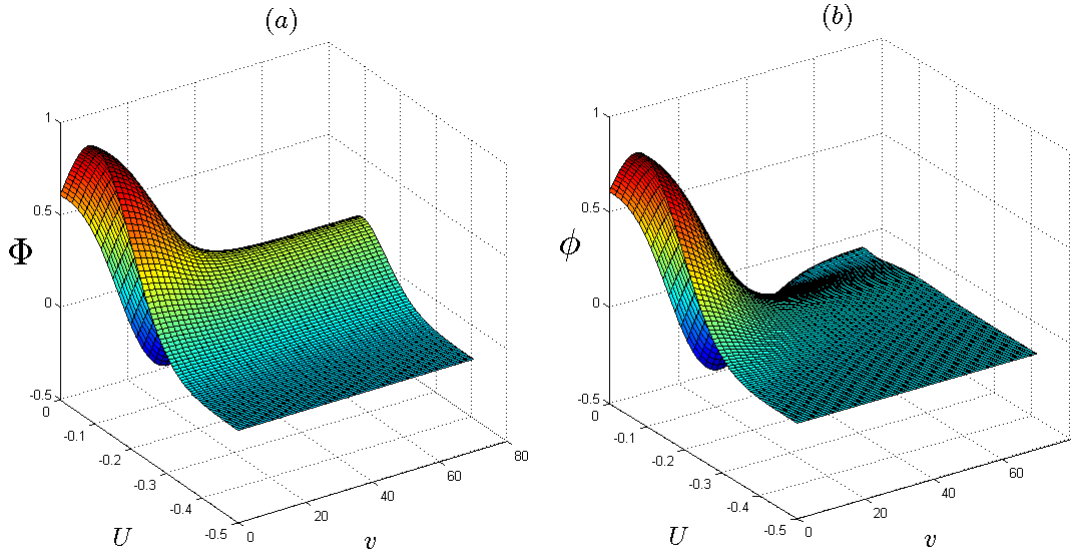


Figure 4.6: Plots of $\Phi(U, v)$ and of the scalar field $\phi(U, v) = \dfrac{\Phi}{r}$ for an outgoing wave with $(\sigma, \mu) = (0.1, -0.1)$ in a fixed extreme background. We observe on (b) that ϕ vanishes on and outside the event horizon, which was a main result proved by Aretakis in [3] that we exploited to exhibit the instability.

Once we have computed the field over the whole domain, it is easy to calculate its derivatives with respect to U. We are however interested in the r-derivatives of the field. In the (v, r) coordinate system, $\partial_r \Phi$ is the variation of the field in the r direction **at constant** v. Hence we get the following relation:

$$\partial_r \Phi = \frac{\partial U}{\partial r}|_v \partial_U \Phi \tag{4.16}$$

Considering eq.(4.14) at constant v, we get:

$$\frac{\partial U}{\partial r}|_v = -\frac{F(r_+ - U)}{F(r)} \tag{4.17}$$

hence

$$\partial_r \Phi = -\frac{F(r_+ - U)}{F(r)} \partial_U \Phi. \tag{4.18}$$

Note that on the horizon: $\partial_r \Phi = -\partial_U \Phi$. Likewise for the second derivative with respect to r:

$$\partial_r^2 \Phi = \partial_r \left(-\frac{F(r_+ - U)}{F(r)} \partial_U \Phi \right) = \left(\frac{F(r_+ - U)}{F(r)} \right)^2 \partial_U^2 \Phi + \frac{F'(r) - F'(r_+ - U)}{F(r)} \partial_U \Phi. \tag{4.19}$$

Thus using eq.(4.18) we are able to plot $\partial_r \Phi(U, v)$. On fig.(4.7), we can clearly examine the differences between keeping the extreme background fixed or including backreaction on the metric. In the first case $\partial_r \Phi(U, v)$ is constant on the horizon but vanishes outside which yields the divergence of the second derivative; in the second case $[\partial_r \Phi]_{r=r_+}$ undergoes a slow exponential decay which will turn the instability of $[\partial_r^2 \Phi]_{r=r_+}$ off.

