Milan J. Math. Vol. 83 (2015) 295–311
DOI 10.1007/s00032-015-0241-2
Published online July 26, 2015

Milan Journal of Mathematics

Myers and Hawking Theorems: Geometry for the Limits of the Universe

Pablo Morales Álvarez and Miguel Sánchez

Abstract. It is known that the celebrated theorem by Hawking which assures the existence of a *Big-Bang* under physically motivated hypotheses, uses geometric ideas inspired in classical Myers theorem. Our aim here is to go a step further: first, a result which can be interpreted as the exact analogy in pure Riemannian geometry to Hawking theorem will be proven and, then, the isomorphic role of the hypotheses in both theorems will be analyzed. This will provide some interesting links between Riemannian and Lorentzian geometries, as well as an introduction to the latter.

The reader interested only in Riemannian Geometry can regard this new result as a simple application of Myers theorem combined with the properties of focal points. However, readers with broader perspectives will learn that when a geometer thinks in our space as a *complete* Riemannian manifold, a relativist may think in our spacetime as *predictable*, or that suitable bounds on the Ricci tensor will force geodesics either to converge in the space or to join in the time. Moreover, the limitation of the distance from any point to a hypersurface in a Riemannian manifold, may turn out into a catastrophic relativistic limit for the duration of our physical Universe.

Mathematics Subject Classification. Primary 53C50, 53C20; Secondary 53C21, 83C75.

Keywords. Focal points, expanding hypersurface, positive Ricci curvature, singularity theorems, separating and Cauchy hypersurfaces, timelike convergence.

1. Introduction

One of the most classical theorems in Riemannian Geometry is Myers' one [16], which can be stated as follows:

Theorem 1.1 (Myers). *Let (M, g_R) be a (connected) complete Riemannian manifold. If its Ricci tensor is lower bounded by some constant $(n-1)k > 0$ (i.e., $Ric(v, v) \geq (n-1)kg_R(v, v)$ for all tangent vectors v), then its diameter is at most $\pi/\sqrt{k}$.*

Birkhäuser

Myers theorem is a refinement of a previous one by Bonnet, which yields the same conclusion under the stronger assumption that the sectional curvatures are lower bounded by k. Known consequences of Myers theorem are that only compact manifolds can admit complete metrics with a positive lower bound for the Ricci curvature, and that the fundamental group of such manifolds must be finite (see for example the nice exposition in [9]).

A natural extension of Riemannian Geometry is Lorentzian one, where the (positive definite) Riemannian metric g_R is replaced by a (index one) Lorentzian metric g_L. Such a change may be scaring for a pure Riemannian geometer, as most of the power of Riemannian geometry relies on the positive character of the metric. However, the applicability of Lorentzian Geometry to Einstein's General Relativity has led many physicists, as well as a quickly increasing number of mathematicians, to focus in this geometry.

A Lorentzian manifold (M, g_L) admits a standard Levi-Civita connection as in the Riemannian case and, then, local concepts such a geodesics, curvature tensor or covariant derivative. Nevertheless, difficulties to translate global Riemannian tools to the Lorentzian setting appear from the very beginning. In fact, there is no any distance associated to a Lorentzian metric and, thus, no any analogy to Hopf-Rinow theorem exists —notice that this theorem is the elementary starting point at any global Riemannian result. So, the name "complete" for a Lorentzian metric can be used only in the sense of "geodesically complete" but, beware, even a compact Lorentzian manifold can be (geodesically) incomplete [17, 7.16].

Nevertheless, intuitions about the structure of our physical Universe have led to the development of a nowadays well-settled *Global Lorentzian Geometry*. This is broadly inspired in the Riemannian one, but it also presents striking differences [3, 14, 17]. Among the results in this geometry, a specially representative example is Hawking's celebrated theorem about singularities on spacetimes (see the detailed expositions in [12] or [17]). This is commonly regarded as a strong support for the existence of the *Big-Bang*. Its precise statement is the following:

Theorem 1.2 (Hawking). *Let (M, g_L) be a spacetime such that:*

(i) *It is globally hyperbolic.*

(ii) *Some spacelike Cauchy hypersurface Σ is strictly expanding, i.e., $H \geq C > 0$ for some constant C, where H is the future mean curvature (that is, $\vec{H} = H\vec{n}$ with $\vec{n}$ the future unit normal).*

(iii) *Strong energy (or timelike convergence) condition holds: $Ric(v, v) \geq 0$ for any timelike tangent vector v.*

Then, the time-separation d_L satisfies $d_L(p, \Sigma) < 1/C$ for any $p \in I^-(\Sigma)$. In particular, any past-directed timelike geodesic γ is incomplete.

This may seem beyond comprehension to non-initiated people, but even geometers who have some acquaintance with Lorentzian Geometry cannot see any relation to Riemannian Geometry. In fact, only the appearance of the Ricci tensor in the third hypotheses suggests some link with Myers'; neither the other hypotheses nor the conclusion seem to be related.

Our purpose here is to fill this gap, by proving a purely Riemannian result close in spirit to Hawking's, and analyzing the dual role of the hypotheses. Concretely:

Theorem 1.3. *Let (M, g_R) be a connected Riemannian manifold satisfying the following conditions:*

(i) *g_R is complete.*

(ii) *There exists some embedded hypersurface S "separating" M in M_+ and M_- (definition* 3.1*), and with an infimum $C > 0$ of its expansion towards M_+, i.e., its mean curvature vector $\vec{H} = H\vec{n}$ satisfies $H \geq C > 0$, where $\vec{n}$ is the unit vector field normal to S which points out M_-.*

(iii) *$Ric(v, v) \geq 0$ for every tangent vector v.*

Then, the distance to S satisfies $d(p, S) \leq 1/C$ for every $p \in M_-$.

As we will see, this result can be obtained by using standard techniques in the study of Myers theorem and focal points. However, we are not aware of a similar proven result in the literature, and a proof (plus a discussion about its applicability, Remark 3.2) will be provided in Section 3. Previously, in Section 2, we recall briefly the background needed for the proof. This background will be introduced in the general semi-Riemannian setting, which includes both, Riemannian and Lorentzian geometries, since it will be also used for (Lorentzian) Hawking theorem.

In Section 4.1, the Lorentzian setting of relativistic spacetimes will be briefly introduced. No previous knowledge on Lorentzian Geometry is assumed and, so, this may serve as a first contact with this geometry. The proof of Hawking's theorem will be carried out in Section 4.2, by following exactly the same pattern as in the Riemannian version. The similarities between the proofs of Theorem 1.2 and 1.3 will be discussed further in Section 5. In fact, we will stress the isomorphic role of each one of the hypothesis (i), (ii) and (iii) in the corresponding proofs.

