

A new derivation of singularity theorems with weakened energy hypotheses

Christopher J Fewster[✉] and Eleni-Alexandra Kontou[✉]

Department of Mathematics, University of York, Heslington, York YO10 5DD, United Kingdom

E-mail: chris.fewster@york.ac.uk and eleni.kontou@york.ac.uk

Received 13 August 2019, revised 17 December 2019

Accepted for publication 7 January 2020

Published 18 February 2020



CrossMark

Abstract

The original singularity theorems of Penrose and Hawking were proved for matter obeying the null energy condition or strong energy condition, respectively. Various authors have proved versions of these results under weakened hypotheses, by considering the Riccati inequality obtained from Raychaudhuri's equation. Here, we give a different derivation that avoids the Raychaudhuri equation but instead makes use of index form methods. We show how our results improve over existing methods and how they can be applied to hypotheses inspired by quantum energy inequalities. In this last case, we make quantitative estimates of the initial conditions required for our singularity theorems to apply.

Keywords: general relativity, singularity theorems, energy conditions, quantum inequalities

1. Introduction

A central question in gravitational physics is to determine conditions under which singularities arise either as the endpoint of gravitational collapse or at the origin of an expanding universe. Initial efforts to answer this question were restricted to spacetimes with high symmetry or simple matter models. A major breakthrough occurred with the proof of general singularity theorems by Penrose [39] and Hawking [28] in the mid-1960's. There, for the first time, it was proved that a singularity is inevitable if the spacetime and matter obey a series of general assumptions. It is striking that these revolutionary theorems were proven about 5 years before the identification of the first black hole candidate [4, 47] and within months of the cosmic microwave background (CMB) discovery [40] (the link with singularity theorems was



Original content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

discussed in [29]). Today, though observation now provides direct images of a black hole [11] and detailed CMB measurements [2], the question of whether singularities exist in our physical universe remains open. Of course it is impossible to observationally detect singularities and a full theoretical answer requires a consistent theory of quantum gravity. In the absence of such a theory, semiclassical gravity provides the most accurate current model of the universe. It is therefore important to seek improved versions of singularity theorems that take into account the properties of quantised matter in order to shed light in the necessary conditions that lead to singular spacetimes in a semiclassical approximation.

In the context of singularity theorems a spacetime is defined as singular if it contains at least one causal geodesic that is inextendible and incomplete to the future (or past). That is, it cannot be extended arbitrarily far to the future as an affinely parameterised geodesic. This situation need not be associated with a curvature singularity. The basic ingredients of the singularity theorems may be grouped under three headings [42]: causality assumptions on the spacetime, energy conditions on its matter content, and an initial condition at some surface S . The causality assumptions are used to show that a future-geodesically complete spacetime must contain a future-complete geodesic that emerges normally from S and has no focal points to S . The precise definition of a focal point will be recalled below in section 2; for the moment it is enough to think of it as a point where nearby geodesics emanating normally from S focus (to a good approximation). On the other hand, the energy conditions and the initial condition are used to show that every future-complete geodesic emerging normally from S must contain at least one focal point. It follows that no spacetime obeying the causality, energy condition and initial conditions can be future-geodesically complete¹.

This paper is concerned with the link between energy conditions and focal points. The standard approach, as presented in [28, 30, 35, 39, 42, 45, 48] for example, links the existence of focal points to the behaviour of certain geodesic congruences leaving S ; a focal point exists if the expansion of the congruence diverges at finite affine parameter. The energy conditions are then used in conjunction with Raychaudhuri's equation to prove that such a divergence occurs, provided that the congruence is initially converging. Here, the energy conditions employed are typically the null energy condition (NEC), that $T_{\mu\nu}U^\mu U^\nu \geq 0$ for all null U^μ at all points in spacetime, or the strong energy condition (SEC), that $T_{\mu\nu}U^\mu U^\nu - T/(n-2) \geq 0$ for all timelike unit vectors U^μ at all points of spacetime, where n is the spacetime dimension. In this paper we will always incorporate the cosmological constant into the stress–energy tensor. Then Einstein's equations directly connect the strong and NECs on the stress–energy tensor with the geometric conditions of timelike and null convergence.