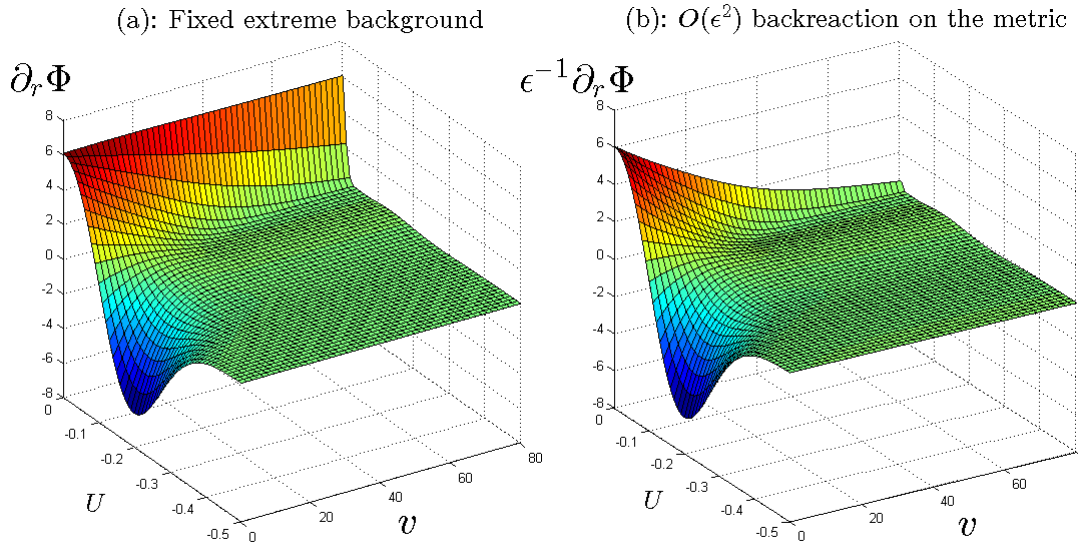


Figure 4.7: Plots of $\partial_r \Phi(U, v)$ in the extreme case and of $\epsilon^{-1} \partial_r \Phi(U, v)$ in the non extreme one ($\epsilon = 0.02$) for an outgoing wave with $(\sigma, \mu) = (0.1, -0.1)$. We can clearly see on (a) that $\partial_r \Phi$ is constant on the horizon at $U = 0$, but because it vanishes outside it becomes steeper and steeper; hence $\partial_r^2 \Phi$ blows up on the horizon. On the contrary on (b), when the backreaction is included, $\partial_r \Phi(U, v)$ decays slowly on the horizon and after some time it suffices to contain the instability on the second derivative.

4.3 Case $H_0 = 0$ in a fixed extreme background

We will approach the understanding of this case thanks to some simulations. A theoretical approach is obviously possible but will not be delt with here; the reader should refer to [1].

When $H_0 = 0$, eq.(3.24) tells us that $[\partial_r^2\Phi]_{r=M}$ tends to be constant as $v \to \infty$. However we also know that $\partial_r^2\Phi$ vanishes right outside the horizon, hence we face the same problem as before with $\partial_r\Phi$ but with $\partial_r^2\Phi$ now : the variation between the value on the horizon and that right outside becomes steeper and steeper. We should expect the same instability as before but a derivative further, that is to say on $[\partial_r^3\Phi]_{r=M}$.

The case $H_0 = 0$ can be approached for both outgoing and ingoing wave. For an outgoing wave to have $H_0 = 0$, i.e. $[\partial_r\Phi]_{r=M} = 0$, one either has to set $\mu = 0$ or set a large value for it such that the tail of the gaussian profile is already very flat on the horizon. In the case of an ingoing wavepacket, $[\partial_r\Phi]_{r=M}$ never gets a very high value compared to the outgoing situtation. Indeed because its inital profile is set on $U = constant$ line, an ingoing wave initially varies only a little with respect to U and hence r, thus $[\partial_r\Phi](0, v)$ is close to 0 and so is H_0.

Fig.(4.8) illustrates this case for both kinds of wavepackets and shows the divergence of the third derivative which prooves the instability. The scheme of the instability is easily understandable from that figure. Indeed note that, for the initial data we have chosen, $[\partial_r^2\Phi]_{r=M}$ tends to a positive constant value for the ingoing wavepacket, a negative constant for the outgoing one. This leads to a positive slope in the divergence of $[\partial_r^3\Phi]_{r=M}$ for an ingoing wave and a negative slope in the other case. If we were to find even more specific initial data such that $[\partial_r^2\Phi]_{r=M}$ was to tend towards a constant value equal to 0, then $[\partial_r^3\Phi]_{r=M}$ would not diverge but be constant and there is little doubt that the instability would be on the fourth derivative: $[\partial_r^4\Phi]_{r=M}$.