We remark that, in spite of the similarities between the proofs of both theorems, the conclusions admit very different interpretations. Roughly, in Riemannian Theorem 1.3 the conclusion is just that the distances from M_- to S must be bounded by $1/C$; that is, there is a limit for the "width" of M_-. However, in Lorentzian Theorem 1.2, the most striking conclusion asserts geodesic incompleteness. As we will see, this conclusion can be interpreted as the existence of some "singularity" in our past, that is, that our Universe has an initial limit with some sort of Big-Bang. After the proof of each theorem (1.2 or 1.3), a remark will be devoted to outline this geometrical or physical meaning respectively.

2. Semi-Riemannian background

Let us start bringing together the main classical results to be used. The aim is to set notation and let the reader know the roots of our theorem. As explained above, this section will be *semi-Riemannian*, since these results will be used to prove both, the Lorentzian theorem by Hawking and our Riemannian version. A thorough exposition of the semi-Riemannian setting can be found in O'Neill's book [17], our main reference along this section. In what follows, the reader interested only in the Riemannian result can simply consider that the metric g below is Riemannian and ignore some observations for the indefinite case.

Let M be a (smooth, connected, Hausdorff, necessarily paracompact) manifold endowed with a semi-Riemannian metric g, i.e., g is a non-degenerate two covariant symmetric tensor, possibly Riemannian (g is definite positive), Lorentzian (the index of g is one $(-,+,\dots,+)$) or with a higher index and, thus, with a canonical Levi-Civita connection. Consider a piecewise smooth curve $\sigma : [a,b] \to M$, and a (piecewise smooth) variation of σ, that is, $\sigma_s(t) \equiv f(s,t)$ with $f : (-\delta,\delta)\times[a,b] \to M$ and $\sigma_0 = \sigma$ (see for example [17], above its Proposition 10.2). The variation f has an associated variational field

$$V(t) = \frac{\partial f}{\partial s}(0,t)$$

and viceversa, any (piecewise smooth) vector field on σ, $V \in \mathfrak{X}(\sigma)$, is a variational field for some variation. One can define the functional "length" as:

$$\mathcal{L} : (-\delta,\delta) \to \mathbb{R} \qquad s \mapsto \text{length}(\sigma_s) := \int_a^b \sqrt{\left| g\left(\frac{\partial f}{\partial t}(s,t), \frac{\partial f}{\partial t}(s,t)\right)\right|}\mathrm{d}t$$

where the absolute value takes into account that g is not necessarily positive definite. Consider an (embedded) submanifold P of (M,g) which is nondegenerate, i.e., the restriction of g to the bundle TP is not degenerate. This implies that P can be also regarded as a semi-Riemannian manifold endowed with this restriction, the orthogonal bundle $TP^\perp$ satisfies $TP \oplus TP^\perp \hookrightarrow TM$ and the second fundamental form $\mathrm{I\!I} : \mathfrak{X}(P) \times \mathfrak{X}(P) \to \mathfrak{X}(P)^\perp$ of P is well defined. For any $q \in M$, put:

$$\Omega(P,q) = \{\text{piecewise smooth curves running from } P \text{ to } q\}.$$

In particular, given a curve $\sigma \in \Omega(P,q)$ we will consider (P,q)*-variations of* σ, that is, any variation such that all its *longitudinal curves* σ_s belong to $\Omega(P,q)$. Such a variation can be regarded as *a curve in* $\Omega(P,q)$ $(s \mapsto \sigma_s)$ passing through $\sigma = \sigma_0$. Then, its variational field V can be seen as a tangent vector to $\Omega(P,q)$ in σ. This suggests to denote, for any $\sigma \in \Omega(P,q)$:

$$T_\sigma\Omega(P,q) = \{\text{piecewise smooth fields along } \sigma \text{ with } V(a) \in T_{\sigma(a)}P \text{ and } V(b) = 0\}.$$

This set has a natural structure of vectorial space and, consistently, it is easy to check that there exists a (P,q)-variation with variational vector field V for every $V \in T_\sigma\Omega(P,q)$ (see [17, Lemma 10.49]). These concepts allow to give the following variational characterization of a geodesic normal to a submanifold (see [17, Corollary 10.26]), which extends trivially a well-known Riemannian one.

Proposition 2.1. *Let (M,g) be a semi-Riemannian manifold, P a nondegenerate submanifold, and $q \in M$. Let $\sigma : [a,b] \to M$ be a curve of $\Omega(P,q)$ with constant sign $\varepsilon \in \{-1,+1\}$ of $g(\sigma',\sigma')$ and constant speed $c = \sqrt{|g(\sigma',\sigma')|} > 0$. Then, the following statements are equivalent:*

(i) *σ is a geodesic normal to P (that is, $\sigma'(a) \perp T_{\sigma(a)}P$); in particular, σ is smooth on all $[a,b]$.*

(ii) *σ is a critical point of the length for any (P,q)-variation, that is, $\mathcal{L}'(0) = 0$ for any such variation of σ.*

Remark 2.2. Recall that, by assumption, $g(\sigma', \sigma')$ cannot vanish, that is, σ is *non-null*; by continuity, this property will hold for nearby longitudinal curves too. Moreover, we will be interested in the case that σ is a *co-spacelike curve*, that is, the orthogonal of σ is positive definite and, therefore, either g is Riemannian (case $\varepsilon = 1$) or g is Lorentzian ($\varepsilon = -1$); in the latter case, σ is called *timelike*, consistently with the definitions in Section 4.1. As a consequence, P will be *spacelike*, in the sense that g restricted to TP is positive definite.

For these curves with the property of being critical points for any (P,q)-variation, it is natural to consider the second derivative $\mathcal{L}''(0)$. That is the origin of the next concept:

Definition 2.3 (Index form). Let (M,g) be a semi-Riemannian manifold, P a nondegenerate submanifold, $q \in M$ and $\sigma \in \Omega(P,q)$ be a nonnull geodesic which is normal to P at its origin. The *index form of* σ is the unique symmetric bilinear form

$$I_\sigma : T_\sigma\Omega(P,q) \times T_\sigma\Omega(P,q) \to \mathbb{R}$$

satisfying $I_\sigma(V,V) = \mathcal{L}''(0)$, where $\mathcal{L}$ is the length functional associated to any (P,q)-variation of σ with variational field V.