There are good reasons to seek generalisations of these results with weaker conditions on the matter. At the microscale, for example, it is known that matter described by quantum fields cannot obey any pointwise energy conditions [10]. Meanwhile, on the macroscale, the SEC fails in the current era of our universe, due to the dominant effect of dark energy. Several authors have considered generalisations of the singularity theorems in this direction, starting from work of Tipler [43, 44] (see also [7]) in which various averaged energy conditions are employed in place of the pointwise versions. Examples include [5, 41, 46] and [6, 15]. Most of these references involve the analysis of a Riccati inequality

$$\frac{D\theta}{dt} \leq R_{\mu\nu}U^\mu U^\nu - \frac{\theta^2}{n-r} \quad (1)$$

¹ There are also singularity theorems, such as the Hawking–Penrose theorem [31] that turn on the existence of a pair of conjugate points along a geodesic, but which we will not consider here.

derived from the Raychaudhuri equation, using results such as those of [26]. Here θ is the expansion of the geodesic congruence with velocity field U^μ and $r = 1$ (resp., $r = 2$) for timelike (resp., null) geodesic congruences; our geometric conventions are stated at the end of this section. Similar techniques may be used to establish generalised versions of other results in mathematical relativity, e.g. the area theorem [34].

In this paper, we will point out a more direct method for obtaining such results, which avoids the use of the Raychaudhuri equation and Riccati inequalities. Instead it is based on the study of the *index form*, which arises as the second variational derivative of the length functional about a geodesic. Our treatment has been influenced by O’Neill’s discussion of the standard singularity theorems [37]. Actually, index form methods were used by Chicone and Ehrlich [7] to prove the existence of conjugate points along complete geodesics, using the averaged null energy condition (ANEC) or the comparable condition for the SEC, but this seems to be the only occasion on which index forms have been used to establish singularity theorems under weakened conditions. Our treatment will make a much wider use of these techniques, and will prove results for local energy averages that are analogous to, but improve upon and are simpler to prove than, the results of [6, 15].

There are two basic advantages to this approach. The first is that it works well with weakened integral energy conditions, more easily satisfied by classical and quantum fields. The second is that it gives an estimate of the proper time or affine parameter where the focal point is formed. That means we can estimate the location of the singularity, as well as the minimum contraction required to guarantee its existence.

The averaged energy conditions we will study include some that are inspired by known quantum energy inequalities (QEIs) satisfied by quantum fields (which, as mentioned, cannot satisfy pointwise energy conditions in general). QEIs were first introduced by Ford [24] and have since been established in a number of quantum field theory models in flat and curved spacetimes (see [13] for a recent review). Often such inequalities take the general form

$$\int_{\gamma} \langle \rho \rangle_{\omega} f(\tau)^2 d\tau \geq -|||f|||^2, \quad (2)$$

for f a smooth compactly supported real-valued function. Here ρ is the renormalized energy density or similar quantity along a timelike curve γ , ω is a Hadamard state, and $|||f|||^2$ is a finite positive linear combination of squared L^2 -norms of f and its derivatives; thus $|||\cdot|||$ is dominated by a Sobolev norm.

It has been long known [14] that the quantized minimally coupled scalar field admits a bound of this form with $|||f|||^2 = (16\pi^2)^{-1} \|f''\|^2$ for averaging along a timelike geodesic in four-dimensional Minkowski space. QEIs established for this field in curved spacetimes, e.g. [12, 21] can be rewritten in this form. For example, [33] computed the QEI of [21] by perturbative methods, writing the bound in terms of L^2 -norms of derivatives of the averaging function and constants that depend on the upper bound of the curvature. The non-minimally coupled scalar field obeys (state-dependent) QEI bounds [16, 19] that could potentially also be brought into the same form. Similar remarks apply to other fields for which QEIs have been found—see [13] for references.

