Proof of the area–angular momentum–charge inequality for axisymmetric black holes

2013 Class. Quantum Grav. 30 065017

(http://iopscience.iop.org/0264-9381/30/6/065017)

IOP Publishing Classical and Quantum Gravity

Class. Quantum Grav. **30** (2013) 065017 (29pp) doi:10.1088/0264-9381/30/6/065017

Proof of the area–angular momentum–charge inequality for axisymmetric black holes

María E Gabach Clement, José Luis Jaramillo and Martín Reiris

Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, D-14476 Golm, Germany

E-mail: gabach@aei.mpg.de, Jose-Luis.Jaramillo@aei.mpg.de and martin.reiris@aei.mpg.de

Received 28 August 2012, in final form 4 February 2013
Published 1 March 2013
Online at stacks.iop.org/CQG/30/065017

Abstract

We give a comprehensive discussion, including a detailed proof, of the area–angular momentum–charge inequality for axisymmetric black holes. We analyze the inequality from several viewpoints, in particular including aspects with a theoretical interest well beyond the Einstein–Maxwell theory.

PACS numbers: 04.70.Bw, 04.40.Nr, 04.20.Cv

1. Introduction

The main result of this paper is the following theorem.

Theorem 1.1. *Let S be either*

1. *a smooth stable axisymmetric marginally outer trapped surface (MOTS) embedded in a spacetime, satisfying the dominant energy condition, with the non-negative cosmological constant Λ, angular momentum J, charges Q_E and Q_M and area A, or,*
2. *a smooth stable axisymmetric minimal surface in a maximal data set, with non-negative scalar curvature, with angular momentum J, charges Q_E and Q_M and area A and non-negative Λ.*

   Then,

$$A^2 \geqslant 16\pi^2\left[4J^2 + \left(Q_E^2 + Q_M^2\right)^2\right]. \quad (1)$$

Moreover, the equality in (1) is achieved if and only if the surface is the extreme Kerr–Newman sphere (see section 3.2).

This type of relation among physical parameters of black holes plays a relevant role in the context of the standard picture of classical gravitational collapse [41]. In this sense, the works of Penrose (see the review [37] on Penrose inequality) offer a paradigmatic example with the proposal of a lower bound for the total mass in terms of the size (area) of the black hole in the form $m^2 \geqslant A/16\pi$. The efforts to formulate similar geometrical inequalities incorporating the angular momentum of the black hole have led to two different lines of research. The first one, started in [19] (see also [18]), followed in [16, 17] and recently extended and improved in [45], is of global nature and provides a lower bound to the total mass in terms of the angular momentum and charges in a vacuum black hole spacetime

$$m^2 \geqslant \frac{|J|^2}{m^2} + Q_{\rm E}^2 + Q_{\rm M}^2. \tag{2}$$

The second line of research leads to inequality (1), which presents a quasilocal character in the sense that only the geometry on a closed surface is involved in the analysis.

The first explicit lower bounds for the area solely in terms of black hole physical parameters, including the angular momentum, were given in [30, 31, 5] (see also [6]) in stationary black holes, and later in [1, 20, 34] within dynamical scenarios (see also [24, 21, 46, 33, 25, 35] and the review article [22] on the subject).

In the recent article [25], a first straightforward approach to prove inequality (1) in the dynamical case was presented. It consists in matching the variational problem discussed in [31] for the stationary axisymmetric case with the dynamical quasilocal treatment in [34] (see also [15, 38] for further clarification on the relation between the stationary and the dynamical quasilocal approaches). More specifically, as shown in [25], the proof of the strict case in point 1 of theorem 1.1 with vanishing magnetic charge $Q_{\rm M} = 0$ follows directly from the proof in [31] under the assumption of strict stability. We note that the rigidity result is lacking in [25]. We would like to mention that as this paper was written, we have learned that the inclusion of the marginally stable case may also be done [14] following the same procedure as in [31], and whose resolution would lead to the full inequality (1).

It is remarkable that both inequalities (1) and (2) can be obtained via a variational principle involving energy *flux* functionals [35]. Although both procedures can be carried over without reference to one another, the similarity between the functionals suggests a deeper relation between them, and ultimately, a possible relation between the inequalities themselves.

In this paper, we pursue three goals. The first one is to establish and give a detailed proof of the AJQ inequality, completing and extending the analysis in [25]. This is relevant for several reasons, namely it gives information about the allowed values of the physical parameters for black holes. In particular, it shows that, even in non-vacuum dynamical scenarios, the relations between these basic parameters remain simple. Also, it puts in evidence the special role of extreme Kerr–Newman black hole, a fact that might shed light on the stability of black holes. Finally, the relevance of this type of inequalities in the study of multiple black hole configurations and as a powerful tool to probe known solutions was made clear in the work of Neugebauer and Hennig [40] (see also [15]), where they strongly use the uncharged version of (1) to prove by contradiction that two rotating black holes do not exist in equilibrium.

The second goal of this paper is to gain insights into the underlying mechanisms leading to such an inequality as (1). In this respect, we expose two different approaches to the AJQ inequality. One of them relaxes to certain extent the axial symmetry assumption and makes use of harmonic maps between the surface and the complex hyperbolic space $\mathbb{H}^2_{\mathbb{C}}$. The second approach makes use of geodesics in $\mathbb{H}^2_{\mathbb{C}}$. Both approaches implement a minimization procedure which leads to theorem 1.1, a procedure which seems to be needed due to the presence of the angular momentum (cf [21, 46], where an inequality between area and electric and magnetic charges is obtained without axial symmetry and where no variational problem is formulated).

The third goal of this paper is to show how stable MOTSs and stable minimal surfaces over maximal surfaces can be treated on the same footing in the study of these quasilocal inequalities. Since the global characterization of a black hole in terms of notions such as the event horizon is of little practical use in the present quasilocal context, we must resort to quasilocal objects to represent black holes [28, 10, 11, 27]. Both MOTSs and minimal surfaces have been extensively studied and used in the literature as signatures of the presence of a black hole region, at least in strongly predictable spacetimes [29], and more precisely, in the study of quasilocal inequalities, but, as far as we know, no link was established between the two types of surfaces in this context. Although in the generic case there are fundamental differences between minimal surfaces and MOTS [2], in this paper, we point out that their respective notions of stability crucially lead (in axisymmetry) to the same integral characterization and ultimately to the same inequality.

Although much has been done in the last few years in the field of geometrical inequalities for black holes, there are still many open questions to be studied. One of them is the possible explicit inclusion of the cosmological constant into the inequalities in the presence of angular momentum (the area-charge case has already been addressed in [46]). We emphasize that our result, theorem 1.1, allows the spacetime to have a non-negative Λ, but this quantity does not enter into the inequality (1). So we wonder how is inequality (1) modified by its explicit introduction, and moreover, what happens with the negative Lambda case. Results in [46] provide a first step in this direction.

Another issue that must be better understood is the connection between the two types of inequalities mentioned above, i.e. (1) and (2). We give some insights into this paper (see the appendix), but there are many issues that are not entirely clear yet. This is not an easy problem since it involves linking global and quasilocal viewpoints. It would be, however, very desirable since its full resolution would give a concrete probe to compare with the Penrose inequality.

Finally, we want to mention that this type of quasilocal inequalities has been discussed in a broader context lately, mainly by Dain [22, 23], and we are forced to wonder about the universal validity of such a relation. Within the context of electrovacuum black holes, in this paper, we give a first step by studying the case of general surfaces within maximal initial data (that is, surfaces that are not necessarily minimal), and prove its validity. We understand that there is much work to do in order to generalize the results presented here to ordinary objects. Nevertheless, due to the rigidity statement in theorem 1.1 and the special properties of black holes in nature, one might expect that the extreme Kerr–Newman sphere should play a key role also in the general setting. We believe that this will be an active field of research in the coming years.

This paper is organized as follows. In section 2, we introduce the basic elements needed for the statement of our main result. This includes the formal definitions of angular momentum and charges of a surface within the Einstein–Maxwell-matter theory and an outline of stable axially symmetric MOTSs and stable axially symmetric minimal surfaces over maximal slices. In particular, as we mentioned above, we will show that the stability condition for both surfaces leads to the same integral characterization. Finally but crucially, we identify a set of suitable potentials to describe the gravitational and electromagnetic fields which proves to be useful for handling the variational problem needed to establish inequality (1).

In section 3, we present the main partial results leading to theorem 1.1, which are written up in the form of three lemmas, i.e. 3.1, 3.2 and 3.4 and give, respectively, a lower bound to the area in terms of a functional on the 2-sphere, an absolute lower bound to this functional and the rigidity statement. Moreover, in section 3.1, we present an interesting application to black hole initial data which intends to study the general validity of the AJQ inequality for black hole spacetimes. In section 3.2, we study the so-called extreme Kerr–Newman sphere, pointing out the MOTS and minimal surface viewpoints and its connection. We also give an interesting geometric description of the extreme Kerr–Newman horizon geometry in terms of semicircles in the complex hyperbolic space.

In section 4, we present the proof of theorem 3.2. We do so by following two approaches and highlighting different aspects of the underlying structure. The first one, in section 4.1, makes contact with harmonic maps, and the second one, in section 4.2, solves the minimization problem by identifying the minimizers with geodesics in the complex hyperbolic space.

We also include an appendix where we discuss the possible relation between quasilocal and global inequalities.

2. Settings

In this section, we introduce the objects that will be used as the black hole signatures, namely stable MOTSs and stable minimal surfaces. We will expose their main properties and, more importantly, show how, under the axisymmetry hypothesis, the stability notions for both types of surfaces lead to a single inequality from which (1) is obtained. In order to do so, we begin with a brief outline of closed surfaces embedded in a spacetime, their intrinsic and extrinsic geometry and the physical quantities one can associate with them.

Let $(\mathcal{V}, g_{ab})$ be a spacetime satisfying the Einstein equations

$$G_{ab} = 8\pi\left(T_{ab}^{EM} + T_{ab}^{M}\right) - \Lambda g_{ab}, \tag{3}$$

where $G_{ab} := R_{ab} - \frac{1}{2}Rg_{ab}$ is the Einstein tensor, g_{ab} and ∇_a are the spacetime metric and its Levi-Civita connection, respectively, $\Lambda \geqslant 0$ is a non-negative cosmological constant and we have decomposed the stress–energy tensor T_{ab} into its electromagnetic T_{ab}^{EM} and non-electromagnetic T_{ab}^{M} components. We assume that the latter satisfies the dominant energy condition.

Consider a space-like surface S embedded in the spacetime, with induced metric q_{ab} and Levi-Civita connection D_a. Let ℓ^a and k^a be future-oriented null vectors normal to S such that $\ell_a k^a = -1$ and ℓ^a is outward pointing. Regarding the extrinsic curvature elements, we introduce the expansion associated with ℓ^a, $\theta^{(\ell)} := q^{ab}\nabla_a \ell_b$, the shear tensor $\sigma_{ab}^{(\ell)} := q_a^c q_b^d \nabla_c \ell_d - \frac{1}{2}\theta^{(\ell)} q_{ab}$ and the normal fundamental form $\Omega_a^{(\ell)} := -k^c q^d{}_a \nabla_d \ell_c$. It is important to remark that the normalization condition on the null normals ℓ^a, k^a leaves a boost rescaling freedom: $\ell'^a = f\ell^a$, $k'^a = f^{-1}k^a$ under which $\theta^{(\ell)}$ and $\Omega_a^{(\ell)}$ transform, respectively, as $\theta^{(\ell')} = f\theta^{(\ell)}$ and $\Omega_a^{(\ell')} = \Omega_a^{(\ell)} + D_a \ln f$.

Although the main inequality can be understood more naturally in the context of globally axisymmetric black hole spacetimes, it is remarkable that in fact, only very little quasilocal (rather than global) axisymmetry is necessary for its validity. For this reason, we give here the most basic notion of axisymmetry under which (1) is valid.

We say that the closed surface S is *axisymmetric* if there exists a Killing vector field η^a on $\mathcal{S}$, i.e. $\mathcal{L}_\eta q_{ab} = 0$, with closed integral curves and normalized so that its integral curves have an affine length of 2π, and such that

$$\mathcal{L}_\eta \Omega_a^{(\ell)} = \mathcal{L}_\eta \Pi(A_a) = \mathcal{L}_\eta E_\perp = \mathcal{L}_\eta B_\perp = 0. \tag{4}$$

Above A_a is the electromagnetic potential given by $F_{ab} = \nabla_a A_b - \nabla_b A_a$, F_{ab} is the electromagnetic field tensor, $\Pi(A_a)$ is the pullback of the form A_a to the tangent space of S and $E_\perp$ and $B_\perp$ are the electric and magnetic fluxes across S, given by

$$E_\perp := \ell^a k^b F_{ab}, \qquad B_\perp := \ell^a k^b {}^*F_{ab}, \tag{5}$$

where ${}^*F_{ab}$ is the dual of F_{ab}. Note that $E_\perp$ and $B_\perp$ are independent of a conformal rescaling of the null normals.

If S is axisymmetric, then we define the projection of $\Omega_a^{(\ell)}$ along the Killing vector η^a, $\Omega_a^{(\eta)} := \eta^b \Omega_b^{(\ell)} \eta_a/\eta$, where $\eta := \eta^a \eta_a$. Crucially, $\Omega_a^{(\eta)}$ is then divergence-free and therefore invariant under null normal rescalings preserving the axisymmetry.

2.1. Angular momentum and electromagnetic charges

We now introduce three physical quantities[1] associated with a surface S in the context of Einstein–Maxwell-matter theory, namely the charges and the angular momentum.

Following [9], we write the electric and magnetic charges of the surface S, respectively, as

$$Q_{\rm E} = Q_{\rm E}(S) := -\frac{1}{4\pi}\int_S E_\perp \mathrm{d}S, \qquad Q_{\rm M} = Q_{\rm M}(S) := -\frac{1}{4\pi}\int_S B_\perp \mathrm{d}S, \tag{6}$$

where $\mathrm{d}S$ is the area element of S. By integrating Maxwell's equations $j^a = \nabla_b F^{ba}$ and $0 = \nabla_b {}^*F^{ba}$, where j^a is the electric charge current, we have the conservation law

$$Q_{\rm E}(\partial\Sigma) = \int j^a n_a \mathrm{d}V, \quad Q_{\rm M}(\partial\Sigma) = 0, \tag{7}$$

where Σ is a spatial 3-slice with boundary $\partial\Sigma$, n^b is the unit normal vector to Σ and $\mathrm{d}V$ is the volume element in Σ. In particular, this shows that in the absence of matter between two surfaces S, S', the charges are conserved, i.e. $Q(S) = Q(S')$.

If the surface S is axially symmetric with the axial Killing vector η^a, then one can define [7, 12] a *canonical angular momentum* of S given, within the Einstein–Maxwell-matter context, by

$$J = J(S) := -\frac{1}{8\pi}\int_S \Omega_a^{(\ell)}\eta^a \mathrm{d}S - \frac{1}{4\pi}\int_S A_a\eta^a E_\perp \mathrm{d}S. \tag{8}$$

If the axial vector η^a is the restriction of a global spacetime axisymmetric vector, J can be expressed as [13]

$$J(S) = -\frac{1}{16\pi}\int_S \nabla^b\eta^a \mathrm{d}S_{ab} - \frac{1}{4\pi}\int_S A_a\eta^a E_\perp \mathrm{d}S, \tag{9}$$

with $\mathrm{d}S_{ab} = 2\ell_{[a}k_{b]}\mathrm{d}S$. Note that, as given by (9), $J(S)$ is well defined for an arbitrary non-necessarily axisymmetric surface S. The first term is the so-called Komar angular momentum J_K, i.e. $J_K := -\frac{1}{16\pi}\int_S \nabla^b\eta^a \mathrm{d}S_{ab}$. Moreover, we also have [13, 26]

$$J(\partial\Sigma) = -\int_\Sigma T^M_{ab}\eta^a n^b \mathrm{d}V - \int_\Sigma \eta^a A_a j_b n^b \mathrm{d}V. \tag{10}$$

Therefore, in the absence of matter between surfaces S and S', the angular momentum (9) is conserved, $J(S) = J(S')$.

2.2. Stable MOTSs and stable minimal surfaces

We recall here the definitions of stable MOTSs in a given spacetime and stable minimal surfaces over maximal slices, which are the main two types of surfaces we are interested in this paper.