Remark 2.4. Recall that, in order to make consistent this definition, one should check that the value of $\mathcal{L}''(0)$ only depends on the (P,q)-variation through its variational field V. All these standard details hold as in the Riemannian case, and they are exhaustively explained in the first three sections of [17, Chapter 10]. In this reference, an explicit expression for $I_\sigma(V,W)$ (which we will not need here) can be also found. The case of null geodesics and its index forms is studied carefully in [3, Section 10.3].

Finally, let us recall the next two standard concepts, which will be combined with the previous index form in Proposition 2.7.

Definition 2.5 (P-Jacobi field). Let (M,g) be a semi-Riemannian manifold, P a nondegenerate submanifold, and $\sigma : [a,b] \to M$ a geodesic normal to P at the origin. A *P-Jacobi field along* σ is a field $J \in \mathfrak{X}(\sigma)$ satisfying the Jacobi equation:

$$J'' + R(J,\sigma')\sigma' = 0$$

and the next two conditions at the origin:

1. $J(a) \in T_{\sigma(a)}P$.
2. $\tan(J'(a)) = \widetilde{\mathrm{II}}(J(a), \sigma'(a))$, being $\tan(\cdot)$ the tangent component to the subspace $T_{\sigma(a)}P$, and $\widetilde{\mathrm{II}}$ the operator associated to the second fundamental form II of P defined by $g(\widetilde{\mathrm{II}}(V,Z),W) = -g(\mathrm{II}(V,W),Z)$ for all $V,W \in \mathfrak{X}(P), Z \in \mathfrak{X}(P)^\perp$ (see for example [17, Remark 4.39]).

This definition is equal to the one given in [17, Proposition 10.28], and a better geometrical understanding of the two conditions on the origin can be found in this reference. However, we will use the notion of P-Jacobi field just in order to introduce the next classical concept [17, Definition 10.29].

Definition 2.6 (P-focal point). Let (M, g) be a semi-Riemannian manifold, P a nondegenerate submanifold, and $\sigma : [a, b] \to M$ a geodesic normal to P at the origin. A value $r \in (a, b]$ is called *P-focal value* (or $\sigma(r)$ is called *P-focal point*) if there exists a non identically zero P-Jacobi field with $J(r) = 0$.

The next proposition (a standard result of geometrical calculus of variations, see for example [17, Theorem 10.34]) will be essential for the proof of both, Hawking's theorem and our Riemannian version. It deals with a well-known property for geodesics in the Riemannian setting: "among nearby curves, geodesics minimize before their first focal point" (and not beyond it). This type of property can be extended to cospacelike geodesics (in the sense of Remark 2.2). Indeed, a formally analogous computation in the Lorentzian setting shows that "among nearby curves, timelike geodesics maximize the length before their first focal point".

Proposition 2.7 (Focal points theorem). *Let (M, g) be a Riemannian or Lorentzian manifold, P a nondegenerate submanifold and $\sigma : [a, b] \to M$ a cospacelike geodesic normal to P at the origin. If σ has a P-focal value $r \in (a, b)$, then I_σ is indefinite.*

Recall that r must be in the *open* interval (a, b). The geometric meaning of this result is clear: under the assumptions considered for σ, there exist (P, q)-variations of σ for which $\mathcal{L}$ has a minimum at σ and others for which it has a maximum. In particular, *there are strictly shorter and longer curves than σ connecting P and $\sigma(b)$, which can be chosen arbitrarily close to σ.*

To end up, we are going to provide a result that, under certain hypotheses, enables one to find a focal point along a geodesic. Roughly, in Hawking's theorem and our Riemannian version, this result will guarantee the existence of a focal point precisely in the open interval (a, b) of a curve, and then we will be able to apply Proposition 2.7, which will lead to a contradiction (see the detailed proofs in the corresponding sections).

Proposition 2.8 (Existence of focal points). *Let (M, g) be a Riemannian or Lorentzian manifold, P a nondegenerate submanifold and $\sigma : [0, b] \to M$ a cospacelike geodesic normal to P at the origin. Suppose:*

(i) $\mathcal{K} := g\left(\sigma'(0), \vec{H}_{\sigma(0)}\right) > 0$, *where $\vec{H}$ is the mean curvature vector.*

(ii) $Ric(\sigma', \sigma') \geq 0$, *where Ric is the Ricci tensor.*

Then, there is a P-focal value at some $r \in (0, 1/\mathcal{K}]$, provided that σ is defined in this interval, that is, if $1/\mathcal{K} \leq b$.

Remark 2.9. The detailed proof is elementary and can be found in [17, Proposition 10.37]. Remarkably, this proposition can be regarded as a corollary to Myers' theorem (or at least, to its technique): a close look at the proof of this theorem in [9, Theorem

9.3.1] reveals an analogous pattern to the cited proof of Proposition 2.8 in [17]. This influence is pointed out in the graphic summary of Figure 3.

3. Riemannian version of Hawking's theorem

With this classical background settled, we are ready to prove the announced Theorem 1.3: the Riemannian version of Hawking's. Figure 1 may help to follow the geometrical idea of the proof. We insist on the purely Riemannian character of both the result and the proof. Previously, the following definition may clarify the hypothesis (ii) of the theorem.

Definition 3.1. Let M be a connected manifold and S an embedded closed hypersurface. We will call S a *separating* hypersurface if $M - S$ has two connected parts M_- and M_+, with common boundary S. In that situation, it is possible to speak of a transverse vector to S as *pointing out* M_- *or* M_+.

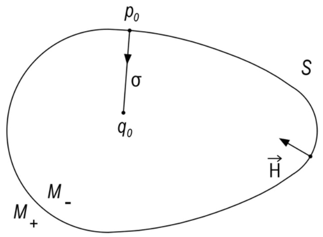

FIGURE 1. Elements in the proof of Theorem 1.3.