This paper is organized as follows. In section 2 we present an overview of index form methods and derive the basic results leading to the proofs of the original Hawking and Penrose singularity theorems. In section 3 we show how the same method can be used to prove singularity theorems with weakened energy conditions. As an example we discuss conditions involving exponentially damped integrals of the stress tensor components for the timelike and null cases. In section 4 we present the main result of the paper which is the derivation of singularity

theorems with energy conditions inspired by QEIs, see equation (2). Approximations for the required initial conditions for geodesic incompleteness in different cases are derived. In section 5 we apply the results of section 4 in the case of the classical non-minimally coupled Einstein–Klein–Gordon theory to show how our results can provide quantitative estimates. Finally we conclude in section 6 with a summary and discussion of future work. For simplicity and as a proof of principle, we have concerned ourselves solely with singularity theorems for globally hyperbolic spacetimes. This also fits well with the fact that QEIs are typically derived under the assumption of global hyperbolicity, and the stability of this condition under small metric perturbations gives some confidence in the relevance of our results for semi-classical gravity. It would also be interesting to apply our methods to singularity theorems that do not assume global hyperbolicity (see, e.g. the discussion and references in [35, section 6.6]).

Conventions. Unless otherwise stated, we consider a general spacetime dimension $n > 2$ and adopt units in which $G = c = 1$. We employ $[-, -, -]$ conventions in the Misner, Thorne and Wheeler classification [36]. That is, the metric signature is $(+, -, -, \dots)$, the Riemann tensor is defined as $R_{\lambda\eta\nu}{}^{\mu} = (\nabla_{\lambda}\nabla_{\eta} - \nabla_{\eta}\nabla_{\lambda})v^{\mu}$, and the Einstein equation is $G_{\mu\nu} = -8\pi T_{\mu\nu}$. The d'Alembertian is written $\square_g = g^{\mu\nu}\nabla_{\mu}\nabla_{\nu}$. A consequence of our conventions is that time-like and null convergence conditions take the form $R_{\mu\nu}U^{\mu}U^{\nu} \leq 0$ for timelike or null vectors U^{μ} respectively.

2. Index form methods and focal points

Here we present the basic theorems concerning the existence of focal points along timelike and null geodesics, in terms of index forms. At the end of each subsection we discuss how these theorems can be used to prove the Hawking and Penrose singularity theorems. Much of this section is closely related to chapters 10 and 14 of [37]. On a point of notation, differentiation with respect to a time parameter will be denoted with a dot, while differentiation with respect to an affine parameter on null geodesic is indicated with a prime.

2.1. Focal points along timelike geodesics

Let $(M, g_{\mu\nu})$ be a smooth Lorentzian spacetime, and let S be a smooth spacelike hypersurface in M . If $\gamma : [0, \tau] \rightarrow M$ is a smooth timelike curve, its length (i.e. the total proper time along γ) is

$$L[\gamma] = \int_0^{\tau} |\dot{\gamma}(t)| dt, \quad |V| := \sqrt{g_{\mu\nu}V^{\mu}V^{\nu}}. \quad (3)$$

Let $q \in I^+(S)$ be fixed; then γ is a critical point of the length functional among unit-speed timelike curves joining S to q if and only if it is an affine geodesic issuing normally from S . In more detail, γ is a geodesic of this type if and only if $dL[\gamma_s]/ds|_{s=0} = 0$ for every smooth 1-parameter family of curves $\gamma_s : [0, \tau] \rightarrow M$, $s \in (-\delta, \delta)$ obeying

$$\gamma_0 = \gamma, \quad \gamma_s(0) \in S, \quad \gamma_s(\tau) = q \quad (4)$$

for all $s \in (-\delta, \delta)$. Introducing some terminology, the partial derivatives of $\gamma_s(t)$ (in an arbitrary system of coordinates) with respect to t and s determine the *longitudinal* and *transverse* vector fields $U^{\mu} = \partial\gamma_s(t)^{\mu}/\partial t$, $V^{\mu} = \partial\gamma_s(t)^{\mu}/\partial s$, obeying

$$\nabla_U V^{\mu} = \nabla_V U^{\mu}, \quad (5)$$

an identity which holds for any smooth 1-parameter family of curves. In particular, the restrictions of U^μ and V^μ to γ yield the velocity vector $U^\mu|_{\gamma(t)} = \dot{\gamma}^\mu(t)$ and *variation vector field* $V^\mu|_{\gamma(t)} = d\gamma_s(t)^\mu/ds|_{s=0}$.