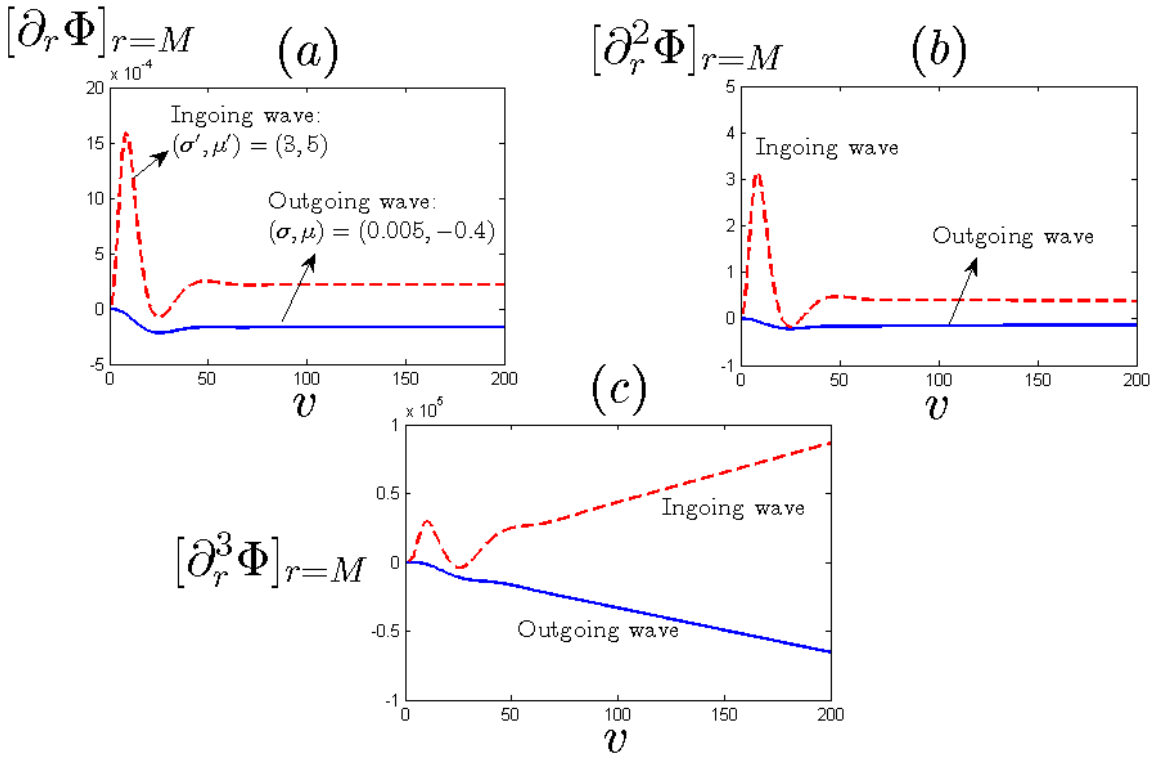


Figure 4.8: Outgoing wavepacket with $(\sigma, \mu) = (0.005, -0.4)$ and ingoing wavepacket with $(\sigma', \mu') = (3, 5)$. For both cases, we see on (a) that H_0 is very close to 0, hence on (b) we observe that $[\partial_r^2 \Phi]_{r=M}$ tends to a constant value and does not diverge as precedently. Even if that value is quite small, we can see on (c) that $[\partial_r^3 \Phi]_{r=M}$ diverges very quickly. Note that although $[\partial_r^2 \Phi]_{r=M}$ is closer to 0 for the outgoing wave than for the ingoing one, the divergence on (c) is of the same strength. This is due to the fact that outgoing waves model radiations on the horizon which are very unstable in the absence of the redshift effect.

5. Conclusion

Black holes are probably the most peculiar physical objects engendered by Einstein's theory of general relativity. Amongst them extreme black holes form an even more singular family as they sit between unstable naked singularities and non-extreme spacetimes known to be stable. Furthermore their key defining property of vanishing surface gravity leads to the absence of the redshift effect which is a major mechanism proving the stability of their non-extreme counterparts. Their study, beyond pure intellectual curiosity we all share, is motivated by the existence of almost-extreme astrophyiscal black holes and our desire to learn more about them.