We say that S is a MOTS if $\theta^{(\ell)} = 0$. Moreover, we say that it is stable (or more precisely, spacetime stably outermost, according to the definition in [34]; see also [3, 4, 28, 42]) if there exists an outgoing vector $X^a = \gamma\ell^a - \psi k^a$, with functions $\gamma \geqslant 0$, $\psi > 0$, such that $\delta_X\theta^{(\ell)} \geqslant 0$. Here, δ_X denotes the deformation operator on S [3, 4, 12] that controls the infinitesimal variations of geometric objects defined on S under an infinitesimal deformation of the surface along the vector X^a. S is a *stable axisymmetric MOTS* if S is axisymmetric and stable with axisymmetric γ, ψ. Given a stable axisymmetric MOTS S with the axial Killing vector η^a, the stability condition for S is translated into the inequality (see [34, 36] for details)

$$\int_S \left[|D\alpha|_q^2 + \frac{R_S}{2}\alpha^2\right]\mathrm{d}S \geqslant \int_S \left[|\Omega^{(\eta)}|_q^2\alpha^2 + |\sigma^{(\ell)}|_q^2\alpha\beta + G_{ab}\alpha\ell^a(\alpha k^b + \beta\ell^b)\right]\mathrm{d}S\,, \tag{11}$$

[1] Note that the sign convention in this paper is consistent with that in [9, 50] and opposite to that in [25, 12]. This does not affect the inequality (1) that involves only quadratic expressions.

valid for all axisymmetric functions α on S. Here, $|\cdot|_q$ is the norm with respect to the 2-metric q_{ab} on S, R_S is the scalar curvature on S and $\beta := \alpha\gamma/\psi$. Use the Einstein equations (3) and insert the expression $8\pi T^{EM}_{ab}\ell^a k^b = E^2_\perp + B^2_\perp$ (see [12, 21]) in (11) to obtain the inequality (see [25])

$$\int_S \left[|D\alpha|^2_q + \frac{R_S}{2}\alpha^2 \right] \mathrm{d}S \geqslant \int_S \left[|\Omega^{(\eta)}|^2_q + E^2_\perp + B^2_\perp \right] \alpha^2 \mathrm{d}S. \tag{12}$$

To pass from the stability condition (11) to inequality (12), we have also discarded the non-negative shear term, the non-electromagnetic matter contribution (due to the energy condition) and the non-negative cosmological constant.

We now introduce the second type of surfaces we are interested in this paper, namely stable axisymmetric minimal surfaces over maximal slices. Consider maximal initial data $(\Sigma, h_{ab}, K_{ab}, E_a, B_a)$ for the Einstein–Maxwell-matter system, where h_{ab}, K_{ab} are the first and second fundamental forms of Σ, respectively, and $E_a := F_{ab}n^b$, $B_a := {}^*F_{ab}n^b$ are the electromagnetic fields on Σ. As the datum is maximal, we have $h^{ab}K_{ab} = 0$. Suppose that $S \subset \Sigma$ is a minimal surface, namely one whose mean curvature (inside Σ) is zero. The surface S is stable if the second variation of the area is non-negative, $\delta^2_{\alpha e_1} A \geqslant 0$ for all functions α. Suppose now that S is an axisymmetric surface in the sense introduced before where $\ell = n + e_1$, $k = n - e_1$ and where n is the spacetime normal to Σ and e_1 is one of the two normals to S in Σ. Then, the axisymmetric surface S is stable if the second variation of the area is non-negative, $\delta^2_{\alpha e_1} A \geqslant 0$ for all axisymmetric functions α. It is worth noting, to conciliate the stability definition for minimal surfaces and MOTS, that if for an axisymmetric minimal surface there is $\gamma > 0$ such that $\delta_{\gamma e_1}\theta \geqslant 0$ (where θ are the mean curvatures in Σ), then the surface is stable in the sense before.

Given a stable axisymmetric minimal surface S in a maximal slice, the stability condition is translated into the standard

$$\int_S \left[|D\alpha|^2 + \frac{R_S}{2}\alpha^2 \right] \mathrm{d}S \geqslant \int_S \frac{1}{2}[R + |\hat{\Theta}|^2]\alpha^2 \mathrm{d}S, \tag{13}$$

where $\hat{\Theta}$ is the trace-less part of the second fundamental form Θ and R is the scalar curvature of the slice. Using the energy constraint $R = |K|^2 + 2(|E|^2 + |B|^2) + 16\pi T^M_{ab}n^a n^b$ and that $\Omega^{(\eta)}_a \eta^a = -K(\eta, e_1)$, we obtain after discarding some non-negative quadratic terms (in $|K|^2$ and $|E|^2 + |B|^2$) and the non-negative term in $T^M_{ab}n^a n^b$, exactly the same inequality (12) which was obtained for stable MOTSs.

2.3. The quasilocal potentials

In this section, we write the relevant components of the intrinsic and extrinsic geometry of S, together with the electromagnetic field in terms of a set of potentials $\mathcal{D} = (\sigma, \omega, \psi, \chi)$ which are appropriate for applying the variational procedure which proves inequality (1).

Let S be either an axisymmetric stable MOTS or an axisymmetric stable minimal surface (in a maximal slice). We assume that over S either J, Q_E or Q_M is non-zero; otherwise, there is nothing to prove and the inequality (1) is trivial. Choosing $\alpha = 1$ in (12), and applying the Gauss–Bonnet theorem, it follows that the Euler characteristic of S is positive and therefore S is topologically a sphere. Thus, the metric over S can be written uniquely in the form (see [8, 20])

$$\mathrm{d}s^2 = \mathrm{e}^{2c-\sigma}\,\mathrm{d}\theta^2 + e^{\sigma}\sin^2\theta\,\mathrm{d}\varphi^2, \tag{14}$$

where c is a constant. With this choice of coordinate system, the area element and area of S are given, respectively, by $\mathrm{d}S = e^c \mathrm{d}S_0$, with $\mathrm{d}S_0 = \sin\theta \mathrm{d}\theta \mathrm{d}\varphi$, and $A = 4\pi e^c$. Moreover, the regularity of the metric at the poles requires $\sigma|_{\theta=0,\pi} = c$. In addition, the squared norm η of the axial Killing vector η^a is given by $\eta = e^{\sigma}\sin^2\theta$.

Regarding the 1-form $\Omega_a^{(\ell)}$, we write its Hodge decomposition in divergence-free and exact parts (see [34, 35])

$$\Omega_a^{(\ell)} = \epsilon_{ab}D^b\tilde{\omega} + D_a\lambda, \tag{15}$$

for some regular functions $\tilde{\omega}$ and λ on S. From the axisymmetry of S, it follows that this divergence-free part is given by $\Omega_a^{(\eta)}$. Explicitly,

$$\Omega_\theta^{(\eta)} = 0, \qquad \Omega_\varphi^{(\eta)} = -e^{\sigma-c}\sin\theta\,\tilde{\omega}', \tag{16}$$

where the prime denotes derivative with respect to the variable θ. Next, let ψ, χ, ω be regular functions of θ defined through the following expressions:

$$\psi' = -E_\perp e^c\sin\theta, \qquad \chi' = -B_\perp e^c\sin\theta, \tag{17}$$

$$\omega' = 2\eta\tilde{\omega}' - 2\chi\psi' + 2\psi\chi'. \tag{18}$$

It is remarkable that with this choice of potentials, the charges and angular momentum are given by the boundary values of ψ, χ and ω. To see this, use (6) and (17) to obtain

$$Q_{\mathrm{E}} = \frac{\psi(\pi) - \psi(0)}{2}, \qquad Q_{\mathrm{M}} = \frac{\chi(\pi) - \chi(0)}{2} \tag{19}$$

and use (8) and (18) to find

$$J = \frac{\omega(\pi) - \omega(0)}{8}. \tag{20}$$

Moreover, since the potentials ψ, χ, ω are defined up to an additive constant, we assume, without loss of generality, that $\psi(\pi) = -\psi(0) = Q_{\mathrm{E}}$, $\chi(\pi) = -\chi(0) = Q_{\mathrm{M}}$ and $\omega(\pi) = -\omega(0) = 4J$.

Finally, note that the function ω_K defined through $\omega_K' = 2\eta\tilde{\omega}'$ provides a potential for the Komar angular momentum J_K (cf [34]): $\omega_K(\pi) = -\omega_K(0) = 4J_K$.

3. Discussion of the main results

We present here the main results leading to the AJQ inequality, a discussion about its more general validity and a detailed study of the unique minimizer for the area, namely the extreme Kerr–Newman sphere.

We begin by stating three lemmas which, together, prove theorem 1.1. Lemma 3.1 establishes a lower bound for the area in terms of a bounded functional $\mathcal{M}$. This result comes up simply by rewriting the stability condition (12) for the surface S in terms of the set $\mathcal{D}$. The second statement, lemma 3.2, gives an explicit sharp bound for the functional $\mathcal{M}$ in terms of the angular momentum and charges. These two lemmas prove the AJQ inequality. The final statement, lemma 3.4, proves that the AJQ inequality is saturated by a unique set $\mathcal{D}_0$ called the extreme Kerr–Newman sphere. We will discuss this special set in section 3.2.

Consider the stability condition (12) valid for axisymmetric MOTSs and minimal surfaces (over maximal slices). Since it holds for any axisymmetric function α, we take, as in [21, 34, 25], the probe function

$$\alpha = e^{c-\sigma/2}. \tag{21}$$

Some insights into the reason for this choice of α are provided in section 3.2. Then, rewrite (12) in terms of the potentials (17)–(18) and use $A = 4\pi e^c$ to arrive, following [20, 34, 25], at a fundamental inequality which is summarized in the following lemma [25].

Lemma 3.1. *Let S be an axisymmetric stable MOTS or an axisymmetric stable minimal surface in a maximal slice. Then,*

$$A \geqslant 4\pi e^{\frac{\mathcal{M}-8}{8}}, \tag{22}$$

where $\mathcal{M}$ is given by

$$\mathcal{M} := \frac{1}{2\pi}\int\left[4\sigma + |D\sigma|^2 + \frac{|D\omega + 2\chi D\psi - 2\psi D\chi|^2}{\eta^2} + 4\frac{|D\psi|^2 + |D\chi|^2}{\eta}\right] dS_0, \tag{23}$$

and the norm $|\cdot|$ is taken with respect to the standard round metric on S^2.

A fundamental sharp lower bound for the functional $\mathcal{M}$ is stated in the following lemma.

Lemma 3.2. *Let $\mathcal{D} = (\sigma, \omega, \psi, \chi)$ be a regular set on S^2 with fixed values of J, Q_E and Q_M. Then,*

$$e^{\frac{\mathcal{M}-8}{4}} \geqslant 4J^2 + Q^4, \tag{24}$$

with $Q^2 = Q_E^2 + Q_M^2$.

The proof of this result involves a minimization problem and can be approached in different ways, that will be discussed in full detail in section 4.

We want to emphasize that lemma 3.2 does not assume axisymmetry on the set $\mathcal{D} = (\sigma, \omega, \psi, \chi)$. One of the proofs will deal with these non-necessarily axisymmetric sets (see section 4.1). Finally, we give the precise definition of regular set mentioned in the lemma.

Definition 3.3. *The set $\mathcal{D} = (\sigma, \omega, \psi, \chi)$ on S^2 is a regular set if the functions σ, ω, ψ and χ are C^∞ on S^2, and moreover, we have the following behavior near the poles:*

(i) $\omega = \pm 4|J| + O(\sin^2\theta)$, $\psi = \pm Q_E + O(\sin^2\theta)$, $\chi = \pm Q_M + O(\sin^2\theta)$, *where the signs $+, -$ refer to the values at $\theta = \pi, 0$, respectively.*
(ii) $|D\omega + 2\chi D\psi - 2\psi D\chi| = O(\sin^3\theta)$.

We remark that if the functions ω, ψ, χ arise from a smooth set of axisymmetric fields $\Omega_a^{(\eta)}$, $E_\perp$, $B_\perp$ via equations (16)–(18), then they satisfy items (i) and (ii) of definition 3.3 automatically.

It is also important to stress that lemma 3.2 is also valid for smooth functions σ, ω, ψ, χ such that they satisfy condition (i) in the above definition and $\mathcal{M}$ is finite, condition (ii) being no longer necessary. We will come back to this point in section 4.1.

The final result we present concerns the uniqueness of a regular set saturating inequality (1).

Lemma 3.4. *There exists a unique regular set $\mathcal{D}$ saturating the AJQ inequality (1), with $A = 4\pi e^{\sigma}\big|_{\theta=0,\pi}$ and it is the extreme Kerr–Newman sphere set $\mathcal{D}_0 = (\sigma_0, \omega_0, \psi_0, \chi_0)$ given by*

$$\sigma_0 = \ln\frac{\left(2a_0^2 + Q^2\right)^2}{\Sigma_0}, \tag{25}$$

$$\omega_0 = -4J\frac{2a_0^2 + Q^2}{\Sigma_0}\cos\theta, \tag{26}$$

$$\psi_0 = -\frac{Q_E\left(2a_0^2 + Q^2\right)\cos\theta - Q_M a_0\sqrt{a_0^2 + Q^2}\sin^2\theta}{\Sigma_0}, \tag{27}$$

$$\chi_0 = -\frac{Q_{\rm M}\left(2a_0^2 + Q^2\right)\cos\theta + Q_{\rm E}a_0\sqrt{a_0^2+Q^2}\sin^2\theta}{\Sigma_0} \tag{28}$$

with

$$J = a_0 m_0 = a_0\sqrt{a_0^2+Q^2}, \quad m_0 = \sqrt{\frac{Q^2+\sqrt{4J^2+Q^4}}{2}}, \tag{29}$$

$$\Sigma_0 = Q^2 + a_0^2(1+\cos^2\theta). \tag{30}$$

In section 3.2, we will discuss the properties of the minimizer set $\mathcal{D}_0$ and show that this special datum appears in two non-equivalent important and concrete contexts:

O1. on a MOTS in the horizon of the extreme Kerr–Newman solution, and,
O2. on a minimal sphere in the *extreme Kerr–Newman throat*, which is a maximal initial datum.

3.1. On the general validity of the AJQ inequality

Case 2 in theorem 1.1 allows to show that, in some situations, the *AJQ*-inequality (1) is valid for any surface and not just for stable minimal axisymmetric surfaces over a maximal slice. The particular situations in consideration will be those of 'trumpet' and 'doubly asymptotically flat (AF)' axisymmetric initial data.

This result is interesting in the light of the conjecture that the foliation by maximal slices, whose leaves are all of the same type (trumpet or doubly AF), is believed to cover the whole domain of outer communication of the black holes. For this type of solutions, one expects the inequality to hold over a large variety of surfaces in the whole domain of outer communication. It is worth stressing that one does not expect the equality in (1) to be achieved at any surface in the trumpet or doubly AF maximal slice [43].

To define axisymmetric 'trumpet' initial data sets, we follow [48] and refer the reader to this paper for more details. A 'trumpet' initial data set for the Einstein–Maxwell equations is a maximal and axisymmetric electrovacuum initial datum $(\Sigma; h, K; E, B)$, with $\Sigma \approx \mathbb{R}^3\setminus\{0\}$ and $\Sigma/_{U(1)} \approx [0, 1] \times \mathbb{R}$ (in particular with an axis having two connected components) and with particular asymptotics at the origin and at infinity. Precisely, let x^i be the standard coordinates on $\mathbb{R}^3$, we require asymptotic flatness at infinity (of $\mathbb{R}^3$), and at the origin (of $\mathbb{R}^3$) requiring h to approach a cylindrical metric in the following sense: there exists a diffeomorphism Φ between, say, $B_{1/2} \setminus \{0\}$ and $(T, \infty) \times S^2$ so that $(\Phi_*\bar{h})_{ij} - \bar{h}_{ij} = o(1)$ as $t \to \infty$, where $\bar{h}$ denotes the cylindrical metric of the form $\bar{h} = f^2 dt^2 + q$, with q being a Riemannian metric on S^2. We refer to the origin as a cylindrical end. 'Doubly AF' initial data are defined in exactly the same way but now with two AF ends, at infinity and at the origin of $\mathbb{R}^3$. We prove below the following proposition.