Proof of Theorem 1.3. Assume that there exists some $q_0 \in M_-$ such that $b_0 := d(q_0, S) > 1/C$. After some steps, this will lead to a contradiction:

1. We claim the existence of a point $p_0 \in S$ such that the distance d associated to g_R satisfies $d(q_0, p_0) = b_0$. In fact, let B be the closed ball $\bar{B}(q_0, 2b_0)$ centered at q_0 with radius $2b_0$. As B is closed and bounded in (M, g_R), it is also compact by the completeness of g_R and Hopf-Rinow theorem. Moreover, by hypothesis, S is closed and, so, $B \cap S \neq \emptyset$ is compact. This and the continuity of the distance d, yield a point $p_0 \in B \cap S$ such that $d(q_0, p_0) = d(q_0, B \cap S)$. Furthermore, the definition of B implies that $d(q_0, B \cap S) = d(q_0, S) = b_0$.
2. Since (M, g_R) is Riemannian, complete and connected, Hopf-Rinow theorem provides a minimizing geodesic $\sigma : [0, b_0] \to M$ with constant speed 1 such that $\sigma(0) = p_0$, $\sigma(b_0) = q_0$.
3. We also claim that σ is normal to S at p_0 and points out M_-, i.e., $\sigma'(0)$ is equal to $\vec{n}(p_0)$, the unit normal vector pointing out at M_-. For the first assertion, simply, use Proposition 2.1, as $L(\sigma) = d(q_0, S)$ and, thus, σ is a minimum for the length functional $\mathcal{L}$. For the second one, notice that, from the proven properties, $\sigma'(0)$ must be either $\vec{n}(p_0)$ or $-\vec{n}(p_0)$. Nevertheless, the latter option can be discarded as, otherwise, there would exist some $\delta > 0$ such that $\sigma(0, \delta) \subseteq M_+$.

However, as $\sigma(b_0) = q_0 \in M_-$, there would be some $t_0 \geq \delta$ such that $\sigma(t_0) \in S$. This would make $\sigma\big|_{[t_0,b_0]}$ be a curve connecting S and q_0, *strictly shorter than* σ, a contradiction.

4. Let $\vec{H}_{p_0} = H(p_0)\vec{n}(p_0)$ be the mean curvature vector. Let us check that if $1/H(p_0) < b_0$, then σ has a focal value at some $r \in (0, b_0)$. This is a straightforward consequence of Proposition 2.8, as $\mathrm{Ric}(\sigma', \sigma') \geq 0$ by the hypothesis (iii) of Theorem 1.3 and, using the previous item,

$$\mathcal{K} = g_R\left(\sigma'(0), \vec{H}_{p_0}\right) = g_R\left(\vec{n}(p_0), H(p_0)\vec{n}(p_0)\right) = H(p_0).$$

5. The last point is applicable, that is, σ must have a focal point in $r \in (0, b)$. In fact, $1/H(p_0) < b_0$, as hypothesis (ii) of the theorem says $H(p_0) \geq C$ and we are assuming $b_0 > 1/C$.
6. This focal point yields a contradiction, as Proposition 2.7 guarantees the existence of another curve joining p_0 and q_0 strictly shorter than σ, i.e., $d(q_0, p_0) < b_0 := d(q_0, S)$, in contradiction with the choice of p_0.

Remark 3.2. A simple application of Theorem 1.3 is to consider a hypersurface S embedded in Euclidean space $\mathbb{R}^n$, so that the hypotheses (i) and (iii) of the theorem are automatically satisfied.

If S is a topological sphere, Jordan-Brouwer theorem assures that it separates $\mathbb{R}^n$ in two regions as required, the inner one M_- and the outer one M_+. Moreover, it is well known that there exists an (elliptic) point $p_0 \in S$ such that the second fundamental form is positive definite with respect to $\vec{n}(p_0)$ (just take any point $p_- \in M_-$ and choose $p_0 \in S$ with $d(p_-, p_0) = d(p_-, S)$), and thus $H(p_0) > 0$. In the case of S being an ovaloid, we have $H(p) > 0$ for all $p \in S$, and then $H \geq C > 0$ by the compactness of S. Thus, Theorem 1.3 becomes applicable, and the optimality of the bound for $d(p_-, S)$ is attained, for instance, when p_- is the center of a round sphere. The case of a big square with rounded corners shows that the inequality $H \geq 0$ is not enough to obtain a bound in the spirit of Theorem 1.3.

The cases when S is a cylinder or a paraboloid in $\mathbb{R}^3$ show, respectively, the applicability of Theorem 1.3 for a non compact hypersurface and the necessity to strenghten $H > 0$ into a positive lower bound for H. The case of the cylinder also shows that Theorem 1.3 does not yield always an optimal bound; nevertheless, one should take into account that Theorem 1.3 is applicable when M is nonnegatively Ricci curved (sharper bounds could be found for spaces with nonnegative sectional curvature, as $\mathbb{R}^n$ itself).

4. Hawking's original theorem

Before sketching the proof of Hawking's result, we recall some specific Lorentzian notions, which complete the semi-Riemannian ones provided in Section 2.

4.1. Framework of Lorentzian manifolds and relativistic spacetimes

For the description of the next Lorentzian concepts, we follow mostly the conventions of [17] and include some recent developments surveyed in [14].

A Lorentzian metric g_L on a manifold M yields a *Lorentzian scalar product* $(g_L)_p$ with signature $(-, +, \ldots, +)$ at each tangent space T_pM, $p \in M$. A non-zero tangent vector $v_p \in T_pM \setminus \{0_p\}$ is called *timelike* (resp. *lightlike*; *causal*; *spacelike*) if $g_L(v_p, v_p) < 0$ (resp. $g_L(v_p, v_p) = 0$; v_p is timelike or lightlike i.e., $g_L(v_p, v_p) \leq 0$; $g_L(v_p, v_p) > 0$). The timelike vectors of T_pM are distributed in two connected parts, each one called a *time cone*. When one chooses one of these time cones and declares that it is the *future cone* (and, so, the non-chosen one is the *past cone*) then $(T_pM, (g_L)_p)$ is *time-oriented*. A *time-orientation* in a Lorentzian manifold (M, g_L) is a smooth choice of a future timelike cone for every T_pM, $p \in M$. Here, *smooth* means that the time-orientation at each point is provided by some timelike vector field X on M. Not all the Lorentzian metrics admit a time-orientation, and neither all the smooth manifolds admit a Lorentzian metric [17, Prop. 6.37].

Definition 4.1. A (relativistic) *spacetime* is a connected Lorentzian manifold (M, g_L) endowed with a time-orientation.

In a spacetime, a (piecewise smooth) curve γ is called *future or past directed* timelike/causal when so is its velocity everywhere (no change of cone is allowed at the possible breaks of γ). In Relativity, the reader, as well as any massive particle, is represented by means of a future-directed timelike curve.