The above discussion may be generalised to piecewise smooth curves, with the result that γ is a critical point of L among unit-speed piecewise smooth curves if and only if it is an unbroken timelike geodesic in its proper time parametrisation. The variation field arising from a piecewise smooth variation of γ is continuous and piecewise smooth, while the velocity field may have discontinuities (see [3, section 10.1] for further background on piecewise smooth variations).

Now suppose that γ is an affine geodesic emanating normally from S . A point p on γ is a *focal point to S along γ* if there is a nontrivial variation of γ , among affine geodesics issuing normally from S , with a variation field that vanishes at p . The corresponding variation vector field therefore obeys the equation of geodesic deviation

$$\frac{D^2 V^\mu}{dt^2} + R^\mu{}_{\nu\alpha\beta} U^\nu U^\alpha V^\beta = 0, \quad (6)$$

i.e. V^μ is a Jacobi field. The other conditions on the variation imply that

$$V^\mu|_p = 0, \quad V^\mu U_\mu|_{\gamma(0)} = 0, \quad K_{\mu\nu} V^\mu V^\nu|_{\gamma(0)} = V_\mu \nabla_U V^\mu|_{\gamma(0)}, \quad (7)$$

where $K_{\mu\nu} = \nabla_\mu \xi_\nu$ is the extrinsic curvature tensor of S and ξ^μ is the normalised longitudinal vector field of this variation (in particular $U^\mu|_\gamma = \xi^\mu|_\gamma$). In fact, the first two conditions in equation (7), combined with equation (6), imply that $U_\mu V^\mu$ vanishes identically along γ .

The existence of focal points along γ is closely related to the question of whether γ is a local maximum of the length functional, among constant speed curves joining S to q (not necessarily geodesics). The analysis starts from a formula for the second derivative of the length functional for a variation γ_s of γ in which each γ_s is a piecewise smooth constant speed curve² joining S to q and γ_0 has unit speed:

$$\left. \frac{d^2}{ds^2} L[\gamma_s] \right|_{s=0} = I[V] := \int_0^\tau \left(\frac{DV^\mu}{dt} \frac{DV_\mu}{dt} - R_{\mu\nu\alpha\beta} U^\mu V^\nu V^\alpha U^\beta \right) dt + K_{\mu\nu} V^\mu V^\nu|_{\gamma(0)}. \quad (8)$$

The quantity $I[V]$ is called the *index form*; it is usually presented in a polarised form as a bilinear form in two vector fields, but we will not need to do that here.

Owing to the conditions placed on γ_s , we have

$$V^\mu U_\mu|_{\gamma(0)} = 0, \quad V^\mu|_{\gamma(\tau)} = 0, \quad U_\mu \frac{DV^\mu}{dt} \equiv 0, \quad (9)$$

where the first two conditions reflect the boundary conditions that $\gamma_s(0) \in S$, $\gamma_s(\tau) = q$, while the third is due to the restriction to constant speed curves. (To see this, note that

$$U_\mu \nabla_U V^\mu = U_\mu \nabla_V U^\mu = \frac{1}{2} \nabla_V (U^\mu U_\mu) = \frac{1}{2\tau} \frac{d}{ds} L[\gamma_s]^2 \Big|_{s=0} = 0 \quad (10)$$

using constant speed parametrisation and the fact that γ_0 is geodesic.) As γ is an affine geodesic, these conditions imply that $U_\mu V^\mu \equiv 0$. If V is smooth, integrating by parts in (8) and rearranging gives

$$I[V] = - \int_0^\tau V_\mu \left(\frac{D^2 V^\mu}{dt^2} + R^\mu{}_{\nu\alpha\beta} U^\nu U^\alpha V^\beta \right) dt + V_\mu \nabla_U V^\mu|_{\gamma(\tau)} + (K_{\mu\nu} V^\mu V^\nu - V_\mu \nabla_U V^\mu)|_{\gamma(0)}, \quad (11)$$

² The speed $|\dot{\gamma}_s(t)|$ may vary with s , but not t .