In the context of linear stability, Stephanos Aretakis initiated this analysis and proved the existence of a somewhat subtle and intriguing instability occuring on the event horizon of an extreme Reissner-Nordström black hole perturbed by a massless and spherically symmetric scalar field ϕ. Using a major result of his: the decay of ϕ on the horizon at late times, we could prove the conservation of $\partial_r\phi$ on the horizon, and then show the linear divergence of $[\partial_r^2\phi]_{r=M}$. However, a more detailed analysis including the backreaction of the scalar field on the metric tells us that the instability is only present during a time of order $1/\kappa$ before decaying. We underline the importance of the model chosen for the backreaction: weak variations of the mass parameter recover Aretakis' result, whilst stronger variations require looking at the third derivative $\partial_r^3\phi$ to observe the instability. Using an appropriate change of coordinates, $(v, r) \to (U, v)$, I have been able to reproduce numerically these behaviours with good accuracy. I nevertheless encountered a numerical instability due to a very high sensitivity of the potential on the horizon with respect to variation of the ϵ parameter. Further investigation regarding its relation with the gap which exists between extremality and subextremality would be of interest.

In [1], the authors have extended Aretakis' result to massive or non-spherically symmetric scalar fields. Purely looking at the near-horizon geometry of an extreme Reissner-Nordström black hole, using $AdS_2 \times S^2$ spacetime, they have analysed the dynamic of the l^{th} multipole of a massless scalar field ϕ, and proved the conservation of $\partial_r^{l+1}\phi$ on the horizon, hence leading to the divergence of $\partial_r^{l+2}\phi$ at late times. Likewise, they have demonstrated that a massive and spherically symmetric scalar field exhibits similar conserved quantities and thus similar instability. Another improvement was accomplished in [20] where Aretakis' result has been generalised to *any* extreme black hole perturbed by a massless and spherically symmetric scalar field.

Nonetheless, it should be kept in mind that the scalar perturbation approach appears as a

simplified model for linearised gravitational perturbations. Therefore it is worth emphasising once again the result achieved in [20], where, by considering such perturbations, the authors have managed to give a true proof for linear instability of an extreme Reissner-Nordström black hole. This result brings us a step closer to solving the complete nonlinear problem steming from Einstein-Maxwell's equations, our ultimate goal in the end.

Appendices

A. Algorithm used

Here's a description of the explicit method used to solve eq.(4.10). Remember we first have to integrate eq.(4.15) so as to be able to compute $V(U, v)$ on the whole domain. Let's h_U and h_v be the resolution intervals of the grid along U and v. Therefore $h_U = U_{i+1} - U_i$ and $h_v = v_{j+1} - v_j$. Defining then $\Phi(U_i, v_j) = \Phi_{i,j}$ we can write the left hand term of eq.(4.10) as follow:

$$\partial_v \partial_U \Phi = \frac{\Phi_{i+1,j+1} - \Phi_{i,j+1} - \Phi_{i+1,j} + \Phi_{i,j}}{h_U h_v} + O(h_U^3 h_v, h_U h_v^3) \tag{A.1}$$

However this double derivative does not "live" on a point (i, j) of the grid, it lives in the middle of the square formed by the four points: (i, j), $(i + 1, j)$, $(i, j + 1)$, $(i + 1, j + 1)$. Hence we cannot naively equate it to $V_{i,j}\Phi_{i,j}$. We have to take its average over the four points. Considering only the first order for the derivatives, eq.(4.10) becomes:

$$-4\frac{\Phi_{i+1,j+1} - \Phi_{i,j+1} - \Phi_{i+1,j} + \Phi_{i,j}}{h_U h_v} = \frac{V_{i+1,j+1}\Phi_{i+1,j+1} + V_{i+1,j}\Phi_{i+1,j} + V_{i,j+1}\Phi_{i,j+1} + V_{i,j}\Phi_{i,j}}{4} \tag{A.2}$$

The point of our explicit method is to express $\Phi_{i+1,j+1}$ in terms of all the other variables, hence enabling us to compute one point after the other the value of the field solution to the equation of motion. Solving the above equation for $\Phi_{i+1,j+1}$ we obtain:

$$\Phi_{i+1,j+1} = \left(1 + \frac{h_U h_v}{16} V_{i+1,j+1}\right)^{-1} \left(\Phi_{i,j+1} + \Phi_{i+1,j} - \Phi_{i,j} - \frac{h_U h_v}{4} \frac{V_{i+1,j}\Phi_{i+1,j} + V_{i,j+1}\Phi_{i,j+1} + V_{i,j}\Phi_{i,j}}{4}\right) \tag{A.3}$$