The following binary relations $\ll, <, \leq$ between the points of a spacetime (M, g_L) become important:

$$\begin{aligned} \textit{Chronological: } p \ll q &\iff \exists \text{ a future directed timelike curve from } p \text{ to } q. \\ \textit{Strict causal: } p < q &\iff \exists \text{ a future directed causal curve from } p \text{ to } q. \\ \textit{Causal: } p \leq q &\iff p < q \text{ or } p = q. \end{aligned}$$

The *chronological* and *causal futures* of a point p are defined, resp., as:

$$I^+(p) = \{q \in M : p \ll q\} \qquad J^+(p) = \{q \in M : p \leq q\}$$

and analogously are defined the past notions $I^-(p), J^-(p)$. For a subset $A \subseteq M$, one writes, for example, $I^+(A) := \cup_{p\in A} I^+(p)$; easily, $I^{\pm}(A)$ is always an open subset of M (see the first section of [17, Chapter 14] or [14]).

These relations of causality generate a whole branch of Lorentzian Geometry, the *Causality Theory*, which is conformally invariant. Global conditions of causality are natural conditions imposed on the causality of a spacetime in order to make it both, physically more realistic and mathematically more interesting. We will only use the following two conditions (a complete study can be found in [3] or [14]).

Definition 4.2. A spacetime (M, g_L) is:

- *Causal*, if it does not contain any closed causal curve.

- *Globally hyperbolic*, if it is causal[1] and $J^+(p) \cap J^-(q)$ is compact for every $p, q \in M$ (i.e., there exist no "naked singularities").

An important characterization of global hiperbolicity is carried out in terms of *Cauchy hypersurfaces*, which appear in the statement of Hawking theorem.

Definition 4.3. Let (M, g_L) be a spacetime.

A *spacelike Cauchy hypersurface* is a smooth hypersurface Σ which is spacelike (i.e., g_L restricted to $T\Sigma$ is positive definite) and which is crossed exactly once by any inextendible timelike curve.

A *Cauchy temporal function* is an onto smooth function $\tau : M \to \mathbb{R}$ such that all its levels $\Sigma_a := \tau^{-1}(a), a \in \mathbb{R}$, are spacelike Cauchy hypersurfaces and τ satisfies that $\tau \circ \gamma$ is increasing for any future-directed causal curve γ.

The link between these three notions comes from the following result[2]:

Theorem 4.4 (Characterization of global hyperbolicity). *Let (M, g_L) be a spacetime. The following properties are equivalent:*

1. (M, g_L) *is globally hyperbolic.*
2. (M, g_L) *admits a spacelike Cauchy hypersurface.*
3. (M, g_L) *admits a Cauchy temporal function.*

Remark 4.5. Let us discuss briefly the three alternative properties in the previous theorem (a more detailed discussion with further references can be found in [15]).

The definition of global hyperbolicity in Definition 4.2 comprises two physical requirements: the impossibility for matter or energy to travel to its own past (i.e., to be causal) and the property of absence of naked singularities. The latter means that if a singularity existed (in the sense that some matter or energy "appears" or "disappears" in the spacetime) then this is not visible for any observer. The typical singularities of black holes or Big-Bang models are not naked, i.e., they do not violate global hyperbolicity. Intuitively, the existence of a naked singularity would produce an unpredictable spacetime.

Moreover, the existence of a Cauchy hypersurface Σ is linked to the *predictability* of the spacetime in the following sense: the conditions imposed on Σ by definition, allow one to ensure the existence and uniqueness of solutions to hyperbolic equations (as Einstein's one) for well-posed data on Σ.

Remarkably, this is equivalent to the existence of a (highly non-unique) Cauchy temporal function τ, which allows one to split the spacetime in "space" and "time".

[1] In classical references such as [3, 12, 17] a more restrictive and technical condition, *strong causality*, is used instead of causality. However, the possibility to optimize the definition of global hyperbolicity by using just causality was proved recently in [6] (see also [14, Section 3.11]).

[2] In Definition 4.3 and Theorem 4.4 we are following the explanation in [14, Section 3.11], instead of more classical references such as [3, 17]. The reason is that a *topological* version of Theorem 4.4 was proved by Geroch in a celebrated article [11] published in 1970. The question whether the "smooth" version stated here holds, remained open until the papers [4, 5], which show the equivalences (1) $\Longleftrightarrow$ (2) and (1) $\Longleftrightarrow$ (3), resp.

Moreover, any Cauchy hypersurface Σ can be understood as a level of such a τ and, so, Σ becomes “the full space at an instant of time”.

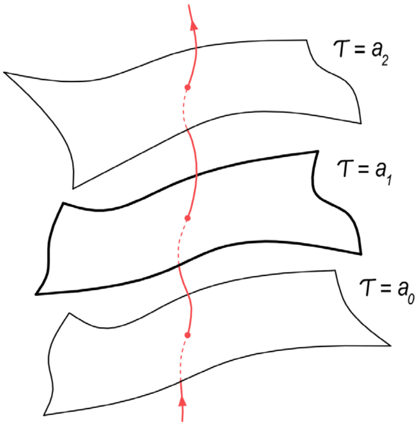

FIGURE 2. In globally hyperbolic spacetimes one has a (non-unique) choice of time such that the levels $\tau =$ constant are spacelike Cauchy hypersurfaces and, then, the spacetime can be predicted from suitable data in any of them.

The last notion to be recalled is the Lorentzian distance or time-separation of a spacetime (see [17, Chapter 14]). For this purpose, denote:

$$\mathscr{C}^c_{pq} = \{\text{piecewise smooth future directed causal curves from } p \text{ to } q\}$$

Definition 4.6 (Time separation). Let (M, g_L) be a spacetime and $p, q \in M$. The *Lorentzian distance* or *time-separation* from p to q is defined as:

$$d_L(p,q) = \left\{ \begin{array}{ll} \sup\{L(\alpha) : \alpha \in \mathscr{C}^c_{pq}\} & \text{if } \mathscr{C}^c_{pq} \neq \emptyset \\ 0 & \text{if } \mathscr{C}^c_{pq} = \emptyset \end{array} \right\} \in [0, \infty].$$

Analogously, if Σ is a subset of M, the *Lorentzian distance or time separation from* p *to* Σ, denoted $d_L(p, \Sigma)$, is the supremum of the lengths of the future-directed causal curves from p to Σ (if there are not such curves, $d_L(p, \Sigma) := 0$).