Table 1. Summary of theorem 10.34 in [37]. Here ‘for some/all V^μ ’ is to be interpreted as ‘for some/all piecewise smooth V^μ obeying the conditions (9)’.

\nexists focal point in $(0, \tau]$	\implies	$I[V] < 0$ for all V^μ
\exists focal point in $(0, \tau)$	\implies	$I[V] > 0$ for some V^μ
Only focal point in $(0, \tau]$ is τ	\implies	$I[V] \leq 0$ for all V^μ , and $I[V] = 0$ for some V^μ

and it is clear from equations (6) and (7) that $I[V] = 0$ if $\gamma(\tau)$ is a focal point to S along γ and V^μ is the corresponding Jacobi field (which necessarily satisfies the conditions (9)). In this situation the index form is not negative definite, so either the second derivative test is inconclusive or γ is not a local maximum of L . Table 1 summarises a more detailed relationship between the index form and focal points presented in theorem 10.34 in [37]. From these logical relationships, it may be seen that the first two implications may be replaced by ‘if and only if’ statements. For if $I[V] < 0$ for all V^μ then neither of the second or third conclusions holds, and therefore neither of the hypotheses on these lines can hold. Therefore there is no focal point in $(0, \tau]$ and so the first implication admits a converse. Similarly, if the second conclusion holds then neither of the first or third can and we deduce that there must be a focal point in $(0, \tau)$, so the implication on the second line also admits its converse. In particular, we have:

Proposition 2.1. *There exists a focal point $\gamma(r)$ to S along γ for $r \in (0, \tau]$ (resp., $r \in (0, \tau)$) if and only if $I[V] \geq 0$ (resp., $I[V] > 0$) for some piecewise smooth V^μ obeying (9). Consequently: if γ is length-maximising, then there is no focal point in $(0, \tau)$; if γ is not length-maximising, then there is a focal point in $(0, \tau]$; if there is a focal point in $(0, \tau)$, then γ is not length-maximising; if there is no focal point in $(0, \tau]$ then γ is length-maximising.*

Next, suppose v^μ is a unit spacelike vector tangent to S at $\gamma(0)$, extended along γ by parallel transport. Then, any continuous piecewise smooth function f obeying $f(0) = 1$, $f(\tau) = 0$ determines a continuous piecewise smooth variation field obeying (9) by

$$V^\mu = f v^\mu. \quad (12)$$

Noting that $DV^\mu/dt = \dot{f}v^\mu$, which is spacelike, we have

$$I[V] = \int_0^\tau \left(-\dot{f}^2 - f^2 R_{\mu\nu\alpha\beta} U^\mu v^\nu v^\alpha U^\beta \right) dt + K_{\mu\nu} v^\mu v^\nu|_{\gamma(0)}. \quad (13)$$

But now consider an orthonormal basis e_μ ($\mu = 0, \dots, n-1$) for the tangent space to S at $\gamma(0)$, in which $e_0^\mu = U^\mu$, and apply the preceding argument to each element e_i ($i = 1, \dots, n-1$), keeping the same scalar function f . Summing, we find

$$\sum_{i=1}^{n-1} I[f e_i] = - \int_0^\tau \left((n-1) \dot{f}^2 + f^2 R_{\mu\nu} U^\mu U^\nu \right) dt - K|_{\gamma(0)}, \quad (14)$$

where we have used the fact that $U^\mu U^\nu K_{\mu\nu} = 0$, so that the extrinsic curvature $K = g^{\mu\nu} K_{\mu\nu} = -\sum_i K(e_i, e_i)$. If there is no focal point in $(0, \tau]$ (resp., $(0, \tau)$) then each term on the left-hand side is negative (resp., nonpositive) for all f obeying the boundary conditions $f(0) = 1$, $f(\tau) = 0$. Conversely, if the right-hand side is nonnegative (resp., positive) for some such f , then the same must be true of at least one of the terms on the left, and it follows that there is a focal point in $(0, \tau]$ (resp., $(0, \tau)$). In other words, we have sufficient conditions for the presence of a focal point in $(0, \tau)$ or $(0, \tau]$ (and consequently some necessary conditions for their absence).