Proposition 3.5. *Consider either an axisymmetric 'doubly AF' or a 'trumpet' maximal initial datum* $(\Sigma; (g, K); (E, B))$ *for the electrovacuum system, with total angular momentum and charges* $\mathcal{J}$, $\mathcal{Q}_{\rm E}$ *and* $\mathcal{Q}_{\rm M}$. *Then for any oriented, non-necessarily axisymmetric embedded surface S of arbitrary topology, its angular momentum and charges are given by one of the following two possibilities:*

$$J = 0, \; Q_{\rm E} = 0, \; Q_{\rm M} = 0, \quad \text{or} \tag{31}$$

$$J = \mathcal{J}, \; Q_{\rm E} = \mathcal{Q}_{\rm E}, \; Q_{\rm M} = \mathcal{Q}_{\rm M}. \tag{32}$$

Moreover, the AJQ-inequality (1) holds.

Proof. To better visualize the proof, let us assume that we choose a diffeomorphism between $\mathbb{R}^3 \setminus \{0\}$ and Σ in such a way that the orbits of the Killing field, as seen in $\mathbb{R}^3 \setminus \{0\}$, are exactly those circles which are the rotations of points around the z-axis. In this way, the two components of the axis are given by $\{(x, y, z), x = y = 0, z > 0\}$ and $\{(x, y, z), x = y = 0, z < 0\}$.

Let S be an oriented surface. As a surface in $\mathbb{R}^3 \setminus \{0\} \subset \mathbb{R}^3$, it divides $\mathbb{R}^3$ into two connected components. If the unbounded component contains the origin $\{0\}$, then S encloses (including S) a compact region in $\mathbb{R}^3 \setminus \{0\}$ and therefore (by Gauss theorem) J, Q_E and Q_M are zero, namely their values are as in (31). In this case, (1) is trivial. We assume therefore that it is the bounded component that contains the origin. In this case, the values of J, Q_E and Q_M are (by Gauss theorem again) those of the end, namely as in (32).

Now, in order to prove that the AJQ inequality (1) is satisfied, assume by contradiction that (1) does not hold. Following [39], there are surfaces[2] (possibly repeated) $S_1, \ldots, S_m$ realizing the infimum of the areas $A(\tilde{S})$ where $\tilde{S}$ is isotopic to S, namely $\sum A(S_i) = \inf\{A(\tilde{S}), \tilde{S} \sim S\}$, where $\tilde{S} \sim S$ means that $\tilde{S}$ is isotopic to S (one can see that the infimum is non-zero). Moreover, the surfaces are non-contractible (to a point) in $\mathbb{R}^3 \setminus \{0\}$ and are also embedded. It follows that they are orientable (otherwise are contractible) and stable. As the manifold (Σ, g) is axisymmetric (complete) and non-compact, every S_i, $i = 1, \ldots, m$ is known to be axisymmetric. So each of them is either an axisymmetric sphere or an axisymmetric torus (there are no axisymmetric surfaces of higher genus). But, any axisymmetric torus is contractible to a point in $\mathbb{R}^3 \setminus \{0\}$, which is not possible. Therefore, all S_i's are axisymmetric spheres and as they are non-contractible (to a point) in $\mathbb{R}^3 \setminus \{0\}$, they all must enclose the origin. Thus, the angular momentum and charges of, say, S_1, are the given $\mathcal{J}$, $\mathcal{Q}_E$ and $\mathcal{Q}_M$. Therefore, we have

$$A^2(S) \geqslant A^2(S_1) \geqslant 16\pi^2\left[4\mathcal{J}^2 + \left(\mathcal{Q}_E^2 + \mathcal{Q}_M^2\right)^2\right] = 16\pi^2\left[4J^2 + \left(Q_E^2 + Q_M^2\right)^2\right] \tag{33}$$

as desired. □

3.2. A discussion on the extreme Kerr–Newman sphere

We have seen above that the set $\mathcal{D}_0$ given by equations (25)–(28) plays a crucial role in bounding the area of an axisymmetric MOTS or minimal surface (over a maximal slice), and moreover, due to proposition 3.5, in bounding the area of any surface in axially symmetric electrovacuum initial data. We show here how this set is related to extreme Kerr–Newman solution, from where it takes the name *extreme Kerr–Newman sphere set*.

It is well known that the Kerr–Newman solution is parametrized by four quantities: the mass m, the angular momentum J and the electromagnetic charges Q_E, Q_M. Of these parameters, let J, Q_E, Q_M be fixed and decrease the remaining parameter m as $m \downarrow m_0$. If we denote by $\mathcal{D}_m$ the set on the bifurcating sphere (for each m), then the limit $\lim_{m\downarrow m_0} \mathcal{D}_m = \mathcal{D}_0$ is attained. In other words, we take the limit of $\mathcal{D}_m$ as the black holes become extremal to obtain $\mathcal{D}_0$ in (25)–(28). As we discuss below, this way of finding $\mathcal{D}_0$ allows one to see how this particular kind of datum arises in the contexts **O1** and **O2** mentioned in section 3.

The spacetime metrics for the Kerr–Newman solutions, in the usual Boyer–Lindquist coordinates, are given by (see [13])

$$\begin{aligned} g_{ab}\mathrm{d}x^a\mathrm{d}x^b = &-\frac{\Delta - a^2\sin^2\theta}{\Sigma}\mathrm{d}t^2 - \frac{2a\sin^2\theta}{\Sigma}(r^2 + a^2 - \Delta)\,\mathrm{d}t\,\mathrm{d}\phi \\ &+ \frac{(r^2+a^2)^2 - \Delta a^2\sin^2\theta}{\Sigma}\sin^2\theta\mathrm{d}\phi^2 + \frac{\Sigma}{\Delta}dr^2 + \Sigma\mathrm{d}\theta^2, \end{aligned} \tag{34}$$

[2] The conclusion is direct for 'doubly AF' initial data. For 'trumpet' data, it requires a little more effort but feasible by taking into account that the 'asymptotic spheres' over the cylindrical end satisfy (1).

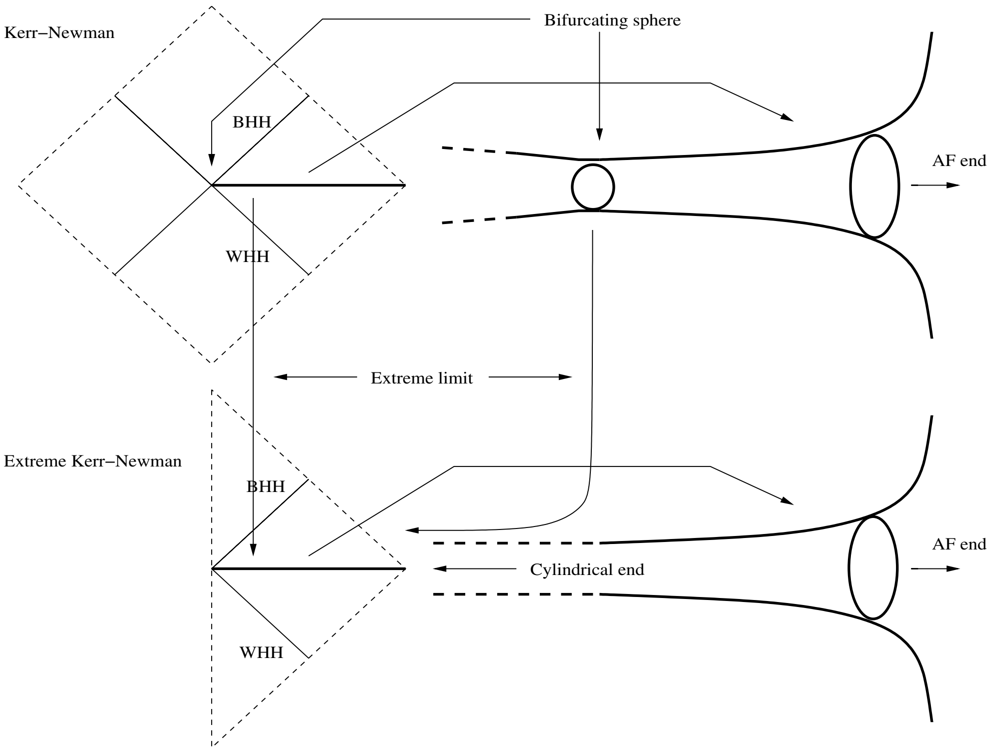

Figure 1. Penrose diagram of the Kerr–Newman solution and its $\{t = 0\}$ slice. Below, the Penrose diagram of the extreme Kerr–Newman solution and its $\{t = 0\}$ slice displaying the cylindrical end and the AF end.

where

$$\Sigma := r^2 + a^2 \cos^2 \theta, \qquad \Delta := r^2 + a^2 + Q^2 - 2mr. \tag{35}$$

The parameter $a = J/m$ is the angular momentum per unit mass and again $Q^2 = Q_{\rm E}^2 + Q_{\rm M}^2$. The electromagnetic part of the solution is encoded in the potential A_a which is given explicitly by (see [13])

$$A_a = -\frac{Q_{\rm E} r}{\Sigma}[(\mathrm{d}t)_a - a\sin^2\theta(\mathrm{d}\phi)_a] + \frac{Q_{\rm M}\cos\theta}{\Sigma}[a(\mathrm{d}t)_a - (r^2 + a^2)(\mathrm{d}\phi)_a]. \tag{36}$$

The *subextremal* Kerr–Newman black holes are those solutions with $m^2 > \frac{Q^2+\sqrt{4J^2+Q^4}}{2}$. The *extreme Kerr–Newman black holes* are those solutions with $m^2 = \frac{Q^2+\sqrt{4J^2+Q^4}}{2}$. Let us concentrate on non-extreme Kerr–Newman black holes. Let r_H be the greatest root of $\Delta = 0$ (corresponding to the event horizon), explicitly $r_H = m + \sqrt{m^2 - a^2 - Q^2}$. The range of coordinates $\{r \geqslant r_H\}$ $(t \in \mathbb{R}, \theta \in [0, \pi), \varphi \in [0, 2\pi)$ arbitrary) covers exactly the whole domain of outer communication and its boundary $\{r = r_H\}$ consists of a bifurcating sphere, a black hole horizon and a white hole horizon (respectively, BHH and WHH, see figure 1). The bifurcation surface is located at $\{r = r_H\}$ over the maximal slice $\{t = 0\}$. It has a dual character: it is at the same time a strictly stable minimal surface over the (doubling of the) maximal slice $\{t = 0\}$ and a strictly stable MOTS on the spacetime. The area of the bifurcating sphere is easily calculated (use that $\Delta(r_H) = 0$) as

$$A(S_H) = 4\pi\left(r_H^2 + a^2\right) > 4\pi\sqrt{4J^2 + Q^4}, \tag{37}$$

and we have

$$A(S_H) \downarrow 4\pi\sqrt{4J^2 + Q^4}, \tag{38}$$

as $m \downarrow m_0$.

We make now some claims, crucial to link the MOTS and minimal surface perspectives and show why the limit $\mathcal{D}_m \to \mathcal{D}_0$ allows us to see the set $\mathcal{D}_0$ as in **O1** and **O2**.

1. The set $\mathcal{D}_m$ over S_H is the same as the set on any axisymmetric sphere S embedded in the black (white) hole horizon. This can be seen as follows. The past (future) spacetime flow generated by the stationary Killing field pushes any surface S over the black (white) hole toward S_H and the convergence is smooth. As the flow by the Killing vector field is an isometry (also leaving F_{ab} invariant) and the components $(\sigma, \omega, \psi, \chi)$ of the set on the surface S are intrinsic to the surface, it follows by this and continuity that the set over S_H or over any axisymmetric sphere S must be the same.
2. The black (white) hole horizon of the extreme Kerr–Newman solution is the limit of the black (white) hole horizon of the Kerr–Newman black hole solutions as $m \downarrow m_0$. To see this, just take the point-wise limit of expression (34). In this limit, the horizons $\{r = r_H\}$ approach the extreme horizon $\{r = m_0 = \sqrt{Q^2 + a_0^2}\}$.
3. For every m, consider the initial data over $\Sigma = \{t = 0\}$, $(\Sigma, h_m, K_m; E_m, B_m)$ where we put a subindex m to emphasize that the initial datum is parametrized by m. 'Following' the initial data around S_H as $m \downarrow m_0$, a smooth limit initial datum is obtained (see below for details on how to perform the limit). It is the so-called *extreme Kerr–Newman throat*, which is a maximal electrovacuum initial datum on $\mathbb{R} \times S^2$ with the explicit form

$$h_T = \Sigma_0\, \mathrm{d}\tilde{r}^2 + \Sigma_0\, \mathrm{d}\theta^2 + \frac{(4J^2 + Q^4)}{\Sigma_0} \sin^2\theta\, \mathrm{d}\varphi^2, \tag{39}$$

$$K_T = -J\big((J/a_0)^2 + a_0^2\big)\frac{\sin^2\theta}{\Sigma_0^{\frac{3}{2}}}(\mathrm{d}\tilde{r}\,\mathrm{d}t + \mathrm{d}t\,\mathrm{d}\tilde{r}), \tag{40}$$

$$E_T = -\left[Q_{\mathrm{E}}\big(Q^2 + a_0^2\sin^2\theta\big) - Q_{\mathrm{M}}2J\cos\theta\right]\frac{\mathrm{d}\tilde{r}}{\Sigma_0^{\frac{3}{2}}}, \tag{41}$$

$$B_T = -\left[Q_{\mathrm{M}}\big(Q^2 + a_0^2\sin^2\theta\big) + Q_{\mathrm{E}}2J\cos\theta\right]\frac{\mathrm{d}\tilde{r}}{\Sigma_0^{\frac{3}{2}}}. \tag{42}$$

The solution is independent of $\tilde{r}$ (the coordinate in the $\mathbb{R}$ factor; see below), which implies that $\partial_{\tilde{r}}$ is a Killing field. For this reason, the initial datum has the same form if we replace $\tilde{r}$ by $\tilde{r} + c$, where c is a constant. In particular, the coordinate can be chosen in such a way that the bifurcating sphere S_H (for given m) converges (as $m \downarrow m_0$) to the minimal sphere $S_0 = \{\tilde{r} = 0\}$ that we define as an *extreme Kerr–Newman throat sphere* and which because of (38) satisfies (1). Of course, any other sphere with constant $\tilde{r}$ has the same set of potentials $\mathcal{D}_0$.

We emphasize that the calculations leading to the extreme Kerr–Newman throat initial datum (39)–(42) are long but straightforward if one follows a simple procedure. From (34)–(36), obtain the explicit expressions of h, K, E and B, over $\{t = 0\}$ in the coordinates $\{r, \theta, \varphi\}$ $(r \geqslant r_H)$. Then, make the change of the radial coordinate r to $\tilde{r}$ as

$$\tilde{r}(r) = \int_{r_H}^{r} \frac{1}{\sqrt{\Delta(\bar{r})}}\, \mathrm{d}\bar{r}. \tag{43}$$

Of course, $\tilde{r}(r_H) = 0$. Express h, K, E, B whose components were given in terms of $\{(r, \theta, \varphi)\}$ in the coordinates $\{(\tilde{r}, \theta, \varphi)\}$. Note that now the range of the coordinates

$\{(\tilde{r},\theta,\varphi)\}$ is $[0,\infty)\times[0,\pi)\times[0,2\pi)$. Then *in this domain* take the point-wise limit $m\downarrow m_0$ of each component of the fields (in the $\{(\tilde{r},\theta,\varphi)\}$ coordinates). The result is (39)–(42).

Summarizing, from 1, 2 and 3, one obtains that the set $\mathcal{D}_0=\lim_{m\downarrow m_0}\mathcal{D}_m$ verifying (1) can be achieved as the set on a MOTS inside a spacetime (more precisely on an axisymmetric surface over the horizon of the extreme Kerr–Newman solution), or as the set endowed on stable minimal surfaces over maximal slices (more precisely over the extreme Kerr throat initial datum).