In some way, the Lorentzian distance will play a role analogous to the natural distance associated to a Riemannian metric. However, there are substantial differences, since the Lorentzian distance is not even an abstract distance —in fact, it is non-symmetric in any causal spacetime. However, it satisfies a sort of reversed triangle inequality, namely:

$$p \leq q \leq r \Rightarrow d(p,q) + d(p,r) \leq d(p,r).$$

Finally, let us recall two Lorentzian results to be used. In some way, the first one will play the role of Hopf-Rinow theorem for the Lorentzian case (see the discussion in Section 5). It is a standard result of causality theory, whose proof can be seen in [17, Propositions 14.19 and 14.21].

Theorem 4.7 (Avez-Seifert). *Let (M, g_L) be a globally hyperbolic spacetime.*

1. *If $p, q \in M$ satisfy $p < q$, there exists a future directed causal geodesic from p to q with length equal to $d_L(p, q)$ (i.e., the geodesic is maximizing).*
2. *The time-separation $d_L : M \times M \to [0, \infty]$ is finite-valued and continuous.*

The second result is just a technical lemma, but it will be crucial at a step of Hawking's theorem. As stressed in Section 5, this lemma is a major difference between the proofs of Lorentzian and Riemannian versions. A detailed proof can be found in [13, Lemma 3.71]; it is also straightforward from the results on limit curves in [3], and it is implicit in the results along [17, Chapter 14].

Lemma 4.8. *Let (M, g_L) be a globally hyperbolic spacetime and Σ a Cauchy hypersurface of M. Then, for every $p \in I^-(\Sigma)$ the set $J^+(p) \cap \Sigma$ is compact.*

4.2. Proof of Hawking's Theorem

In this subsection, we are going to sketch the proof of classical Hawking's theorem, following the same steps as in our Riemannian version, in order to emphasize the analogy and our source of inspiration.

Notice that the thesis of the theorem states a bound about the time separation, and this thesis implies easily the last assertion on timelike geodesic incompleteness. In fact, any past-directed timelike geodesic γ must cross Σ (as it is Cauchy) and enter in $I^-(\Sigma)$. However, the length of the part of γ in $I^-(\Sigma)$ must be smaller than $1/C$ because, otherwise, the inequality $d_L(p, \Sigma) < 1/C$ would be violated. So, we concentrate on the main thesis of the theorem.

Proof of $d_L(p, \Sigma) < 1/C$ in Theorem 1.2. Suppose that there exists some $q_0 \in I^-(\Sigma)$ such that $b_0 := d_L(q_0, \Sigma) \geq 1/C$. This will lead us to a contradiction:

1. We claim the existence of a point $p_0 \in \Sigma$ such that $d_L(q_0, p_0) = b_0$ $(< \infty)$. Certainly, $b_0 = d_L(q_0, \Sigma) = d_L(q_0, J^+(q_0) \cap \Sigma)$. Under our hypothesis of global hyperbolicity, Avez-Seifert theorem guarantees the continuity and finiteness of d_L, and Lemma 4.8 ensures the compactness of $J^+(q_0) \cap \Sigma$. Thus, we find the required $p_0 \in J^+(q_0) \cap \Sigma$.
2. Since (M, g_L) is globally hyperbolic, Avez-Seifert theorem provides (after a reparametrizacion) a maximizing timelike geodesic $\sigma : [0, b_0] \to M$ with constant speed 1 such that $\sigma(0) = p_0$, $\sigma(b_0) = q_0$. Recall that, by hypotheses, we know $b_0 \geq 1/C$. Even more, we can assume $b_0 > 1/C$ because, in the case of equality, we could extend the geodesic σ to $[0, b_0 + \varepsilon]$ for some $\varepsilon > 0$. Then, it would be enough to reason with $q_0' := \sigma(b_0 + \varepsilon)$ instead of q_0 (the corresponding b_0' would satisfy $b_0' > b_0 = 1/C$).
3. We also claim that σ is normal to Σ at p_0 and points out $I^-(\Sigma)$, i.e., $\sigma'(0)$ is equal to $-\vec{n}(p_0)$ (recall that we took $\vec{n}$ as future-directed). For the first assertion, use Proposition 2.1, as $L(\sigma) = d_L(q_0, \Sigma)$ (i.e., σ is a local maximum for the length functional $\mathcal{L}$). For the second one, notice that, from the proven properties, $\sigma'(0)$ must be either $\vec{n}(p_0)$ or $-\vec{n}(p_0)$. However, the first option can be discarded as, otherwise, there would exist some $\delta > 0$ such that $\sigma(0, \delta) \subseteq I^+(\Sigma)$. However, as $\sigma(b_0) = q_0 \in I^-(\Sigma)$, there would be some $t_0 \geq \delta$ such that

$\sigma(t_0) \in \Sigma$. Then, σ would be a timelike curve that crosses *twice* the Cauchy hypersurface Σ, a contradiction.

4. Let $\vec{H}_{p_0} = H(p_0)\vec{n}(p_0)$ be the mean curvature vector. Let us check that if $1/H(p_0) < b_0$, then σ has a focal value at some $r \in (0, b_0)$. This is a straightforward consequence of Proposition 2.8, as $\mathrm{Ric}(\sigma', \sigma') \geq 0$ by the timelike convergence hypothesis (iii) of Theorem 1.2 and, using the previous item,

$$\mathcal{K} = g_L\left(\sigma'(0), \vec{H}_{p_0}\right) = -g_L\left(\vec{n}(p_0), H(p_0)\vec{n}(p_0)\right) = H(p_0),$$

the latter as $\vec{n}(p_0)$ is timelike and unit.

5. The last point is applicable, that is, σ must have a focal point in $r \in (0, b_0)$. In fact, $1/H(p_0) < b_0$, as hypothesis (ii) of the theorem says $H(p_0) \geq C$, and by step (2) we have $b_0 > 1/C$.
6. This focal point yields a contradiction, as Proposition 2.7 guarantees the existence of another curve γ joining p_0 and q_0, close to σ (so that γ can be chosen timelike) and strictly longer than σ, i.e., $d_L(q_0, p_0) > b_0 := d_L(q_0, \Sigma)$, in contradiction with the choice of p_0. □

Remark 4.9. As pointed out in the introduction, Theorem 1.2 admits an interesting cosmological interpretation, in addition to its geometrical implications.

Let us begin with the interpretation of the hypotheses. The hypothesis (i) is just global hyperbolicity (already explained in Remark 4.5). This is reasonable as the predictability of our Universe becomes very appealing both, physically and philosophically. Moreover, as physicists think that the temperature of the Universe is decreasing on average, the inverse of the temperature (in absolute Kelvin scale) would seem a good tool to construct a Cauchy temporal function.