Note that it can be interesting if not necessary to use an adaptative grid to have a better resolution on the U axis. Indeed, the potential $V(U, v)$ tends to be extremely sharp as $U \to 0$ as shown on fig.(A.1). The computation of $\partial_r \Phi$ does not suffer much from a rather poor resolution, but that of $\partial_r^2 \Phi$ does, especially in the non-extreme case. Even though I have refined my grid closer to the horizon, it doesn't look like I have been able to get rid of this numerical instability. The latter becomes even more important when dealing with the case of a first order perturbation of the metric.

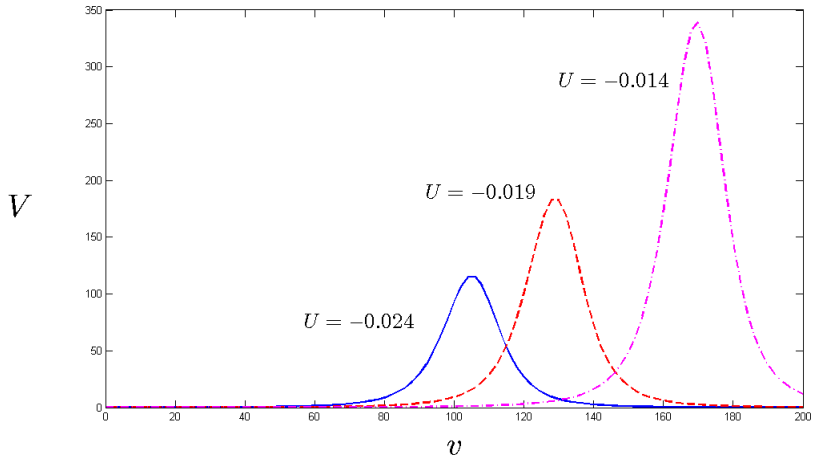


Figure A.1: Plot of the potential V over v for different fixed values of U close to the horizon, located at $U = 0$, in the extreme case. Even though the three values of U considered for the plots are very close to each other, the potential profile increases a lot which demonstrates a high sharpness. Therefore it becomes necessary to consider a more refined grid on the U-axis in order to get a better resolution and avoid numerical errors.

Bibliography

[1] J. Lucietti, K. Murata, H.S. Reall, N. Tanahashi, "On the horizon instability of an extreme Reissner-Nordström black hole", (2013) [arXiv:1212.2557v2 [gr-qc]]

[2] K. Murata, H.S. Real, N. Tanahashi, "What happens at the horizon(s) of an extreme black hole?", (2013), [arXiv:1307.6800v1 [gr-qc]]

[3] S. Aretakis, "Stability and instability of extreme Reissner-Nordström black holes spacetimes for linear scalar perturbations I", (2011) [arXiv:1110.2007v1 [gr-qc]]

[4] S. Aretakis, "Stability and instability of extreme Reissner-Nordström black holes spacetimes for linear scalar perturbations II", (2011) [arXiv:1110.2009v1 [gr-qc]]

[5] S. Aretakis, "The Wave Equation on Extreme Reissner-Nordström Black Hole Spacetimes: Stability and Instability Results", (2011) [arXiv:1006.0283v2 [math.AP]]

[6] M. Dafermos, I. Rodnianski, "Lectures on black holes and linear waves", (2008) [arXiv:0811.0354v1 [gr-qc]]

[7] M. Dafermos, I. Rodnianski, "The black hole stability problem for linear scalar perturbations", (2010) [arXiv:1010.5137v1 [gr-qc]]

[8] M. Dafermos, I. Rodnianski, Y. Shlapentokh, "Decay for solutions of the wave equation on Kerr exterior spacetimes III: the full subextremal case $|a| < M$", (2014) [arXiv:1402.7034v1 [gr-qc]]

[9] M. Dafermos, I. Rodnianski, "The red-shift effect and radiation decay on black hole spacetimes", (2008) [arXiv:gr-qc/0512119v1]

[10] M. Dafermos, "The evolution problem in general relativity, Current developments in mathematics", 2008, 1-66, Int. Press, Somerville, MA

[11] S. Dain, G. Dotti, "The wave equation on the extreme Reissner-Nordström black hole", (2013) [arXiv:1209.0213v4 [gr-qc]]