Proposition 2.2. *Let $\gamma : [0, \tau] \rightarrow M$ be a unit-speed timelike geodesic emanating normally from a smooth spacelike hypersurface S . If there exists a continuous, piecewise smooth f on $[0, \tau]$ obeying $f(0) = 1$, $f(\tau) = 0$ and*

$$\int_0^\tau \left((n-1)\dot{f}^2 + f^2 R_{\mu\nu} U^\mu U^\nu \right) dt \leq -K|_{\gamma(0)}, \quad (15)$$

then there is a focal point to S along γ . If the inequality (15) holds with a strict inequality then the focal point lies before $\gamma(\tau)$.

This result may be used to deduce the existence of focal points, without using the Raychaudhuri equation [37, proposition 10.37]. Recall that the timelike convergence condition asserts that $R_{\mu\nu} U^\mu U^\nu \leq 0$ for all timelike U^μ , which is equivalent to the SEC for solutions to the Einstein equations³.

Corollary 2.3. *If $K|_{\gamma(0)} < 0$, $\tau \geq (n-1)/|K|_{\gamma(0)}$, and the Ricci tensor obeys the timelike convergence criterion, then there is a focal point to S along γ , i.e. by proper time τ at the latest. If $\tau > (n-1)/|K|_{\gamma(0)}$ the focal point occurs before time τ .*

Proof. Apply proposition 2.2 using the function $f(t) = 1 - t/\tau$, noting that the left-hand side of (15) is less than or equal to $(n-1)/\tau$ which is less than or equal to $-K|_{\gamma(0)}$ by assumption. This gives the first stated result; the second is a trivial modification. \square

At this point it is helpful to explain the relationship between the index form method and the traditional approaches based on the Raychaudhuri equation. One way to determine whether the condition given in proposition 2.2 holds is to solve the variational problem of minimising the right-hand side of equation (15), treated as a functional $J[f]$, over smooth f obeying $f(0) = 1$, $f(\tau) = 0$. Considering variations $f + \epsilon g$ where g is smooth and obeys $g(0) = g(\tau) = 0$, one finds easily that

$$J[f + \epsilon g] = J[f] + 2\epsilon \int_0^\tau \left(-(n-1)\dot{f} + f\rho \right) g dt + \epsilon^2 J[g], \quad (16)$$

where we now write $\rho(t) = R_{\mu\nu} U^\mu U^\nu|_{\gamma(t)}$. There is at most one stationary point, namely the solution to

$$-(n-1)\ddot{f} + f\rho = 0, \quad f(0) = 1, f(\tau) = 0, \quad (17)$$

if it exists (which it does unless there is a solution to the same equation with $f(0) = f(\tau) = 0$). For this solution, one finds $J[f] = (n-1)\dot{f}(0)$ using an integration by parts⁴. Therefore a sufficient condition for (15) to hold is that the solution to (17) has $\dot{f}(0) \leq -K|_{\gamma(0)}/(n-1)$.

Assuming that f is nonvanishing in $(0, \tau)$, we may now set $\theta = (n-1)\dot{f}/f$ and note that the Euler–Lagrange equation (17) may be rewritten as the Riccati equation

$$\dot{\theta} = -\frac{\theta^2}{n-1} + \rho, \quad \theta(0) = (n-1)\dot{f}(0), \quad (18)$$

with $\theta \rightarrow -\infty$ as $t \rightarrow \tau^-$. (If f has an interior zero then we repeat the reasoning on the interval between $t = 0$ and the first zero.) We see that a sufficient condition for the existence of a focal

³ Recall that the cosmological constant has been incorporated into the stress–energy tensor.

⁴ This stationary point is the global minimum of J provided that $J[g] \geq 0$ for all smooth g obeying Dirichlet boundary conditions, which can be analysed as an eigenvalue problem. However we will not need to do this.

point along γ is that the above Riccati equation fails to have a solution beyond $[0, \tau]$ for initial data with $\theta(0) \leq -K|_{\gamma(0)}$.