To see that $\mathcal{D}_0$ is given by (25)–(28), proceed as follows. Making $\tilde{r}=0$ in (39), one obtains the 2-metric of the extreme Kerr–Newman sphere to be

$$\mathrm{d}s^2=\Sigma_0\,\mathrm{d}\theta^2+\frac{(4J^2+Q^4)}{\Sigma_0}\sin^2\theta\,\mathrm{d}\varphi^2. \tag{44}$$

From the definition of σ in (14), one obtains (25). To obtain (27) and (28), use (41) and (42) and the definitions (17). We discuss now how to obtain (26). Over any 2-sphere $\{r=r_1>r_H, t=0\}$ on a Kerr–Newman black hole, one uses the null vectors [47]

$$\ell^a=(r^2+a^2)(\partial_t)^a+a(\partial_\phi)^a+\Delta(\partial_r)^a \tag{45}$$

$$k^a=\left(\frac{r^2+a^2}{2\Delta\Sigma}\right)(\partial_t)^a+\frac{a}{2\Delta\Sigma}(\partial_\phi)^a-\frac{1}{2\Sigma}(\partial_r)^a, \tag{46}$$

normalized such that $\ell^ak_a=-1$ to calculate $\Omega_a^{(\ell)}$. Taking the limit $r_1\to r_H$ and then the limit $m\downarrow m_0$, one obtains a limit form over the extreme Kerr throat, which can be calculated to be

$$\Omega^{(\ell)}=-\frac{1}{(2\Sigma_0)^2}\left(2a_0^2\Sigma_0\sin(2\theta)\mathrm{d}\theta+4a_0\sqrt{a_0^2+Q^2}\left(2a_0^2+Q^2\right)\sin^2(\theta)\mathrm{d}\phi\right). \tag{47}$$

From (15) and axial symmetry[3], by solving $\partial_\theta\tilde{\omega}=\Omega_\phi^{(\ell)}$ and $\partial_\theta\lambda=\Omega_\theta^{(\ell)}$, and taking into account $\omega'_{K0}=2\eta\tilde{\omega}'_0$, we obtain

$$\omega_{K0}(\theta)=\frac{\left(2a_0^2+Q^2\right)^2Q^2}{a_0^2\left(a_0^2+Q^2\right)}\arctan\left(\frac{a_0\cos\theta}{\sqrt{a_0^2+Q^2}}\right)-\cos\theta\frac{\left(2a_0^2+Q^2\right)^3}{\sqrt{a_0^2+Q^2}a_0\Sigma_0}$$
$$\lambda(\theta)=\ln[\sqrt{2\Sigma_0}]. \tag{48}$$

Moreover, we verify $\omega_{K0}(\theta=0)=-\omega_{K0}(\theta=\pi)=-4J_K$, where J_K is the Komar contribution to the total angular momentum

$$J_K=\frac{\left(2a_0^2+Q^2\right)^2}{4a_0^2\left(a_0^2+Q^2\right)}\left[a_0\sqrt{a_0^2+Q^2}-Q^2\arctan\left(\frac{a_0}{\sqrt{a_0^2+Q^2}}\right)\right]. \tag{49}$$

Using expression (48), together with (27) and (28) into (18), we obtain ω_0 in (26), and thus, complete the derivation of the set $\mathcal{D}_0$.

Finally, we present two remarks concerning the extreme Kerr–Newman sphere:

- $\mathcal{D}_0$ in $\mathbb{H}^2_{\mathbb{C}}$. There is an interesting description of the geometry of the extreme Kerr–Newman sphere with vanishing magnetic charge ($Q_{\mathrm{M}}=0$), which shows the underlying connection with the complex hyperbolic space. This connection will be exposed in section 4.2 and arises when one studies the critical point of the functional $\mathcal{M}$. What we want to show here is that the set $\mathcal{D}_0$ can be visualized as two arcs of circles in $\mathbb{H}^2_{\mathbb{C}}$ (cf figure 2). In order

[3] More generally, one can fix $\tilde{\omega}$ and λ by solving the second-order system: $D^aD_a\tilde{\omega}=-f$, $D^aD_a\lambda=D^a\Omega_a^{(\ell)}$, where $(\mathrm{d}\Omega^{(\ell)})_{ab}=f\epsilon_{ab}$.

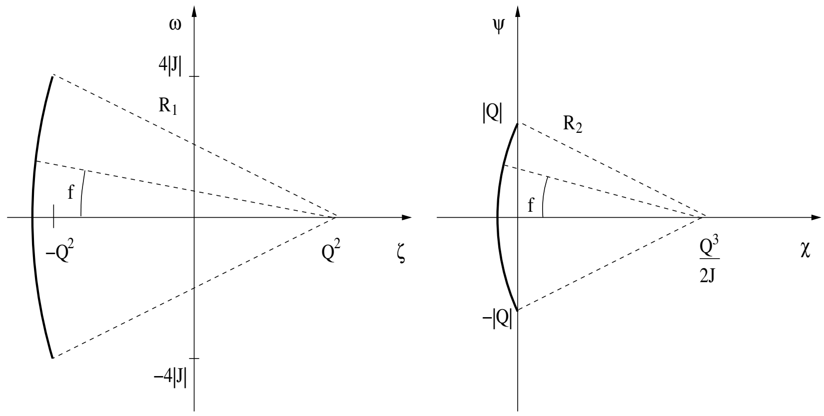

Figure 2. $\mathcal{D}_0$ as arcs of circles.

to describe these arcs, we consider, instead of the quadruple $(\sigma, \omega, \psi, \chi)$, the two pairs (ζ, ω) and (ψ, χ) where

$$\zeta = -(\eta + \psi^2 + \chi^2). \tag{50}$$

Then, whether by working with the Euler–Lagrange equations of $\mathcal{M}$ (as done in section 4.2, precisely, the form of the rhs of (51) comes from the last two equations of (112), while the form of lhs comes from the first equation of (112) and (119)) or with the explicit expression for the potentials, equations (25)–(28), we find the following remarkable relations:

$$\zeta + i\omega = R_1 e^{if} + B_1, \qquad \chi + i\psi = R_2 e^{if} + B_2, \tag{51}$$

where the angle to the center f is given by (see (101))

$$f = 2\arctan\left(\frac{\sqrt{Q^4 + 4J^2} - Q^2}{2J}\cos\theta\right) \tag{52}$$

and $R_1 = -2\sqrt{4J^2 + Q^4}$, $B_1 = Q^2$, $R_2 = -\frac{Q}{2J}\sqrt{4J^2 + Q^4}$, $B_2 = Q^3/2J$.
This shows that the first arc, in the (ζ, ω) plane, starts at $\left(-Q^2, -4|J|\right)$ and ends at $\left(-Q^2, 4|J|\right)$ (this can be obtained by evaluating the pair at the values $0, \pi$, respectively). The center of the circle to which the arc belongs lies on the $\zeta = 0$ axis and its radius is R_1. The arc in the (ψ, χ) plane starts at $(-Q, 0)$ and ends at $(Q, 0)$. The center of the circle to which the arc belongs lies on the $\psi = 0$ axis and its radius is R_2.

- *On the choice of the probe function* α. Here, we want to give some insights into the choice of the function α, equation (21), entering in the stability condition (12). In particular, it has the nature of a rescaling factor between null normals and we show that it is related to the minimizing set $\mathcal{D}_0$. From the transformation properties of $\Omega_a^{(\ell)}$ under a rescaling of the null normals ℓ^a, k^a, we note that the null vector $\ell_o^a = e^{-\lambda} \cdot \ell^a$, with λ given by the expression in (48), is such that the associated fundamental form $\Omega_a^{(\ell_o)}$ is divergence-free, i.e. $D^a\Omega_a^{(\ell_o)} = 0$. This provides a natural quasilocal normalization for the outgoing null vector on S. On the other hand, evaluating α in (21) with the expressions in (25), we can check $\alpha\ell_o^a = \text{const} \cdot \ell^a_{\text{Killing}}$, where $\ell^a_{\text{Killing}} = (\partial_t)^a - \Omega\partial_\phi$ is the only null vector on S (up to constant) that extends as a Killing vector in a spacetime neighborhood of S (here, Ω is the constant horizon angular velocity). In other words, our choice of α in (21) provides precisely the rescaling from the canonical quasilocal choice ℓ_o^a on S, with a divergence-free fundamental form, to the globally defined Killing vector of the Kerr–Newman spacetime that becomes null on the horizon. This remark is explained by the rigidity results in [36] (see also the analysis in [38]).

4. Different avenues to prove the AJQ inequality

The AJQ inequality (1) is obtained from two ingredients, namely from the stability condition, leading to lemma 3.1, and from the resolution of the naturally associated minimizing problem, leading to lemma 3.2. In this section, we show two different ways to approach the variational principle.

It is also worth mentioning here that an interesting approach to the AJQ inequality might be inspired by the recent work of Schoen and Zhou [45]. In that article, the mass-angular momentum inequalities for black hole initial data within the context of Einstein–Maxwell theory are proven. Moreover, via a careful study of the asymptotic conditions on the data, they obtain a lower bound on the gap between the general data and the extreme Kerr–Newman data. This gives not only a stronger control on the geometry of the initial data, but also a characterization of the Kerr–Newman solution as a border case, a very important issue that was lacking in [16, 17]. Since the reduced harmonic energy on initial data studied by Schoen and Zhou is on a similar footing as the functional $\mathcal{M}$,(23), on potentials on the 2-surface (see the appendix for a discussion on the link between the functional $\mathcal{M}$ and an analogous functional on initial data potentials), it is very likely that a similar bound on the difference between general potentials and the extreme Kerr–Newman sphere potentials can be found. This would imply a refinement and improvement of the AJQ inequality. Works along this line are under current research and will be presented elsewhere.

Before addressing these points, a remark on the implications of the analysis of the AJQ inequality in the stationary case is in order. In [30, 31, 5], the strict version of inequality (1), with vanishing magnetic charge, is proved for Killing horizons in axisymmetric spacetimes. The scheme of that proof shares the two ingredients of the analysis in this section: first, use of a stability condition in the form of a horizon (outer) subextremal assumption [28, 12] from which an *integral stability condition* for axisymmetric Killing horizons is derived; second, definition of a variational problem from such an integral stability condition, whose resolution leads to the strict (1). Remarkably, in [25], it is explicitly shown that the first ingredient, namely the integral stability condition, can be derived directly from the quasilocal (strict) stability of axisymmetric MOTS, in particular from the strict version of inequality (11). Further geometrical insight into the relation between the stationary axisymmetric black hole condition and the quasilocal MOTS stability is provided in [15, 38]. As a consequence, the variational analysis in [30, 31, 5] can be exactly applied to strictly stable axisymmetric MOTS, so that the proof in [30, 31, 5] of the strict inequality (1) with $Q_{\rm M} = 0$ extends straightforwardly from the stationary setting to the dynamical case with arbitrary standard matter [25] (namely the strict version of item 1 in theorem 1.1). The extension of the variational problem in [30, 31, 5] to include the equality case and the rigidity analysis is under research [14].

Following a different rationale, the two approaches to the variational problem discussed in this section aim at enriching the understanding of the geometric structure underlying theorem 1.1. We believe that each of the perspectives presented here gives important insights into this problem.

The first approach, in section 4.1, deals with non-necessarily axisymmetric sets of potentials $\mathcal{D}$ and its associated functional $\mathcal{M}$. Although the main result, theorem 1.1, holds in the physical scenario of axisymmetric surfaces S, the fact that the variational problem can be stated and solved outside axisymmetry shows that the extreme Kerr–Newman sphere plays a special role among a wider class of sets $\mathcal{D}$. Inspired by this generalization, one is tempted to think about the possibility of extending inequality (1) to other non-necessarily axisymmetric physical situations. This, however, is not an easy task, mainly because it is not clear how to give a satisfactory canonical definition of angular momentum outside axial symmetry. Nevertheless, if such a statement can be made, the functional $\mathcal{M}$ and its properties studied here might be of relevance.

The second approach, in section 4.2, is restricted to axisymmetry and therefore, when solving the minimization problem for $\mathcal{M}$, the Euler–Lagrange equations reduce to a system of ordinary differential equations which can be solved explicitly. Then, a remarkable point that comes up when studying these equations is that the boundary conditions J, Q_E and Q_M for the minimizer of $\mathcal{M}$ determine uniquely the boundary conditions for the remaining potential σ. This is the key fact under the sharpness of inequality (1). Actually, an important consequence of this is that we can prove uniqueness for the minimizer of $\mathcal{M}$ with given values of J, Q_E, Q_M without any reference to the boundary values of σ. This is a difference to what we do in the non-axisymmetric case, where the boundary values of σ are prescribed from $A = 4\pi e^{\sigma}|_{\theta=0}$.

4.1. Proof from harmonic maps

In this section, we prove lemmas 3.2 and 3.4 by exploiting the connection between $\mathcal{M}$ and harmonic energy for maps from the sphere into the complex hyperbolic space. The first lemma follows closely the arguments given by Acena *et al* [1]. To prove the rigidity in inequality (1), we use certain properties of the distance between harmonic maps in the complex hyperbolic space.

Proof. (*Lemma* 3.2)

To prove our claim, we follow the lines and arguments of [1] and refer to that article for more details. The key points in the argument are the following:

1. The extreme Kerr–Newman sphere, i.e. the set $\mathcal{D}_0$, satisfies the Euler–Lagrange equations for the functional $\mathcal{M}$:

$$\Delta\sigma - 2 = -\frac{(D\omega + 2\chi D\psi - 2\psi D\chi)^2}{\eta^2} - \frac{2}{\eta}((D\psi)^2 + (D\chi^2))^2, \quad (53)$$

$$D_a\left(\frac{D^a\omega + 2\chi D^a\psi - 2\psi D^a\chi}{\eta^2}\right) = 0 \quad (54)$$

$$D_a\left(\frac{D^a\chi}{\eta}\right) - \frac{1}{\eta^2}D_a\psi(D^a\omega + 2\chi D^a\psi - 2\psi D^a\chi) = 0 \quad (55)$$

$$D_a\left(\frac{D^a\psi}{\eta}\right) + \frac{1}{\eta^2}D_a\chi(D^a\omega + 2\chi D^a\psi - 2\psi D^a\chi) = 0, \quad (56)$$

   where indices are moved with the standard round metric on S^2.
2. The functional $\mathcal{M}$ is related to the harmonic energy $\tilde{\mathcal{M}}_\Omega$ for maps $(\eta, \omega, \chi, \psi)$ from a subset $\Omega \subset S^2 \setminus \{\theta = 0, \pi\}$ into the complex hyperbolic space $\mathbb{H}^2_{\mathbb{C}}$ which is equipped with the metric

$$g_H = \frac{d\eta^2}{\eta^2} + \frac{(d\omega + 2\chi d\psi - 2\psi d\chi)^2}{\eta^2} + 4\frac{d\chi^2 + d\psi^2}{\eta}, \quad (57)$$

and is given by

$$\tilde{\mathcal{M}}_\Omega = \frac{1}{2\pi}\int_\Omega \frac{|D\eta|^2}{\eta^2} + \frac{|D\omega + 2\chi D\psi - 2\psi D\chi|^2}{\eta^2} + 4\frac{|D\chi|^2 + |D\psi|^2}{\eta}\mathrm{d}S_0. \tag{58}$$

Now restrict the integral in the definition of $\mathcal{M}$, (23) to compact regions with the smooth boundary $\Omega \subset S^2 \setminus \{\theta = 0, \pi\}$ and denote the resulting functional as $\mathcal{M}_\Omega$. We have

$$\tilde{\mathcal{M}}_\Omega = \mathcal{M}_\Omega + 4\int_\Omega \ln\sin\theta \mathrm{d}S + \oint_{\partial\Omega}(4\sigma + \ln\sin\theta)\frac{\partial \ln\sin\theta}{\partial n}\mathrm{d}l, \tag{59}$$

where n is the exterior unit normal to the boundary $\partial\Omega$ of Ω and $\mathrm{d}l$ is the measure element on $\partial\Omega$. Since the difference between $\tilde{\mathcal{M}}_\Omega$ and $\mathcal{M}_\Omega$ is a constant plus a boundary term, both functionals have the same Euler–Lagrange equations.

3. A result of Hildebrandt *et al* [32] states that if the domain for the map is compact, is connected, with a non-void boundary and the target manifold has a negative sectional curvature, then a minimizer of the harmonic energy with Dirichlet boundary conditions exists, is unique, is smooth and satisfies the associated Euler–Lagrange equations. That is, harmonic maps are minimizers of the harmonic energy for given Dirichlet boundary conditions.