[12] J. Ybarra, PHZ 6607,
http://www.phys.ufl.edu/courses/phz6607/fall08/JY_black_hole_stability.pdf

[13] D. Christodoulou and S. Klainerman, "The global nonlinear stability of the Minkowski space", Princeton University Press, (1993)

[14] G. Cruciani, "Deriving the Regge-Wheeler and Zerilli equations in the general static spherically symmetric case with MathematicaTM and MathTensorTM(*)", (2000) http://www.icra.it/Publications/Scientific/Files/ncb8653.pdf

[15] T. Regge, J.A. Wheeler, "Stability of a Schwarzschild Singularity", (1957), Physical Review volume 108 number 4

[16] L.A. Edelstein, C.V. Vishveshwara, "Differential Equations for Perturbations on the Scharwzschild Metric", *Physical Review D*, Volume 1 Number 12, (1970) PhysRevD.1.3514

[17] R.M. Wald, "Note on the stability of the Schwarzschild metric", Journal of Mathematical Physics 20, 1056 (1979); DOI:10.1063/1.524181

[18] B.S. Kay, R.S. Wald, "Linear stability of Schwarzschild under perturbations which are non-vanishing on the bifurcation 2-sphere", (1986) doi:10.1088/0264-9381/4/4/022

[19] R.M. Wald, "General Relativity", book, (1984)

[20] J. Lucietti, H.S. Reall, "Gravitational instability of an extreme Kerr black hole", (2012) [arXiv:1208.1437v3 [gr-qc]]

[21] Y. Choquet-Bruhat, R. Geroch, "Global Aspects of the Cauchy Problem in General Relativity", (1969)

[22] H.F. Rúnarsson, "Limits and special symmetries of extremal black hole", Master thesis in theoretical physics, Stockholm University, (2012)

[23] G. Risaliti, F.A. Harrison, K.K. Madsen, D.J. Walton, S.E. Boggs, F.E. Christensen, W.W. Craig, B.W.Grefenstette, C.J. Hailey, E. Nardini, Daniel Stern, W.W. Zhang, "A rapdily spinning black hole at the center of NGC1365", (2013), Volume 494, Nature

[24] M.A Abramowicz, P.C Fragile, "Foundation of black hole accretion disk theory", (2013), [arXiv:1104.5499v3 [astro-ph.HE]]

[25] Jeffrey E. McClintock, Ramesh Narayan1, Shane W. Davis, Lijun Gou, Akshay Kulkarni, Jerome A. Orosz, Robert F. Penna, Ronald A. Remillard, James F. Steiner, "Measuring the spins of accreting black holes", (2011) [arXiv:1101.0811v2 [astro-ph.HE]]

[26] Rebecca Shafee, Jeffrey E. McClintock, Ramesh Narayan, Shane W. Davis, Li-Xin Li, Ronald A. Remillard, "Estimating the Spin of Stellar-Mass Black Holes via Spectral Fitting of the X-ray Continuum", (2005) [arXiv:astro-ph/0508302v2]

[27] J.M. Miller, C.S. Reynolds, A.C. Fabian, G. Miniutti, L.C. Gallo, "Stellar-mass black hole spin constraints from disk reflection and continuum modeling", (2013) [arXiv:0902.2840v1 [astro-ph.HE]]

[28] W.M. Farr, N. Sravan, A. Cantrell, L. Kreidberg, C.D. Baylin, I. Mandel, V. Kalogera, "The Mass Distribution of Stellar-Mass Black Holes", (2011) [arXiv:1011.1459v2 [astro-ph.GA]]

[29] The European Southern Observatory: ESO, www.eso.org/public/videos/eso1332a/

[30] Sean A. Farrell1, Natalie A. Webb, Didier Barret, Olivier Godet Joana M. Rodrigues, "An Intermediate-Mass Black Hole of $\geq$ 500 Solar Masses in the Galaxy ESO 243-49", *Nature*, (2009) [arXiv:1001.0567 [astro-ph.HE]]

[31] J.E. McClintock, R.A. Remillard, "Black hole binaries", (2004) [arXiv:astro-ph/0306213v4]

[32] D.L. Meier, "The Theory and Simulation of Relativistic Jet Formation: Towards a Unified Model For Micro- and Macroquasars", (2003) [arXiv:astro-ph/0312048v1]

[33] S.J. van Tongeren, "Rotating Black Holes", (2009)

[34] W. Israel, Phys. Rev. Lett. 57, 397 (1986).