By contrast, now consider the (irrotational) congruence of unit speed timelike geodesics emanating normally from the S with velocity field U^μ . The Raychaudhuri equation for the expansion $\theta = \nabla_\mu U^\mu$ gives

$$\frac{D\theta}{dt} = R_{\mu\nu}U^\mu U^\nu - 2\sigma^2 - \frac{\theta^2}{n-1}, \quad \theta|_{\gamma(0)} = K|_{\gamma(0)}, \quad (19)$$

along any geodesic in the congruence, where σ is the shear scalar, and so θ obeys the differential inequality

$$\frac{D\theta}{dt} \leq R_{\mu\nu}U^\mu U^\nu - \frac{\theta^2}{n-1}, \quad \theta|_{\gamma(0)} = K|_{\gamma(0)}. \quad (20)$$

Therefore the Raychaudhuri equation leads to a similar analysis to that arising from the index form methods, though with the slight complication of working with a differential inequality rather than a differential equation. More significantly, in the index form approach one need not pass to differential equations at all, but instead try to satisfy (15) by judicious choice of trial functions f . This is exactly what was done in the proof of corollary 2.3 and will be our approach in the rest of this paper.

To complete this section, we now recall one of the simplest links between the presence of focal points and timelike geodesic incompleteness. The following result draws on [27] and [37, theorem 14.55A].

Proposition 2.4. *Let S be a smooth spacelike Cauchy surface in M . (a) If S is compact and every future-complete timelike geodesic emanating normally from S contains a focal point then M is future timelike geodesically incomplete. (b) If every future-directed timelike geodesic emanating normally from S without focal points has length (strictly) less than τ_* then every future-directed timelike curve emanating from S has length (strictly) less than τ_* ; consequently M is future timelike geodesically incomplete.*

Proof.

- (a) Assume that M is future timelike geodesic complete. As discussed in the proof of [15, theorem 5.1], based on results of [27], the hypotheses imply the existence of an S -ray; namely, a future inextendible unit-speed timelike geodesic γ emanating (necessarily normally) from S and which is length-maximising between S and each of its points. Accordingly γ contains no focal points to S , but as it is future-complete by assumption we obtain a contradiction.
- (b) By [37, lemma 14.29 & theorem 14.44], every $q \in D^+(S) \setminus S$ is connected to S by a timelike length-maximising geodesic, which necessarily emanates normally from S and has no focal points before q . By hypothesis, this geodesic has length (strictly) less than τ_* .

Now consider any unit-speed future-directed timelike curve $\gamma : [0, \tau] \rightarrow M$ with $\gamma(0) \in S$, assuming for a contradiction that $\tau > \tau_*$ (or $\tau \geq \tau_*$). But $\gamma(\tau)$ is also joined to S by a length-maximising geodesic of length τ' (strictly) less than τ_* , so we deduce $\tau_* \geq \tau' \geq \tau > \tau_*$ (or $\tau_* > \tau' \geq \tau \geq \tau_*$) thus obtaining a contradiction. In particular, every inextendible future-directed timelike geodesic emanating from S has finite length and is therefore incomplete. \square

Combining corollary 2.3 and proposition 2.4 yields one of Hawking's singularity theorems [28].

Corollary 2.5. *If $\sup_S K < 0$ on S (in particular, if S is compact and $K < 0$) and the time-like convergence condition holds, then M is future-timelike geodesically incomplete.*

The central idea of this paper is that results similar to corollary 2.3, proposition 2.4 and corollary 2.5 may be proved under weaker assumptions on the Ricci tensor, by replacing the linear function f in the proof of corollary 2.3 by a suitable alternative. This will be described in sections 3 and 4 below. Before that, we describe how the above theory can be adapted to null geodesics, again following [37].