The hypothesis (ii) is justified by astronomical observations. In fact, astronomers have measured that, on average, the stars are clearly moving away from us. As they have also measured a high scale regularity, it is natural to assume that some Cauchy hypersurface Σ through our present position p_0 will be expanding with $|H(p) - H(p_0)| < \varepsilon$ for all $p \in \Sigma$ and some small $\varepsilon \in (0, H(p_0))$.

The hypothesis (iii) is called the *timelike convergence condition* because it means that, *on average, gravity attracts.* In fact, a well-known interpretation of the Jacobi equation in Riemannian Geometry yields the following interpretation: the condition $\mathrm{Ric}(v, v) \geq 0$ for all v means that, on average, nearby geodesics are attracted by curvature. In Lorentzian Geometry, to assume this inequality only for timelike (or spacelike) vectors is natural and, then, the geometric interpretation of (iii) becomes: on average, nearby timelike geodesics are attracted by curvature. However, gravity is modelled in Relativity by the curvature of the spacetime, and massive particles in free fall (i.e., only "accelerated by gravity") are modelled by (future-directed) timelike geodesics. In conclusion, (iii) turns into the attractive character of gravity. In principle, this interpretation of (iii) is highly plausible, as everybody learned when child the hypothesis that gravity attracts. Nevertheless, the remarkable discovery of the *accelerated* expansion of the Universe at the end of the XX century[3] questions the reliability of this hypothesis (and opened all type of speculations on dark energy and matter). Summing up, with some caution we can regard (iii) still as a reasonable hypotheses for a first approach.

For the interpretation of the thesis of Theorem 1.2, remind that we, as well as any massive physical particle, are represented by a future-directed timelike curve γ. Moreover, the length of its restriction to some interval $\gamma|_{a,b]}$ represents the "proper time" that we have experienced between the events $\gamma(a)$ and $\gamma(b)$. So, the conclusion of Theorem 1.2 means: *no massive particle could live a proper time bigger than* $1/C$, that is, all the particles when crossing Σ are "younger" than $1/C$. This supports the idea of the "sudden appearance" of the Universe before a time $1/C$ (where C would be close to the measured value of $H(p_0)$ by astronomers). This suggests the idea of a Big-Bang, which is also supported by other physical and philosophical arguments[4].

5. Further comparison and summary

The duality between Hawking's theorem and our Riemannian version has been stressed both, in the statement of their hypothesis (i), (ii), (iii) and in the six steps of their proofs (Sections 3 and 4.2). The next figure 3 summarizes graphically the logical interdependence between the main results involved in these proofs.

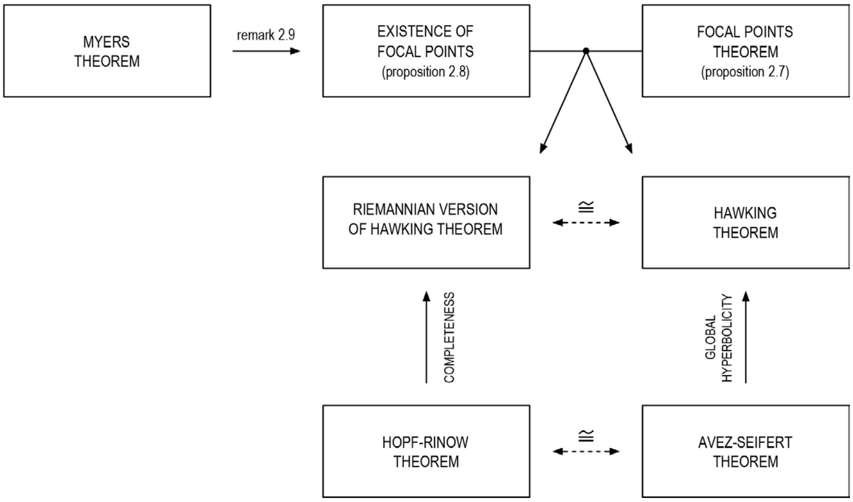
MYERS THEOREM
remark 2.9
EXISTENCE OF FOCAL POINTS (proposition 2.8)
FOCAL POINTS THEOREM (proposition 2.7)
RIEMANNIAN VERSION OF HAWKING THEOREM
≅
HAWKING THEOREM
COMPLETENESS
GLOBAL HYPERBOLICITY
HOPF-RINOW THEOREM
≅
AVEZ-SEIFERT THEOREM

FIGURE 3. Visual summary of the relations between the main results.

To end up, let us summarize the isomorphic role of all the elements in the proofs:

[3] This discovery was awarded with the Nobel Prize in Physics of 2011. See `http://www.nobelprize.org/nobel_prizes/physics/laureates/2011`

[4] Among the latter, one has Penrose's *cosmic censorship hypothesis*, see for example [7, 15].

1. In both theorems a global condition is imposed on the manifold: completeness in the Riemannian case and global hyperbolicity in the Lorentzian. These concepts seem quite different at a first look, but the theorems by Hopf-Rinow and Avez-Seifert allow one to find an analogy in the existence of minimizing and maximizing geodesics (apparent in the second step of the provided proofs).

$$\begin{tikzcd}
\text{Completeness} \arrow[r, phantom, "\leftrightarrow"] \arrow[d, Rightarrow] & \text{Global hiperbolicity} \arrow[d, Rightarrow] \\
\text{(Hopf-Rinow)} \arrow[d, Rightarrow] & \text{(Avez-Seifert)} \arrow[d, Rightarrow] \\
\exists\ \text{minimizing geodesic} \arrow[r, phantom, "\leftrightarrow"] & \exists\ \text{maximizing geodesic}
\end{tikzcd}$$

2. In both theorems, we assume positiveness for the Ricci tensor:

$$Ric \geq 0 \ \ \forall v \in TM \quad \leftrightarrow \quad Ric \geq 0 \ \ \forall v \text{ timelike}$$

   This was a subtle difference in their geometric/physical motivation pointed out in Remark 4.9. Nevertheless, both hypotheses are used at the same moment: in step (4) of the proofs, in order to apply Proposition 2.8. Notice how the notion of cospacelike curve allows us to use the mentioned proposition in both cases.

3. In both theorems we admit the existence of an embedded hypersurface, S or Σ, satisfying analogous hypothesis on its mean curvature:

   (Complete) separating hypersurface ↔ Cauchy hypersurface

   The notion of Cauchy hypersurface requires a Lorentzian framework not available in the Riemannian case. However, it is not difficult to prove that any Cauchy hypersurface is *separating* according to definition 3.1. From the mathematical viewpoint, the separating property is used in the third step of the proof, forcing σ to cross either S or Σ twice.