With the above comments, the proof goes as follows: divide the sphere into three regions as indicated in equations (60). Use a partition function to interpolate the potentials between extreme Kerr–Newman solution in region Ω_{I} and a general solution in region Ω_{III}. This gives a Dirichlet problem in region $\Omega_{\mathrm{IV}} = \Omega_{\mathrm{II}} \cup \Omega_{\mathrm{III}}$, which implies, by point 3. above, that the mass functional for extreme Kerr–Newman is less than or equal to the mass functional for the auxiliary interpolating map in the whole sphere. Finally, we take the limit as Ω_{III} covers the whole sphere and show that the mass functional for the auxiliary maps converges to the mass functional for the original general set.

After giving this general discussion about the proof, we begin with the splitting of the sphere according to

$$\Omega_I = \{\sin\theta \leqslant \mathrm{e}^{-(\log\epsilon)^2}\}, \quad \Omega_{II} = \{\mathrm{e}^{-(\log\epsilon)^2} \leqslant \sin\theta \leqslant \epsilon\}, \quad \Omega_{III} = \{\epsilon \leqslant \sin\theta\}, \tag{60}$$

where $0 < \epsilon < 1$. We also define the region $\Omega_{\mathrm{IV}} = \Omega_{\mathrm{II}} \cup \Omega_{\mathrm{III}}$.

Let $f : \mathbb{R} \to \mathbb{R} \in C^\infty(\mathbb{R})$, $0 \leqslant f \leqslant 1$, be the partition function defined as

$$f(t) = 1 \quad \text{for} \quad t \leqslant 1, \qquad f(t) = 0 \quad \text{for} \quad 2 \leqslant t, \qquad \left|\frac{\mathrm{d}f}{\mathrm{d}t}\right| \leqslant 1 \tag{61}$$

and f_ϵ be

$$f_\epsilon(\rho) = f(t_\epsilon(\rho)), \qquad t_\epsilon(\rho) = \frac{\log(-\log\rho)}{\log(-\log\epsilon)}, \qquad \rho \leqslant 1. \tag{62}$$

Therefore, we have

$$f_\epsilon(\rho) = 0 \quad \text{for} \quad \rho \leqslant \mathrm{e}^{-(\log\epsilon)^2}, \qquad f_\epsilon(\rho) = 1 \quad \text{for} \quad \rho \geqslant \epsilon \tag{63}$$

and

$$\lim_{\epsilon\to 0}\int_0^\infty |\partial_\rho f_\epsilon|^2 \rho \mathrm{d}\rho = 0. \tag{64}$$

Now we define the interpolating functions. Let u represent any of the variables σ, ω, χ, ψ and u_0 represent any of the variables σ_0, ω_0, χ_0, ψ_0 corresponding to the extreme Kerr–Newman sphere set with the same angular momentum and charges. We define u_ϵ to be

$$u_\epsilon = f_\epsilon(\sin\theta)\,u + (1 - f_\epsilon(\sin\theta))\,u_0 = (u - u_0)f_\epsilon(\sin\theta) + u_0. \tag{65}$$

This gives $u_\epsilon|_{\Omega_{\rm I}} = u_0|_{\Omega_{\rm I}}$ and $u_\epsilon|_{\Omega_{\rm III}} = u|_{\Omega_{\rm III}}$. We also define

$$\mathcal{M}^\epsilon = \mathcal{M}(\sigma_\epsilon, \omega_\epsilon, \psi_\epsilon, \chi_\epsilon), \tag{66}$$

and correspondingly $\mathcal{M}^\epsilon_\Omega$ and $\tilde{\mathcal{M}}^\epsilon_\Omega$ when the domain of integration is restricted to some region Ω. We also denote by a superscript '0' these quantities calculated for u_0.

We have all the ingredients needed to make use of the result in [32]. For this, let us consider now a fixed value of ϵ, and the functions $(\sigma, \omega, \chi, \psi)$ on the set $\Omega_{\rm IV}$. By [32], we know that there exists one and only one set of functions that minimizes $\tilde{\mathcal{M}}$ on $\Omega_{\rm IV}$ for given boundary data, and that this function satisfies the Euler–Lagrange equations of $\tilde{\mathcal{M}}$ on $\Omega_{\rm IV}$. By construction of u_ϵ, we have that u_ϵ and u_0 have the same boundary values on $\Omega_{\rm IV}$,

$$u_\epsilon|_{\partial\Omega_{\rm IV}} = u_0|_{\partial\Omega_{\rm IV}}. \tag{67}$$

As we already know that u_0 is a solution of the Euler–Lagrange equations of $\mathcal{M}$, and thus of $\tilde{\mathcal{M}}$ there, then u_0 is the only minimizer of $\tilde{\mathcal{M}}$ on $\Omega_{\rm IV}$ with these boundary conditions. This means that $\tilde{\mathcal{M}}^\epsilon_{\Omega_{\rm IV}} \geqslant \tilde{\mathcal{M}}^0_{\Omega_{\rm IV}}$. Both $\mathcal{M}$ and $\tilde{\mathcal{M}}$ are well defined on $\Omega_{\rm IV}$, and by (59) their difference is just a constant. Therefore, we also have $\mathcal{M}^\epsilon_{\Omega_{\rm IV}} \geqslant \mathcal{M}^0_{\Omega_{\rm IV}}$.

As we have already noted, $u_\epsilon|_{\Omega_{\rm I}} = u_0|_{\Omega_{\rm I}}$, and therefore $\mathcal{M}^\epsilon_{\Omega_{\rm I}} = \mathcal{M}^0_{\Omega_{\rm I}}$. This together with the inequality in $\Omega_{\rm IV}$ found above and the fact that $S^2 = \Omega_{\rm I} \cup \Omega_{\rm IV}$ give

$$\mathcal{M}^\epsilon \geqslant \mathcal{M}^0. \tag{68}$$

Only the last step of the proof is lacking, that is, to show

$$\lim_{\epsilon\to 0} \mathcal{M}^\epsilon = \mathcal{M}. \tag{69}$$

We write

$$\mathcal{M}^\epsilon = \mathcal{M}^\epsilon_{\Omega_{\rm I}} + \mathcal{M}^\epsilon_{\Omega_{\rm II}} + \mathcal{M}^\epsilon_{\Omega_{\rm III}} = \mathcal{M}^0_{\Omega_{\rm I}} + \mathcal{M}^\epsilon_{\Omega_{\rm II}} + \mathcal{M}_{\Omega_{\rm III}}. \tag{70}$$

Using the dominated convergence theorem, it is not hard to see that the first integral in (70) vanishes in the limit $\epsilon \to 0$ since the domain reduces to the poles and we know that $\mathcal{M}_0$ is finite. Also, the third term in (70) tends to $\mathcal{M}$ as $\Omega_{\rm III}$ extends to cover the whole sphere.

To show that the second term in (70) vanishes in the limit $\epsilon \to 0$, we consider its different parts separately. We have

$$\mathcal{M}^\epsilon_{\Omega_{\rm II}} = \frac{1}{2\pi}\int_{\Omega_{\rm II}} \left[|D\sigma_\epsilon|^2 + 4\sigma_\epsilon + \frac{|D\omega_\epsilon + 2\chi_\epsilon D\psi_\epsilon - 2\psi_\epsilon D\chi_\epsilon|^2}{\mathrm{e}^{2\sigma_\epsilon}\sin^4\theta} \right.$$
$$\left. +4\frac{|D\psi_\epsilon|^2 + |D\chi_\epsilon|^2}{\mathrm{e}^{\sigma_\epsilon}\sin^2\theta}\right] \mathrm{d}S_0. \tag{71}$$

Using the definition of u_ϵ (65), we compute

$$Du_\epsilon = (u - u_0)Df_\epsilon + (Du - Du_0)f_\epsilon + Du_0. \tag{72}$$

We see that

$$\sigma_\epsilon \leqslant C \tag{73}$$

because σ and σ_0 are finite on S^2, and $f_\epsilon \leqslant 1$. Here, and in what follows, we denote by C, C_i constants independent of ϵ. Also, because of the regularity of σ and σ_0, we have

$$|D\sigma_\epsilon|^2 \leqslant 3|Df_\epsilon|^2(\sigma - \sigma_0)^2 + 3|D\sigma - D\sigma_0|^2 + 3|D\sigma_0| \leqslant C_1|Df_\epsilon|^2 + C_2. \tag{74}$$

Then, from (73) and (74) and using strongly the property (64) to bound the integral of $|Df_\epsilon|^2$, we conclude that the first two terms in $\mathcal{M}^\epsilon_{\Omega_I I}$ go to zero as $\epsilon \to 0$.

Now we work with the term

$$\int_{\Omega_{\rm II}} \frac{|D\omega_\epsilon + 2\chi_\epsilon D\psi_\epsilon - 2\psi_\epsilon D\chi_\epsilon|^2}{\mathrm{e}^{2\sigma_\epsilon}\sin^4\theta} \mathrm{d}S_0. \tag{75}$$

Using the fact that f_ϵ is bounded, and σ, σ_0 are regular, we have

$$\frac{|D\omega_\epsilon + 2\chi_\epsilon D\psi_\epsilon - 2\psi_\epsilon D\chi_\epsilon|^2}{e^{2\sigma_\epsilon}\sin^4\theta} \leqslant C_1\frac{|Df|^2\,(\omega-\omega_0+2\psi_0\chi-2\chi_0\psi)^2}{\sin^4\theta} + C_2\frac{|D\omega+2\chi_0 D\psi-2\psi_0 D\chi|^2}{\sin^4\theta} + C_3\frac{|D\omega_0+2\chi_0 D\psi_0-2\psi_0 D\chi_0|^2}{\sin^4\theta} + C_4\frac{|D\psi_0|^2(\chi-\chi_0)^2}{\sin^4\theta} + C_5\frac{|D\chi_0|^2(\psi-\psi_0)^2}{\sin^4\theta} + C_6\frac{(\chi-\chi_0)^2|D\psi-D\psi_0|^2}{\sin^4\theta} + C_7\frac{|D\chi-D\chi_0|^2(\psi-\psi_0)^2}{\sin^4\theta}. \tag{76}$$

The term accompanying the constant C_3 is also point-wise bounded in $\Omega_{\rm II}$ because the extreme Kerr–Newman sphere satisfies the regularity item (ii) in definition 3.3.

In virtue of definition 3.3, we find that the remaining terms in (76) are uniformly bounded in $\Omega_{\rm II}$. Altogether we derive

$$\int_{\Omega_{\rm II}}\frac{|D\omega_\epsilon + 2\chi_\epsilon D\psi_\epsilon - 2\psi_\epsilon D\chi_\epsilon|^2}{e^{2\sigma_\epsilon}\sin^4\theta}\mathrm{d}S_0 \leqslant \int_{\Omega_{\rm II}} C_1|Df_\epsilon|^2 + C_2\mathrm{d}S_0. \tag{77}$$

It is important to remark that the lhs in the above inequality is bounded when the potentials ω, ψ, χ are smooth functions on S^2 satisfying condition (i) in definition 3.3 and are such that $\mathcal{M}$ is finite, that is, condition (ii) is no longer necessary.

In the limit $\epsilon \to 0$, this integral vanishes by property (64).

Finally, we look at the term

$$\int_{\Omega_{\rm II}}\frac{|D\chi_\epsilon|^2 + |D\psi_\epsilon|^2}{e^{\sigma_\epsilon}\sin^2\theta}\mathrm{d}S_0. \tag{78}$$

We have, as in (74),

$$\frac{|D\chi_\epsilon|^2}{e^{\sigma_\epsilon}\sin^2\theta} \leqslant C\frac{|Df|^2(\chi-\chi_0)^2 + |D\chi - D\chi_0|^2 + |D\chi_0|^2}{\sin^2\theta} \leqslant C_1|Df_\epsilon|^2 + C_2, \tag{79}$$

where, in the first inequality, we have used the boundedness of σ, σ_0. A similar behavior is found for the second term in (78). Therefore, taking into account the property (64), the limit $\epsilon \to 0$ of (78) gives zero.

We have shown

$$\lim_{\epsilon\to 0}\mathcal{M}^{\epsilon}_{\Omega_{\rm II}} = 0 \tag{80}$$

and thus the limit (69). This, together with (68), completes the proof of the lemma. □

Now we present the proof of lemma 3.4 stating the uniqueness of the minimizer for the area in inequality (1). This is done by exploiting the properties of the distance between harmonic maps in the complex hyperbolic space.

Proof. *(Lemma 3.4)*

We follow the lines of Weinstein [49] and Dain [19]. By contradiction, assume that there exists another regular set $\mathcal{D}^1$ which saturates (1). Denote with a superscript 1 the quantities referred to this set (and with a subscript 0 the quantities referred to $\mathcal{D}_0$). Then, we have

$$A_1 = 4\pi e^{\frac{\mathcal{M}_0 - 8}{8}} = A_0. \tag{81}$$

Then, using $A_1 \geqslant 4\pi e^{\frac{\mathcal{M}_1-8}{8}}$ and (81), we find $\mathcal{M}_1 = \mathcal{M}_0$. This means that $\mathcal{D}_1$ is also a critical point of the functional $\mathcal{M}$, i.e. it is a harmonic map.

By hypothesis, the second solution has the same values of the angular momentum and charges. Let Γ be the poles on S^2, so

$$\omega_1|_\Gamma = \omega_0|_\Gamma = \pm 4J, \quad \chi_1|_\Gamma = \chi_0|_\Gamma = \pm Q_{\rm M}, \quad \psi_1|_\Gamma = \psi_0|_\Gamma = \pm Q_{\rm E}. \tag{82}$$

But also, in virtue of equation (81), we conclude that $\sigma_1|_\Gamma = \sigma_0|_\Gamma = \ln(A_0/4\pi)$ (recall that the area is determined solely by the value of σ on the poles, through the expression $A = 4\pi e^{\sigma(0)}$).

In what follows, we will prove that the distance between these two solutions is in fact zero, and thus that the solutions are identical.

Let $(\eta_1, \omega_1, \chi_1, \psi_1)$ and $(\eta_0, \omega_0, \chi_0, \psi_0)$ be two harmonic maps $S^2 \setminus \Gamma \to \mathbb{H}^2_{\mathbb{C}}$, and consider, for each (θ, ϕ), the corresponding points in $\mathbb{H}^2_{\mathbb{C}}$ equipped with the hyperbolic metric introduced above in equation (57). The distance d between these two points is given by (see [49])

$$\cosh d = 1 + \delta, \tag{83}$$

where

$$\delta = \frac{(\omega_0 - \omega_1 + 2\chi_0\psi_1 - 2\chi_1\psi_0)^2 + ((\chi_0 - \chi_1)^2 + (\psi_0 - \psi_1)^2))^2}{2\eta_1\eta_0} + \left(\frac{1}{\eta_1} + \frac{1}{\eta_0}\right)[(\chi_0 - \chi_1)^2 + (\psi_0 - \psi_1)^2] + \frac{(\eta_0 - \eta_1)^2}{2\eta_1\eta_0}. \tag{84}$$

Therefore, since the functions ω, χ, ψ, σ are regular on S^2, d defines a function $d : S^2 \to \mathbb{R}$. We use the results of Schoen and Yau [44] to deduce that the square distance between harmonic maps is a subharmonic function on S^2, that is

$$\Delta d^2 \geqslant 0, \tag{85}$$

and, since δ is a convex function of d^2, then

$$\Delta \delta \geqslant 0. \tag{86}$$

Let us see now that the distance between the two solutions at Γ is zero. Begin with the first term in (84), from item (ii) in definition 3.3, the following behavior is deduced:

$$\left(\partial^2_\theta \omega + 2\chi\,\partial^2_\theta \psi - 2\psi\,\partial^2_\theta \chi\right)\big|_\Gamma = 0 \tag{87}$$

(note that if the functions ω, ψ, χ satisfy condition (i) in definition 3.3 and $\mathcal{M}$ is finite, then the solutions of the Euler–Lagrange equations of $\mathcal{M}$ necessarily have the above behavior near the poles).

Then, (87) together with the boundary conditions give near the poles

$$\omega_0 - \omega_1 + 2\chi_0\psi_1 - 2\chi_1\psi_0 = O(\sin^3\theta), \tag{88}$$

which implies

$$\left.\frac{(\omega_0 - \omega_1 + 2\chi_0\psi_1 - 2\chi_1\psi_0)^2}{2\eta_1\eta_0}\right|_\Gamma = 0. \tag{89}$$

We look now at the second and third terms in (84). Since by hypothesis $\partial_\theta\psi|_\Gamma = \partial_\theta\chi|_\Gamma = 0$, we find

$$\psi_0 - \psi_1 = \chi_0 - \chi_1 = O(\sin^2\theta). \tag{90}$$

Therefore, we obtain

$$\left.\left[\frac{[(\chi_0 - \chi_1)^2 + (\psi_0 - \psi_1)^2)]^2}{2\eta_1\eta_0} + \left(\frac{1}{\eta_1} + \frac{1}{\eta_0}\right)[(\chi_0 - \chi_1)^2 + (\psi_0 - \psi_1)^2]\right]\right|_\Gamma = 0. \tag{91}$$

The last term we must investigate in (84) is the one involving $(\eta_0 - \eta_1)$. We write it as

$$\frac{(\eta_0 - \eta_1)^2}{2\eta_0\eta_1} = \cosh(\sigma_0 - \sigma_1) - 1, \tag{92}$$

but taking into account the boundary conditions $\sigma_1|_\Gamma = \sigma_0|_\Gamma$, we find

$$\left.\frac{(\eta_0 - \eta_1)^2}{2\eta_0\eta_1}\right|_\Gamma = 0. \tag{93}$$

With conditions (89), (91) and (93), one verifies that

$$\delta|_\Gamma = 0. \tag{94}$$

Then, since δ is continuous (and smooth on $S^2 \setminus \{0\}$) and non-negative, $\delta\big|_\Gamma = 0$, and $\Delta\delta \geqslant 0$ on $S^2 \setminus \{0\}$, we can use the standard maximum principle to conclude that $\delta = 0$ in S^2. Therefore, $d = 0$ and the two maps are identical. This completes the proof of the lemma. □

4.2. Proof from geodesics in $\mathbb{H}^2_{\mathbb{C}}$

We prove now lemmas 3.2 and 3.4 in the axially symmetric case, with zero magnetic charge, namely $Q_{\mathrm{M}} = 0$. We make $Q = Q_{\mathrm{E}}$. The case when magnetic charge is present can be easily obtained by rotating along the (χ, ψ) plane, noting that rotations along the (χ, ψ) plane leave the functional $\mathcal{M}$, (23) invariant. We assume that either J or Q is non-zero; otherwise, there is nothing to prove.

The fundamental fact allowing to prove lemmas 3.2–3.4 only in terms of geodesics in the complex hyperbolic plane $\mathbb{H}^2_{\mathbb{C}}$ is the following identity (see equation (59)):

$$\tilde{\mathcal{M}}_{t_1,t_2} = \mathcal{M}_{\theta_1,\theta_2} + 4\sigma\cos\theta\Big|_{\theta_1}^{\theta_2} + 4\ln\tan\frac{\theta}{2}\Big|_{\theta_1}^{\theta_2}, \tag{95}$$

where $t = \ln\tan\frac{\theta}{2}$ and

$$\mathcal{M}_{\theta_1,\theta_2} := \int_{\theta_1}^{\theta_2}\left(\sigma'^2 + 4\sigma + \frac{(\omega' + 2\chi\psi' - 2\psi\chi')^2}{\eta^2} + 4\frac{\psi'^2 + \chi'^2}{\eta}\right)\sin\theta\,\mathrm{d}\theta, \tag{96}$$

$$\tilde{\mathcal{M}}_{t_1,t_2} := \int_{t_1}^{t_2}\left(\frac{\dot\eta^2}{\eta^2} + \frac{(\dot\omega + 2\chi\dot\psi - 2\psi\dot\chi)^2}{\eta^2} + 4\frac{\dot\chi^2 + \dot\psi^2}{\eta}\right)\mathrm{d}t. \tag{97}$$

Equation (95) shows that for fixed Dirichlet boundary conditions, critical points of $\mathcal{M}_{\theta_1,\theta_2}$ are critical points of $\tilde{\mathcal{M}}_{t_1,t_2}$ and vice versa. Now, consider $\gamma(t) = (\eta, \omega, \psi, \chi)(t)$ in $\mathbb{H}^2_{\mathbb{C}}$ with metric g_H given by (57). Then, we have the remarkable relation

$$\tilde{\mathcal{M}}_{t_1,t_2} = \int_{t_1}^{t_2} g_H(\dot\gamma, \dot\gamma)\,\mathrm{d}t, \tag{98}$$

which shows that critical points of the latter functional are geodesics in the complex hyperbolic plane up to an affine transformation, namely $\gamma(t) = \xi(\alpha t + \beta)$ with $\xi(s)$ being a geodesic parametrized by the arc length s. Moreover, because

$$\tilde{\mathcal{M}}_{t_1,t_2}(\gamma) \geqslant \frac{\mathrm{length}^2_{\mathbb{H}^2_{\mathbb{C}}}(\gamma)}{t_2 - t_1} \geqslant \frac{\mathrm{dist}^2_{\mathbb{H}^2_{\mathbb{C}}}(\gamma_1, \gamma_2)}{t_2 - t_1}, \tag{99}$$

global minimizers are exactly those critical points $\gamma(t) = \xi(\alpha t + \beta)$ for which ξ is a length minimizing geodesic between $\gamma(t_1)$ and $\gamma(t_2)$.

The following lemma, which is constructed on the previous observation, is central to prove lemmas 3.2–3.4. The proofs are given afterward.

Lemma 4.1.

(1) There exists a unique smooth minimizer $\mathcal{D} = (\sigma, \omega, \psi, \chi)$ for the functional $\mathcal{M}_{\theta_1,\theta_2}$ with given Dirichlet boundary conditions $\mathcal{D}(\theta_1)$, $\mathcal{D}(\theta_2)$. Moreover $\gamma(t) = (\eta, \omega, \psi, \chi)(t) = \xi(\alpha t + \beta)$, where $\xi(s)$ is a geodesic of $\mathbb{H}^2_{\mathbb{C}}$ parametrized by the arc length s.

(2) *The general expression for the unique minimizer of* $\tilde{\mathcal{M}}_{t_1,t_2=-t_1}$ *with centered boundary data* $(\eta, \omega, \psi, \chi)|_{t_1} = (\eta, -\omega, -\psi, -\chi)|_{t_2=-t_1}$, $\chi(t_1) = 0$, $\omega(t_1) \neq 0$, *is given by*

$$\eta = \left(\frac{1}{2}c_5^2 + \frac{1}{2}\sqrt{4c_1^2 + c_5^4}\cosh\alpha t\right)^{-1}, \tag{100}$$

$$\chi + \mathrm{i}\psi = c_4 e^{if} + \frac{c_3}{c_1}, \quad \text{with} \quad f = -2\arctan\left(\frac{\sqrt{4c_1^2 + c_5^4} - c_5^2}{2c_1}\tanh\frac{\alpha t}{2}\right), \tag{101}$$

$$\omega = -\frac{\alpha}{2c_1}\sqrt{1 - c_1^2\eta^2 - 4c_1^2c_4^2\eta} - 2\alpha\psi\chi + \frac{\alpha c_3}{c_1}\psi, \tag{102}$$

where $c_1 \neq 0$, c_3, c_4 *and* $c_5 = 2c_1c_4$ *are constants uniquely determined by the boundary conditions at* t_1 *and* $-t_1$.

(3) *For any positive sequence* $\theta^i \to 0$ *and sequence* $\{(\sigma_1^i, \omega_1^i, \chi_1^i, \psi_1^i)\}$, *such that*

$$\lim \sigma_1^i = \sigma_l \neq \infty, \quad \lim \omega_1^i = -4J, \quad \lim \psi_1^i = -Q := -Q_{\mathrm{E}} \quad \text{and,} \tag{103}$$

$$\chi_1^i = 0, \quad \omega_1^i \neq 0, \quad \text{for all } i, \tag{104}$$

the unique minimizer $\mathcal{D}^i(\theta) = (\sigma^i, \omega^i, \chi^i, \psi^i)$ *of* $\mathcal{M}_{\theta^i,\pi-\theta^i}$ *with boundary data*

$$\mathcal{D}^i(\theta^i) = \left(\sigma_1^i, \omega_1^i, \chi_1^i, \psi_1^i\right), \quad \mathcal{D}^i(\pi - \theta^i) = \left(\sigma_1^i, -\omega_1^i, -\chi_1^i, -\psi_1^i\right) \tag{105}$$

has

$$\sigma^i(\theta) = -\ln\left[\tfrac{1}{2}\left(c_5^i\right)^2\sin^2\theta + \tfrac{1}{2}\sqrt{4\left(c_1^i\right)^2 + \left(c_5^i\right)^4}\sin^2\theta\cosh\alpha^i t\right], \tag{106}$$

for constants c_1^i, c_5^i, α^i *and where, as before,* $t = \ln\tan\frac{\theta}{2}$. *Moreover as* $\theta^i \to 0$

$$\lim\alpha^i = 2, \quad \lim\ c_5^i = \frac{-Q}{\sqrt{Q^4 + 4J^2}}, \quad \lim\ c_1^i = \frac{J}{Q^4 + 4J^2}, \tag{107}$$

and if we write

$$\sigma_l = \tfrac{1}{2}\ln(Q^4 + 4J^2) + \Gamma, \tag{108}$$

then

$$\lim\left(\frac{\theta^i}{2}\right)^{\alpha^i - 2} = \mathrm{e}^{\Gamma}. \tag{109}$$

Proof.

(1) As they differ by a constant, a global minimizer for $\mathcal{M}_{\theta_1,\theta_2}$ is a global minimizer for $\tilde{\mathcal{M}}_{t_1,t_2}$. Moreover, as explained above, the latter is of the form $\gamma(t) = \xi(\alpha t + \beta)$ with $\xi(s)$ being a geodesic parametrized by the arc length s and realizing the distance $\mathrm{dist}_{\mathbb{H}^2_{\mathbb{C}}}(\gamma(t_1), \gamma(t_2))$ between $\gamma(t_1)$ and $\gamma(t_2)$. If $\xi(s = 0) = \gamma(t_1)$ and $\xi(s = \mathrm{dist}_{\mathbb{H}^2_{\mathbb{C}}}) = \gamma(t_2)$ (which can always be chosen by redefining s if necessary), then α and β are unique and determined by t_1, t_2 and $\mathrm{dist}_{\mathbb{H}^2_{\mathbb{C}}}(\gamma(t_1), \gamma(t_2))$. But because $\mathbb{H}^2_{\mathbb{C}}$ has a negative sectional curvature and is simply connected, then between two different points $\gamma(t_1)$ and $\gamma(t_2)$, there is always a unique minimizing geodesic $\xi(s)$, with $\xi(0) = \gamma(t_1)$ and $\xi(\mathrm{dist}_{\mathbb{H}^2_{\mathbb{C}}}(\gamma(t_1), \gamma(t_2))) = \gamma(t_2)$. It follows that the global minimizer of $\mathcal{M}_{\theta_1,\theta_2}$ exists, is unique and has the desired form.

(2) We describe how to obtain a general expression for the unique minimizers of $\tilde{\mathcal{M}}_{t_1,t_2=-t_1}$ whose boundary data satisfy

$$(\sigma(t_1), \omega(t_1), \psi(t_1), \chi(t_1)) = (\sigma(t_2), -\omega(t_2), -\psi(t_2), -\chi(t_2)), \;\; \chi(t_1) = 0, \tag{110}$$

with $\omega(t_1) \neq 0$. The Euler–Lagrange equations for $\tilde{\mathcal{M}}_{t_1,t_2}$ are integrable and the first integrals can be obtained as conserved quantities of the form $g_H(X, \dot{\gamma})$ which arise from Killing fields X^a for g_H. The Killing fields we will use are

$$X_1 = \partial_\omega, \qquad X_2 = -2\psi\partial_\omega + \partial_\chi, \qquad X_3 = 2\chi\partial_\omega + \partial_\psi. \tag{111}$$

The corresponding conserved quantities can be combined to give

$$\frac{\dot{\omega} + 2\chi\dot{\psi} - 2\psi\dot{\chi}}{\eta^2} = \alpha c_1, \qquad \alpha c_1\psi - \frac{\dot{\chi}}{\eta} = \alpha c_2, \qquad \alpha c_1\chi + \frac{\dot{\psi}}{\eta} = \alpha c_3, \tag{112}$$

where c_1, c_2 and c_3 are constants and we have inserted explicitly the (positive) constant α (introduced in item 1 before). Note that $c_1 \neq 0$ for if $c_1 = 0$, then (110) and (112) imply ω identically zero which contradicts $\omega(t_1) \neq 0$. To obtain the equation for η (or equivalently, for σ), we use $g_H(\dot{\gamma}, \dot{\gamma}) = \alpha^2$ and thus

$$\frac{\dot{\eta}^2}{\eta^2} + \frac{(\dot{\omega} + 2\chi\dot{\psi} - 2\psi\dot{\chi})^2}{\eta^2} + 4\frac{\dot{\chi}^2 + \dot{\psi}^2}{\eta} = \alpha^2. \tag{113}$$

Equations (112) and (113) are indeed equivalent to the equations of motion obtained from the variation of (96), cf (53)–(56). These equations can be further simplified by using an important property of the variables (ψ, χ). By making $\bar{\psi} = \psi - c_2/c_1$ and $\bar{\chi} = \chi - c_3/c_1$, the second and third equations in (112) reduce to

$$\alpha c_1\bar{\psi} - \frac{\dot{\bar{\chi}}}{\eta} = 0, \qquad \alpha c_1\bar{\chi} + \frac{\dot{\bar{\psi}}}{\eta} = 0. \tag{114}$$

Multiplying these two equations, respectively, by $\bar{\chi}$ and $\bar{\psi}$ and subtracting one from another, we obtain $\bar{\chi}\dot{\bar{\chi}} + \bar{\psi}\dot{\bar{\psi}} = 0$ which implies $\bar{\chi}^2 + \bar{\psi}^2 = c_4^2$, where c_4 is a constant. We write

$$\bar{\chi} + i\bar{\psi} = c_4 e^{if}, \qquad \text{with} \qquad \dot{f} = -c_1\alpha\eta. \tag{115}$$

Now, since $\dot{\bar{\chi}} = \dot{\chi}$ and $\dot{\bar{\psi}} = \dot{\psi}$, then

$$\dot{\chi}^2 + \dot{\psi}^2 = \alpha^2 c_4^2 c_1^2 \eta^2. \tag{116}$$

We use equations (112) and (116) to rewrite (113) as

$$\frac{\dot{\eta}^2}{\eta^2} + \alpha^2 c_1^2\eta^2 + 4\alpha^2 c_1^2 c_4^2\eta = \alpha^2. \tag{117}$$

We now solve equation (117) for η and use $\eta(t_1) = \eta(t_2)$, $t_1 = -t_2$, and find (100) with $c_5^2 := 4c_1^2 c_4^2$.

Now we solve for ψ, χ. Note that in order to have $\chi(t_1) = \chi(t_2) = 0$ and $\psi(t_1) = -\psi(t_2)$ and at the same time $\bar{\psi}^2 + \bar{\chi}^2 = c_4^2$, the only possibility is to have $c_2 = 0$ and therefore

$$\chi + i\psi = c_4 e^{if} + \frac{c_3}{c_1}. \tag{118}$$

We obtain f by integrating the second equation in line (115), using (100) and $f(t_1) = -f(t_2)$, to find (101). To find ψ and χ, use (118) where c_3 is adjusted from c_1, c_5, α to have $\chi(t_1) = \chi(t_2) = 0$. To find ω on the other hand, one could use the expression for $\dot{\omega}$ in (112) and integrate. However, a direct and simple expression for ω arises when using the conserved quantity associated with the Killing field $X_4 = 2\eta\partial_\eta + 2\omega\partial_\omega + \chi\partial_\chi + \psi\partial_\psi$. Explicitly,

$$g_H(X_4, \dot\gamma) = \frac{\dot\eta}{2\eta} + \frac{c_1\alpha}{2}\omega + \frac{\chi\dot\chi + \psi\dot\psi}{\eta} = c. \tag{119}$$

Noting that the above expression is antisymmetric in t around $t = 0$, we deduce that the constant c is zero. Then, from (119), one obtains a direct expression for ω. The expressions that one obtains for ω, ψ, χ in this form are somehow crude but serve well to the purposes of the proof of (3). Summarizing, given $c_1 \neq, c_5, \alpha, \theta_1$ one can associate, following the construction above, a unique solution $\gamma(t)$ satisfying (110) with $-t_2 = t_1 = \ln\tan\frac{\theta_1}{2}$.