4. Along the proofs, the Riemannian distance d and the time-separation d_L are playing isomorphic roles. This is quite surprising, as there are remarkable differences: d_L is defined as a supremum instead of an infimum, and it is not a true distance in spite of its alternative name of Lorentzian distance.

   Riemannian distance ↔ Time separation

These analogies may shed some light on the links between Riemannian and Lorentzian Geometry. However, it is also interesting to stress some differences, emphasizing that we are dealing with very different geometries:

1. The conclusion seems to be similar expressed in terms of the Riemannian/Lorentzian distance. However, as explained in the introduction and along the paper, the geometrical/physical meaning is quite different.
2. In the first step of the proofs, we obtained a point p_0 which materializes the distance from q_0 to the hypersurface. However, the techniques are very different. In the Riemannian case the compactness of closed and bounded subsets (Hopf-Rinow theorem) is used. In the Lorentzian one, we needed the auxiliary Lemma 4.8 in addition to Avez-Seifert theorem. Certainly, these results rely heavily on purely Lorentzian notions.

3. It is also interesting to realize the subtle difference in the third step of the proofs of Theorems 1.3 and 1.2. In both cases we obtained a contradiction, but different facts were contradicted: in the Riemannian case the contradiction is that σ is minimizing, while in the Lorentzian case the contradiction is that Σ is a Cauchy hypersurface.

Finally, we would like to emphasize that singularity theorems not only constituted a classical topic of research in the General Relativity of XX century (see Senovilla [18] for a summary), but also an active field for the future, which includes links with Riemannian and Finslerian Geometries. We refer to Senovilla and Garflinke [19] for the impact of classical Penrose's theorem (the first modern singularity theorem) in current research, and Galloway and Senovilla [10] for a recent result which unifies classical Hawking's, Penrose's and other results. Recently, Aazami and Javaloyes [1] have obtained a result in the spirit of Penrose's for Finsler spacetimes, and Bailleul [2] has introduced a probabilistic viewpoint. The reader is also referred to [8] for an updated review about mathematical relativity and to [7] for a clean introduction to black holes and Penrose inequality (one of the outstanding fields of research linked to singularities), written for an audience of mathematicians.

Acknowledgments

The first-named author is supported by the grant "Beca de Iniciación a la Investigación, plan propio de 2013" (Vicerrectorado de Política Científica e Investigación, Univ. Granada), and both authors are supported by the research project MTM2013-47828-C2-1-P (Spanish MINECO and FEDER funds). The authors also acknowledge the generous help of B. Arch. Ana Rodríguez with the figures.

References

[1] Aazami A., Javaloyes M.A. (2014): *Penrose's singularity theorem in a Finsler spacetime.* Preprint, arxiv: 1410.7595.

[2] Bailleul I. (2011): *A probabilistic view on singularities.* J. Math. Phys., 52, 023520.

[3] Beem J.K., Ehrlich P.E., Easley K.L. (1996): Global Lorentzian geometry. Vol. 202 of Monographs and Textbooks in Pure and Applied Mathematics, Marcel Dekker Inc., New York, second edition.

[4] Bernal A.N., Sánchez M. (2003): *On smooth Cauchy hypersurfaces and Geroch's splitting theorem.* Comm. Math. Phys., 243, pp. 461–470.

[5] Bernal A.N., Sánchez M. (2005): *Smoothness of time functions and the metric splitting of globally hyperbolic spacetimes.* Comm. Math. Phys., 257, pp. 43–50.

[6] Bernal A.N., Sánchez M. (2007): *Globally hyperbolic spacetimes can be defined as 'causal' instead of 'strongly causal'.* Classical Quantum Gravity, 24, pp. 745–749.

[7] Bray, H. (2003): *Black holes, geometric flows, and the Penrose inequality in general relativity.* Not. Am. Math. Soc. 49, 1372–1381.

[8] Chrusciel, P., Galloway, G.J., Pollack, D. (2010): *Mathematical general relativity: a sampler*. Bull. Am. Math. Soc. (N.S.) 47(4), 567–638.

[9] Do Carmo M.P. (1992): Riemannian geometry. Mathematics: Theory and Applications, Birkhäuser Boston Inc., Boston, MA.

[10] Galloway G. J., Senovilla J. M. M. (2010): *Singularity theorems based on trapped submanifolds of arbitrary co-dimension*, Classical Quantum Gravity, 27, no. 15, 152002.

[11] Geroch R. (1970): *Domain of dependence*. J. Mathematical Phys., 11, pp. 437–449.

[12] Hawking S.W., Ellis G.F.R. (1973): The large scale structure of space-time. Cambridge Monographs on Mathematical Physics, No. **1**, Cambridge University Press, London - New York.

[13] Javaloyes M.A., Sánchez M. (2010): An introduction to Lorentzian geometry and its applications. Editorial Universidad de Sao Paulo, Sao Paulo, Brasil. ISBN: 978-85-7656-1.

[14] Minguzzi E., Sánchez M. (2008): *The causal hierarchy of spacetimes*. Recent developments in pseudo-Riemannian geometry, ESI Lect. Math. Phys., Eur. Math. Soc., Zürich, pp. 299–358.

[15] Müller O., Sánchez M. (2014): *An invitation to Lorentzian Geometry*. Jahresbericht der Deutschen Mathematiker-Vereinigung 115 No 3–4: 153–183.

[16] Myers S.B. (1941): *Riemannian manifolds with positive mean curvature*. Duke Mathematical Journal 8 (2): 401–404.

[17] O'Neill B. (1983): Semi-Riemannian geometry with applications to Relativity. Academic Press Inc., New York.

[18] Senovilla, J.M.M. (1997): *Singularity Theorems and Their Consequences*. Gen. Relat. Grav. 29, No. 5 701-848.

[19] Senovilla J.M.M., Garflinke D. (2015): *The 1965 Penrose singularity theorem*. Class. Quantum Grav. 32, 124008, 45pp.

Pablo Morales Álvarez and Miguel Sánchez
Departamento de Geometría y Topología
Universidad de Granada
Facultad de Ciencias, Campus Fuentenueva s/n
E-18071 Granada
Spain
e-mail: pablomorales@correo.ugr.es
sanchezm@ugr.es

Received: March 17, 2015.