(3) *A priori*, to prove item 3, one could calculate the constants (c_1^i, c_5^i, α^i) from the prescribed boundary data at $\theta^i, \pi - \theta^i$ and prove from them, by a direct calculation, the conclusions (107) and (109). Unfortunately, such a procedure is a computational nuisance. For this reason, we follow an alternative argument. Given $\theta_1 > 0$, consider the map $F_{\theta_1} : \mathbb{R}^3 \setminus \{y = 0\} \to \mathbb{R}^3$ that to (Γ, c_1, c_5) associates the boundary values $(\sigma(\theta_1), \omega(\theta_1), \psi(\theta_1))$ of the solution $(\sigma, \omega, \psi, \chi)$ found from the constants $(c_1, c_5, \alpha, \theta_1)$ where α is given by

$$\alpha = 2 + \frac{\Gamma}{\ln\theta_1/2}. \tag{120}$$

Then, if we let $\theta_1 \to 0$, from (100) and the limit $\lim_{\theta_1\to 0}\sin^2\theta_1\cosh(\alpha\ln\frac{\theta_1}{2}) = \frac{2}{\Gamma}$, we obtain $\lim\sigma(\theta_1) = \Gamma - \frac{1}{2}\ln(4c_1^2 + c_5^4)$. Next, using that $\psi(\theta_1) = c_4\sin f(\theta_1)$ and that $f(\theta_1) \to \arctan 2c_1/c_5^2$, we find $\lim\psi(\theta_1) = \frac{c_5}{\sqrt{4c_1^2+c_5^4}}$. Finally, from (119), (113) and the expression

$$\left.\frac{\psi\dot\psi}{\eta}\right|_{\theta_1} = -c_1\alpha c_4^2\sin f\cos f\Big|_{\theta 1} \to -\frac{c_5^4}{4c_1^2 + c_5^4}, \tag{121}$$

we obtain $\lim\omega(\theta_1) = \frac{-4c_1}{4c_1^2+c_5^4}$. This shows that the maps F_{θ_1} converge uniformly on any compact set to a map F_0 given by

$$F_0(\Gamma, c_1, c_5) = \left(\Gamma - \frac{1}{2}\ln\left(4c_1^2 + c_5^4\right), -\frac{4c_1}{4c_1^2 + c_5^4}, \frac{c_5}{4c_1^2 + c_5^4}\right). \tag{122}$$

Moreover, the map F_0 extends to a diffeomorphism from $\mathbb{R}^3 \setminus (\{y = 0\} \cap \{z = 0\})$ into $\mathbb{R}^3 \setminus (\{y = 0\} \cap \{z = 0\})$. A close inspection of the limits above shows that the maps F_θ do extend smoothly too. We note now that given the values $(\sigma_l, J, Q_E = Q, 0)$ (with either J or Q non-zero), if we take

$$\left(\Gamma^\infty, c_1^\infty, c_5^\infty\right) = \left(\sigma_l - \frac{1}{2}\ln[4J^2 + Q^4], \frac{J}{4J^2 + Q^4}, \frac{-Q}{\sqrt{4J^2 + Q^4}}\right), \tag{123}$$

then $F_0(\Gamma^\infty, c_1^\infty, c_5^\infty) = (\sigma_l, J, Q)$. It follows therefore from the above argument that given (σ_l, J, Q) (with either J or Q non-zero) and sequences $\{\theta^i \to 0\}$ and $\{(\sigma^i, \omega^i, \psi^i, \chi^i)\}$ as in the hypothesis of (3), there is a sequence $\{(\Gamma^i, c_1^i, c_5^i)\}$ with limit $(\Gamma^\infty, c_1^\infty, c_5^\infty)$ such that, $F(\Gamma^i, c_1^i, c_5^i) = (\sigma^i, \omega^i, \psi^i)$ and therefore the unique minimizer of $\mathcal{M}_{\theta_1^i, \pi-\theta_1^i}$ with boundary data (105) is the unique solution constructed out of $(c_1^i, c_5^i, \alpha^i, \theta_1^i)$, where $\alpha^i = 2 + \frac{\Gamma^i}{\ln\theta_1^i/2}$. Expressions (106), (107) and (109) are readily checked. This finishes (3) and the proof of the lemma. □

Proof. *(Lemma 3.2)* Consider any sequence $\{\theta_1^i \downarrow 0\}$. Now, we divide the interval $(0, \pi)$ in three parts, and write, for the set $\mathcal{D}$

$$\mathcal{M}(\mathcal{D}) = \mathcal{M}_{0,\theta_1^i} + \mathcal{M}_{\pi-\theta_1^i,\pi} + \mathcal{M}_{\theta_1^i,\pi-\theta_1^i}. \tag{124}$$

Then, we recall the relation between $\tilde{\mathcal{M}}_{\theta_1^i,\pi-\theta_1^i}$ and $\mathcal{M}_{\theta_1^i,\pi-\theta_1^i}$ which is

$$\tilde{\mathcal{M}}_{\theta_1^i,\pi-\theta_1^i}(\gamma) = \mathcal{M}_{\theta_1^i,\pi-\theta_1^i}(\mathcal{D}) + 4\sigma\cos\theta\Big|_{\theta_1^i}^{\pi-\theta_1^i} + 4\cos\theta\Big|_{\theta_1^i}^{\pi-\theta_1^i} + 4\ln\tan\frac{\theta}{2}\Big|_{\theta_1^i}^{\pi-\theta_1^i}, \tag{125}$$

where of course γ represents the same data as $\mathcal{D}$. Using this, we would like to obtain a sharp estimation from below to $\tilde{\mathcal{M}}_{\theta_1^i,\pi-\theta_1^i}$. For this, we proceed as follows. For every i, consider two points in $\mathbb{H}^2_{\mathbb{C}}$, denoted by $\bar{\gamma}_{\theta_1^i}$, $\bar{\gamma}_{\pi-\theta_1^i}$ and given by

$$\bar{\gamma}_{\theta_1^i} = \left(\eta(\theta_1^i), -4J, -Q_{\mathrm{E}}, -Q_{\mathrm{M}} = 0\right), \quad \bar{\gamma}_{\pi-\theta_1^i} = \left(\eta(\theta_1^i), +4J, +Q_{\mathrm{E}}, +Q_{\mathrm{M}} = 0\right) \tag{126}$$

if $J \neq 0$, while if $J = 0$, then we replace $4J$ in the above expressions by $\hat{\omega}_1^i$ tending to zero sufficiently fast (see below). This is because below we will need to use lemma 4.1 (item 3), for the minimizers with boundary data $\bar{\gamma}_{\theta_1^i}$ and $\bar{\gamma}_{\pi-\theta_1^i}$, but lemma 4.1 requires non-zero boundary values for ω. From the regularity at the poles, one easily deduces that (if $J \neq 0$, or if $\hat{\omega}_1^i$ goes to zero sufficiently fast) $\mathrm{dist}_{\mathbb{H}^2_{\mathbb{C}}}(\bar{\gamma}_{\theta_1^i}, \gamma(\theta_1^i)) \to 0$ (see expression (83)) and similarly for $\mathrm{dist}_{\mathbb{H}^2_{\mathbb{C}}}(\bar{\gamma}_{\pi-\theta_1^i}, \gamma(\pi-\theta_1^i))$. Consider now any another sequence $\{\theta_2^i \downarrow 0\}$, such that $\theta_2^i < \theta_1^i$ for every i and

$$\lim \frac{\mathrm{dist}^2_{\mathbb{H}^2_{\mathbb{C}}}\left(\gamma(\theta_1^i), \bar{\gamma}_{\theta_1^i}\right)}{t(\theta_2^i) - t(\theta_1^i)} = \lim \frac{\mathrm{dist}^2_{\mathbb{H}^2_{\mathbb{C}}}\left(\gamma(\pi-\theta_1^i), \bar{\gamma}_{\pi-\theta_1^i}\right)}{t(\theta_2^i) - t(\theta_1^i)} = 0, \tag{127}$$

$$\lim t(\theta_1^i) - t(\theta_2^i) = 0, \tag{128}$$

with, again, $t(\theta) = \ln\tan\frac{\theta}{2}$. Finally, consider the curve in $\mathbb{H}^2_{\mathbb{C}}$, denoted by $\bar{\gamma}^i$, starting at $\bar{\gamma}_{\theta_1^i}$ and ending at $\bar{\gamma}_{\pi-\theta_1^i}$ defined as

1. the minimizer of $\tilde{\mathcal{M}}_{t(\theta_2^i),t(\theta_1^i)}$, with boundary data $\bar{\gamma}_{\theta_1^i}$, $\gamma(t(\theta_1^i))$, if $t \in [t(\theta_2^i), t(\theta_1^i)]$,
2. $\gamma(t)$ if $t \in [t(\theta_1^i), t(\pi-\theta_1^i)]$,
3. the minimizer of $\tilde{\mathcal{M}}_{t(\pi-\theta_1^i),t(\pi-\theta_2^i)}$, with boundary data $\gamma(t(\pi-\theta_1^i))$, $\bar{\gamma}_{\pi-\theta_1^i}$, if $t \in [t(\pi-\theta_1^i), t(\pi-\theta_2^i)]$.

By (99), we can write

$$\tilde{\mathcal{M}}_{t(\theta_2^i),t(\pi-\theta_2^i)}(\bar{\gamma}^i) \geqslant \alpha_i^2\left(t(\pi-\theta_2^i) - t(\theta_2^i)\right) = -2\alpha_i^2\ln\tan\frac{\theta_2^i}{2}, \tag{129}$$

where α_i is the constant associated with the minimizer of $\tilde{\mathcal{M}}_{t(\theta_2^i),t(\pi-\theta_2^i)}$ with boundary data $\bar{\gamma}_{\theta_1^i}$, $\bar{\gamma}_{\pi-\theta_1^i}$, as in lemma 4.1. By (99), (127) and (128), we have

$$\lim \tilde{\mathcal{M}}_{t(\theta_2^i),t(\theta_1^i)}(\bar{\gamma}^i) = \lim \tilde{\mathcal{M}}_{t(\pi-\theta_1^i),t(\pi-\theta_2^i)}(\bar{\gamma}^i) = 0 \tag{130}$$

and finally, of course,

$$\tilde{\mathcal{M}}_{t(\theta_2^i),t(\pi-\theta_2^i)}(\bar{\gamma}^i) = \tilde{\mathcal{M}}_{t(\theta_2^i),t(\theta_1^i)}(\bar{\gamma}^i) + \tilde{\mathcal{M}}_{t(\pi-\theta_1^i),t(\pi-\theta_2^i)}(\bar{\gamma}^i) + \tilde{\mathcal{M}}_{t(\theta_1^i),t(\pi-\theta_1^i)}(\gamma). \tag{131}$$

Collecting (131) and (125) together with the information (128)–(130), we obtain

$$\lim \mathcal{M}_{\theta_1^i,\pi-\theta_1^i}(\mathcal{D}) \geqslant \lim\left[-2\left(\alpha_i^2 - 4\right)\ln\tan\frac{\theta_1^i}{2} - 4\sigma\cos\theta\Big|_{\theta_1^i}^{\pi-\theta_1^i} + 8\right]. \tag{132}$$

But as $\sigma(\theta_1^i) \to \sigma_I = \frac{1}{2}\ln(4J^2 + Q^4) + \Gamma$ with $\lim(\theta_1^i/2)^{\alpha_i^2-4} = \mathrm{e}^{\Gamma}$, we obtain, after a cancellation,

$$\mathcal{M}(\mathcal{D}) = \lim \mathcal{M}_{\theta_1^i,\pi-\theta_1^i} \geqslant 4\ln(Q^4 + 4J^2) + 8 = \mathcal{M}^0. \tag{133}$$

□

We present now the proof of lemma 3.4. This is achieved by making use of the explicit expression for the minimizers of the functional $\mathcal{M}$ with given boundary conditions found above.

Proof. *(Lemma 3.4)*. We know that any critical point of $\mathcal{M}$ is represented in terms of a geodesic $\gamma = (\eta, \omega, \psi, \chi)$ of $\mathbb{H}^2_{\mathbb{C}}$. Regularity implies $\lim \eta'/\eta = 2$ as θ tends to 0 or π. This implies from (117) that $\alpha = 2$. On the other hand, the boundary data imply

$$c_1 = \frac{J}{Q^4 + 4J^2}, \qquad c_5 = \frac{-Q}{\sqrt{Q^4 + 4J^2}}. \tag{134}$$

Then, manipulating (101) while using these values for c_1 and c_5, one obtains

$$f = \arctan \frac{2c\cos\theta}{1 - c^2\cos^2\theta}, \tag{135}$$

with $c = 2J/(\sqrt{4J^2 + Q^4} + Q^2)$. Plugging this expression in 'the law of the two arcs' (51) gives (26)–(28) with $Q_M = 0$. Equation (25) is obtained from (100). □

Acknowledgment

We are very grateful to Walter Simon and Sergio Dain for helpful discussions.

Appendix. Linking global and quasilocal axisymmetric inequalities

In this somehow more informal appendix, we want to show that there might exist a link between the AJQ and MJQ inequalities (see equations (1) and (2)). The MJQ inequality (2) is a global manifestation of the constraints (in the maximal spatial gauge) and in this sense it is a global inequality requiring knowledge of the system as a whole. The AJQ inequality (1) is instead of a quasilocal nature and does not require global information. Despite the different realms in which they manifest, they seem to be closely related. The link that we shall establish could be of help to prove quasilocal inequalities in systems other than Einstein–Maxwell, for which a global three-dimensional mass functional is shown to exist. More concretely, we will point out a relation among (1) and (2) by linking the two-dimensional energy functional $\mathcal{M}$ given in (23) and a three-dimensional energy functional given in [16, 17] (see below) , and whose minimization properties lead to the AJQ and MJQ inequalities, respectively.

Let us first put in parallel how one obtains the MJQ and AJQ inequalities from suitable functionals.

The inequality (2) has been established in [16, 17] following a similar argument as in [19]. The black hole configuration on which (2) has been proved is that of an initial datum with two asymptotically flat ends, where m, J, Q_E, Q_M in the inequality (2) refers to the mass, angular momentum and charges of a selected end. The rationale behind the proof of (2) is the following. One introduces a three-dimensional functional M, defined on such configurations and bounding the mass (of the selected end) from below, i.e. $m \geqslant M$. Moreover, one has $M \geqslant M_0$, where M_0 is the infimum of M among those configurations having (for the selected end) J, Q_E and Q_M fixed. Moreover, M_0 is achieved by the extreme Kerr–Newman solution. The inequalities $m \geqslant M$ and $M \geqslant M_0$ together with the explicit expression for M_0 give (2).

On the other hand, the rationale behind (1) that we have developed in the previous sections was the following. We introduced a functional $\mathcal{M}$ defined on a certain surface (i.e. stable MOTS or stable minimal surface over a maximal slice) and bounding its area from below, more precisely by $A \geqslant 4\pi e^{(\mathcal{M}-8)/8}$. Then, we showed that the extreme Kerr–Newman sphere realizes the absolute minimum of $\mathcal{M}$ among all configurations having J, Q_E and Q_M fixed. Denoting by $\mathcal{M}_0$ the value of $\mathcal{M}$ at the extreme Kerr–Newman sphere, we obtain $A \geqslant 4\pi e^{(\mathcal{M}_0-8)/8}$ which gives (1).

It is clear that the two procedures described above are formally similar and we will see that although both can be carried out without any reference to one another, they are indeed remarkably related. More precisely, we state that the inequality $m \geqslant M \geqslant M_0$ implies that the extreme Kerr–Newman sphere is a critical point of $\mathcal{M}$ (even more, it can be shown from this that the extreme Kerr–Newman sphere is a local minimum for $\mathcal{M}$). However, we do not know at the moment whether the fact that the extreme Kerr–Newman sphere is a global minimizer of $\mathcal{M}$ can be established solely from the inequality $m \geqslant M \geqslant M_0$. This gives a partial connection in the form MJQ $\Rightarrow$ AJQ. In the other direction, namely AJQ $\Rightarrow$ MJQ, we note that with the help of the Penrose inequality $A \leqslant 16\pi m^2$, one obtains for outermost minimal surfaces

$$m^2 \geqslant \frac{A}{16\pi} \geqslant \frac{4J^2+Q^4}{4}, \tag{A.1}$$

which is an inequality slightly worse than (2). Despite these interesting relations, many issues onto the link between the inequalities still remain in shadows.

In order to see how the first implication shows up, we begin by defining the three-dimensional potentials $\bar{\mathcal{D}} = (\bar\sigma, \bar\omega, \bar\psi, \bar\chi)$ over maximal electrovacuum initial data (Σ, h, K, F) and then introduce the functional M together with a crucial minimizing property. We follow [17] on this construction.

We write the spatial metric on the maximal initial datum in the form

$$h = e^{\bar\sigma+2\bar q}\left(\frac{d\bar r^2}{\bar r^2} + d\theta^2\right) + e^{\bar\sigma}\sin^2\theta\,(d\varphi + v_{\bar r}\,d\bar r + v_\theta\,d\theta)^2, \tag{A.2}$$

where $v_{\bar r}, v_\theta$ are functions of $\bar r, \theta$, which spans over $\mathbb{R}^3 \setminus \{0\}$ and defines $\bar\sigma$. For the electromagnetic fields, we have the relations

$$E_a = F_{ab}n^b, \qquad B_a = {}^*F_{ab}n^b, \tag{A.3}$$

$$\partial_a\bar\chi := F_{ab}\eta^b, \qquad \partial_a\bar\psi := {}^*F_{ab}\eta^b, \tag{A.4}$$

where n^a is the unit normal to Σ. These expressions define the potentials $\bar\psi$, $\bar\chi$. Finally, a potential $\bar\omega$ is defined through

$$D_a\bar\omega + 2\bar\chi D_a\bar\psi - 2\bar\psi D_a\bar\chi := 2\epsilon_{abc}K^b{}_d\eta^c\eta^d. \tag{A.5}$$

Observe that because the norm of the axial Killing vector η^a is null over the axis, the differential of the potentials $(\bar\omega, \bar\psi, \bar\chi)$ at the axis is zero and therefore their values remain constant all along them. As they are defined up to a constant, one can take them to be 'centered', namely $\bar\omega|_{\theta=\pi} = -\bar\omega|_{\theta=0}$, $\bar\psi|_{\theta=\pi} = -\bar\psi|_{\theta=0}$ and $\bar\chi|_{\theta=\pi} = -\bar\chi|_{\theta=0}$.

It is interesting and illustrative to see the relation between the potentials $\bar{\mathcal{D}}_0 = (\bar\sigma_0, \bar\omega_0, \bar\psi_0, \bar\chi_0)$ corresponding to the extreme Kerr–Newman solution over, say, the slice $\{t = 0\}$, and the potentials $\mathcal{D}_0 = (\sigma_0, \omega_0, \psi_0, \chi_0)$ defining the extreme Kerr–Newman sphere (25)–(28). The explicit form of the three-dimensional potentials $\bar{\mathcal{D}}_0$ can be found in [13] (see pp 197–204) and we have

$$\bar\sigma_0 = \ln\frac{(\bar r^2 - Q^2 + 2m_0(\bar r + m_0))^2 - \bar r^2 a_0^2\sin^2\theta}{\Sigma}, \tag{A.6}$$

$$2(\bar\sigma_0 + \bar q_0) = \ln\left[(\bar r^2 - Q^2 + 2m_0(\bar r + m_0))^2 - \bar r^2 a_0^2\sin^2\theta\right], \tag{A.7}$$

where $\bar r = r - r_H = r - m_0$. From this, we obtain

$$\lim_{\bar r\to 0}\bar\sigma_0(\bar r, \theta, \varphi) = \sigma_0(\theta, \varphi), \tag{A.8}$$

$$\lim_{\bar{r}\to 0} 2(\bar{\sigma}_0(\bar{r},\theta,\varphi) + \bar{q}_0(\bar{r},\theta,\varphi)) = \ln 4J^2 + Q^4 = \ln \frac{A^2}{16\pi^2} = 2c, \tag{A.9}$$

where σ_0 is given by (25) and (as before) $A = 4\pi e^c$. Together with (A.2), this shows that the $\{(\theta,\varphi)\}$ coordinates on the spheres $\{\bar{r} = \bar{r}_1\}$ become, as $\bar{r}_1 \to 0$, the unique ones for which the induced metric (over $\{\bar{r} = \bar{r}_1\}$) is expressed in the form (14). Moreover from (A.3) and (A.4), it is deduced that over any sphere $\{\bar{r} = \bar{r}_1\}$, it holds

$$E_\perp = \frac{e^{\bar{\sigma}_0+\bar{q}_0}\partial_\theta \bar{\psi}_0}{\sin\theta}, \qquad B_\perp = \frac{e^{\bar{\sigma}_0+\bar{q}_0}\partial_\theta \bar{\chi}_0}{\sin\theta}. \tag{A.10}$$

As explained in section 3.2, $E_\perp$ and $B_\perp$ converge as $\bar{r}_1 \to 0$ to those given by (41) and (42), respectively, that is those of the extreme Kerr–Newman sphere. From this, (17) and (A.9), we deduce that the limit of the potentials $\bar{\psi}$ and $\bar{\chi}$ over the spheres $\{\bar{r} = \bar{r}_1\}$ converges to (27) and (28), respectively, that is

$$\lim_{\bar{r}\to 0} \bar{\psi}(\bar{r},\theta,\varphi) = \psi_0(\theta,\varphi), \qquad \lim_{\bar{r}\to 0} \bar{\chi}(\bar{r},\theta,\varphi) = \chi_0(\theta,\varphi). \tag{A.11}$$

The same property is also obtained for $\bar{\omega}_0$

$$\lim_{\bar{r}\to 0} \bar{\omega}_0(\bar{r},\theta,\varphi) = \omega_0(\theta,\varphi). \tag{A.12}$$

Given the set of three-dimensional potentials $\bar{\mathcal{D}}$, we define, as done in [17], the energy functional M on $\bar{\mathcal{D}}$

$$M = \int_{\mathbb{R}^3} 4\left(|D\bar{U}|^2 + \frac{e^{4\bar{U}}}{\rho^4}\left|\frac{D\bar{\omega}/2 + \bar{\chi}D\bar{\psi} - \bar{\psi}D\bar{\chi}}{2}\right|^2 + \frac{e^{2\bar{U}}}{\rho^2}(|D\bar{\chi}|^2 + |D\bar{\psi}|^2)\right) dV_0, \tag{A.13}$$

where $e^{\bar{\sigma}} = e^{-2\bar{U}}\bar{r}^2$, $dV_0 = \bar{r}^2 \sin\theta\, d\bar{r}\, d\theta\, d\varphi$, D is the Euclidean differential and the norms are Euclidean norms.

Although in principle the functional M was introduced on axisymmetric maximal initial data with two asymptotically flat ends, we will consider it acting on more general sets of functions $\bar{\mathcal{D}}(\bar{\sigma}, \bar{\omega}, \bar{\psi}, \bar{\chi})$ with fixed J, Q_E and Q_M, not necessarily arising from the potentials of an initial state. In this setup, a key property of M, which is deduced from the arguments in [17], is the following. Let $\bar{\mathcal{D}} = (\bar{\sigma}, \bar{v}, \bar{\chi}, \bar{\psi})$ be a set that is the extreme Kerr–Newman set outside a compact set in $\mathbb{R}^3 \setminus \{0\}$. Then, $M(\bar{\mathcal{D}}) \geqslant M(\bar{\mathcal{D}}_0)$, where $\bar{\mathcal{D}}_0$ is the set for the extreme Kerr–Newman solution. In other words, the extreme Kerr–Newman set is a minimum of M under variations of $\bar{\mathcal{D}}_0$ of compact support.

We are ready to explain how to deduce that $\mathcal{D}_0$ is a critical point for $\mathcal{M}$ from the properties of M. Let $\mathcal{D} = (\sigma, \omega, \psi, \chi)$ be a set on the sphere S^2 and $\mathcal{D}_0$ be the extreme Kerr–Newman sphere set, both with the same angular momentum and charges J, Q_E and Q_M. Define the set for the functional $\mathcal{M}$

$$\mathcal{D}_\lambda = \lambda(\mathcal{D} - \mathcal{D}_0). \tag{A.14}$$

Let $\xi(x)$ be a real function equal to 1 for $x \leqslant 0$, equal to $-x+1$ for $x \in [0,1]$ and equal to 0 for $x \geqslant 1$. For every $\epsilon > 0$, define $\xi_\epsilon(\bar{r}) = -1/\bar{r} + 1/\epsilon$. Finally, consider the data for the functional M

$$\bar{\mathcal{D}}_{\lambda,\epsilon}(\bar{r},\theta,\varphi) = \bar{\mathcal{D}}_0(\bar{r},\theta,\varphi) + \mathcal{D}_\lambda(\theta,\phi). \tag{A.15}$$

A long but otherwise straightforward calculation shows

$$\left.\frac{d\mathcal{M}(\mathcal{D}_\lambda)}{d\lambda}\right|_{\lambda=0} = \lim_{\epsilon\to 0}\frac{1}{\epsilon}\left.\frac{dM(\bar{\mathcal{D}}_{\lambda,\epsilon})}{d\lambda}\right|_{\lambda=0}. \tag{A.16}$$

Now, since $M(\bar{\mathcal{D}}) \geqslant M(\bar{\mathcal{D}}_0)$, the right-hand side is zero for every $\epsilon > 0$; therefore, the left-hand side is zero and, because $\mathcal{D}$ was arbitrary, we conclude that $\mathcal{D}_0$ is a critical point of $\mathcal{M}$.

References

[1] Acena A, Dain S and Gabach Clement M E 2011 *Class. Quantum Grav.* **28** 105014 (arXiv:gr-qc/1012.2413)
[2] Andersson L and Metzger J 2009 *Commun. Math. Phys.* **290** 941–72 (arXiv:gr-qc/0708.4252)
[3] Andersson L, Mars M and Simon W 2005 *Phys. Rev. Lett.* **95** 111102 (arXiv:gr-qc/0506013)
[4] Andersson L, Mars M and Simon W 2008 *Adv. Theor. Math. Phys.* **12** 853–88 (arXiv:gr-qc/0704.2889)
[5] Ansorg M, Hennig J and Cederbaum C 2011 *Gen. Rel. Grav.* **43** 1205–10 (arXiv:gr-qc/1005.3128)
[6] Ansorg M and Pfister H 2008 *Class. Quantum Grav.* **25** 035009 (arXiv:gr-qc/0708.4196)
[7] Ashtekar A, Beetle C and Lewandowski J 2001 *Phys. Rev.* D **64** 044016 (arXiv:gr-qc/0103026)
[8] Ashtekar A, Engle J, Pawlowski T and Van Den Broeck C 2004 *Class. Quantum Grav.* **21** 2549 (arXiv:gr-qc/0401114)
[9] Ashtekar A, Fairhurst S and Krishnan B 2000 *Phys. Rev.* D **62** 104025 (arXiv:gr-qc/0005083)
[10] Ashtekar A and Krishnan B 2004 *Living Rev. Relat.* **7** 10 (arXiv:gr-qc/0407042)
[11] Booth I 2005 *Can. J. Phys.* **83** 1073–99 (arXiv:gr-qc/0508107)
[12] Booth I and Fairhurst S 2008 *Phys. Rev.* D **77** 084005 (arXiv:gr-qc/0708.2209)
[13] Carter B ed C DeWitt 1972 Black hole equilibrium states *Black Holes (Les Houches)* pp 125–214
Carter B 2010 *Gen. Rel. Grav.* **42** 653–744
[14] Cederbaum C 2012 Private communication
[15] Chrusciel P, Eckstein M, Nguyen L and Szybka S 2011 *Class. Quantum Grav.* **28** 245017 (arXiv:gr-qc/1111.1448)
[16] Chrusciel P and Lopes Costa J 2009 *Class. Quantum Grav.* **26** 235013 (arXiv:gr-qc/0909.5625)
[17] Costa J L 2009 arXiv:gr-qc/0912.0838
[18] Dain S 2006 *Class. Quantum Grav.* **23** 6857–72 (arXiv:gr-qc/0508061)
[19] Dain S 2008 *J. Diff. Geom.* **79** 33–67 (arXiv:gr-qc/0606105)
[20] Dain S and Reiris M 2011 *Phys. Rev. Lett.* **107** 051101 (arXiv:gr-qc/1102.5215)
[21] Dain S, Jaramillo J L and Reiris M 2012 *Class. Quantum Grav.* **29** 035013 (arXiv:gr-qc/1109.5602)
[22] Dain S 2012 *Class. Quantum Grav.* **29** 073001 (arXiv:gr-qc/1111.3615)
[23] Dain S 2012 Talk given at the *Conf. on 'Relativity and Gravitation, 100 Years after Einstein in Prague' in Prague*, (Czech Republic)
[24] Gabach Clement M E 2011 arXiv:gr-qc/1102.3834
[25] Gabach Clement M E and Jaramillo J L 2012 *Phys. Rev.* D **86** 064021 (arXiv:gr-qc/1111.6248)
[26] Gourgoulhon E 2012 *Lecture Notes in Physics* vol 846 (Berlin: Springer) arXiv:gr-qc/0703035
[27] Gourgoulhon E and Jaramillo J L 2006 *Phys. Rep.* **423** 159 (arXiv:gr-qc/0503113)
[28] Hayward S 1994 *Phys. Rev.* D **49** 6467
[29] Hawking S W and Ellis G F R 1973 *The Large Scale Structure of Space-Time* (Cambridge: Cambridge University Press)
[30] Hennig J, Ansorg M and Cederbaum C 2008 *Class. Quantum Grav.* **25** 162002 (arXiv:gr-qc/0805.4320)
[31] Hennig J, Cederbaum C and Ansorg M 2010 *Commun. Math. Phys.* **293** 449–67 (arXiv:gr-qc/0812.2811)
[32] Hildebrandt S, Kaul H and Widman K 1977 *Acta Math.* **138** 1–16
[33] Hollands S 2012 *Class. Quantum Grav.* **29** 065006 (arXiv:gr-qc/1110.5814)
[34] Jaramillo J L, Reiris M and Dain S 2011 *Phys. Rev.* D **84** 121503 (arXiv:gr-qc/1106.3743)
[35] Jaramillo J L 2013 Area inequalities for stable marginally trapped surfaces *Recent Trends in Lorentzian Geometry* ed M Sánchez, M Ortega and A Romero *Springer Proc. Math. Stat.* **26** 139–61 (arXiv:1201.2054)
[36] Jaramillo J L 2012 *Class. Quantum Grav.* **29** 177001 (arXiv:gr-qc/1206.1271)
[37] Mars M 2009 *Class. Quantum Grav.* **26** 193001 (arXiv:gr-qc/0906.5566)
[38] Mars M 2012 *Class. Quantum Grav.* **29** 145019 (arXiv:gr-qc/1205.1724)
[39] Meeks W, Simon L and Yau S T 1982 *Ann. Math.* **116** 621–59
[40] Neugebauer G and Hennig J 2012 *J. Geom. Phys.* **62** 613 (arXiv:gr-qc/1105.5830)
[41] Penrose R 1973 *Ann. New York Acad. Sci.* **224** 125
[42] Racz I 2008 *Class. Quantum Grav.* **25** 162001 (arXiv:gr-qc/0806.4373)
[43] Reiris M 2012 arXiv:1209.4530
[44] Schoen R and Yau S T 1979 *Topology* **18** 361–80
[45] Schoen R and Zhou X 2012 arXiv:gr-qc/1209.0019
[46] Simon W 2012 *Class. Quantum Grav.* **29** 062001 (arXiv:gr-qc/1109.6140)
[47] Wald R 1984 *General Relativity* (Chicago, IL: University of Chicago Press)
[48] Waxenegger G, Beig R and Murchadha N Ó 2011 *Class. Quantum Grav.* **28** 24500 (arXiv:gr-qc/1107.3083)
[49] Weinstein G 1990 *Commun. Pure Appl. Math.* **43** 903–48
[50] Weinstein G 1996 *Commun. Partial Diff. Eqns.* **21** 1389–430 (arXiv:gr-qc/9412036)