2.2. Focal points along null geodesics

Let $\gamma : [0, \ell] \rightarrow M$ be a piecewise smooth curve. Then the integral

$$E[\gamma] = \frac{1}{2} \int_0^\ell g(\gamma'(\lambda), \gamma'(\lambda)) d\lambda, \quad (21)$$

is called the energy or action integral, and is the natural quantity to consider in the variational theory of null geodesics because, unlike the situation for L , $E[\gamma_s]$ varies smoothly in s for any piecewise smooth variation of γ , regardless of the causal nature of γ_s . Let P be a smooth semi-Riemannian submanifold of M , $q \in M \setminus P$ and $\Omega(P, q)$ the manifold of all piecewise smooth curve segments from P to q . Then the critical points of E among curves in $\Omega(P, q)$ are geodesics emanating normally from P , and the second derivative of E in a smooth variation γ_s of such a geodesic $\gamma = \gamma_0$ is [37, proposition 10.39]

$$\left. \frac{\partial^2 E[\gamma_s]}{\partial s^2} \right|_{s=0} = \int_0^\ell [(\nabla_U V_\mu)(\nabla_U V^\mu) - R_{\mu\nu\alpha\beta} U^\mu V^\nu V^\alpha U^\beta] d\lambda + (U_\mu \nabla_V V^\mu) \Big|_0^\ell, \quad (22)$$

where $U^\mu = \gamma'(\lambda)$ and $V^\mu|_{\gamma(\lambda)} = d\gamma_s(\lambda)^\mu / ds|_{s=0}$, and we assume that γ is affinely parametrised. The right-hand side of (22) is, by definition, the Hessian $\mathcal{H}[V]$ of E at critical points; it may also be written

$$\mathcal{H}[V] = \int_0^\ell \left(\frac{DV^\mu}{d\lambda} \frac{DV_\mu}{d\lambda} - R_{\mu\nu\alpha\beta} U^\mu V^\nu V^\alpha U^\beta \right) d\lambda - U_\mu \mathbb{I}^\mu(V, V) \Big|_{\gamma(0)}, \quad (23)$$

where \mathbb{I} is the shape tensor or second fundamental form, defined so that $\mathbb{I}^\mu(V, W)$ is the projection of $\nabla_V W^\mu$ onto the subspace of vectors normal to P , for vector fields V, W tangential to P . The following result is proved as proposition 10.41 of [37].

Proposition 2.6. *Let P be a spacelike submanifold of M . If there are no focal points of P along a normal null geodesic $\gamma \in \Omega(P, q)$, then $\mathcal{H}[V]$ is positive semidefinite when restricted to piecewise smooth V^μ obeying $U_\mu V^\mu \equiv 0$. If $\mathcal{H}[V] = 0$ for some such V^μ , then V^μ is tangent to γ .*

Now let $\gamma : [0, \ell] \rightarrow M$ be a null geodesic affinely parametrized by λ and P a spacelike $(n-2)$ -dimensional submanifold of M . Let e_i ($i = 1, \dots, n-2$) be an orthonormal basis at $T_{\gamma(0)}(P)$. We parallel transport these vectors along γ to get E_i ($i = 1, \dots, n-2$). Let f be a smooth function with $f(0) = 1$ and $f(\ell) = 0$. Then

$$H[fE_i] = \int_0^\ell (-f'^2 - f^2 R_{\mu\nu\alpha\beta} U^\mu E_i^\nu E_i^\alpha U^\beta) d\lambda - U_\mu \mathbb{I}^\mu(E_i, E_i) \Big|_{\gamma(0)}. \quad (24)$$

Now sum over all $i = 1, \dots, n-2$, noting that $g^{\nu\alpha} = U^{(\nu} W^{\alpha)} - \sum_{i=1}^{n-2} E_i^\nu E_i^\alpha$ for a suitably chosen null vector W , to obtain

$$\sum_{i=1}^{n-2} \mathcal{H}[fE_i] = - \int_0^\ell ((n-2)f'^2 - f^2 R_{\mu\nu} U^\mu U^\nu) d\lambda - (n-2)U_\mu H^\mu|_{\gamma(0)}, \quad (25)$$

where

$$H^\mu = \frac{1}{n-2} \sum_{i=1}^{n-2} \mathbb{I}^\mu(E_i, E_i) \quad (26)$$

is the mean normal curvature vector field of P .

This calculation may be used in conjunction with proposition 2.6 to give a sufficient condition for the existence of a focal point along γ , just as in the derivation of proposition 2.2. As there is no unique natural parametrisation of a null geodesic, it is convenient to state the result in an invariant form, regarding γ as an unparametrised 1-dimensional submanifold of M . The notation $d\gamma^\mu$ denotes the line element 1-form on γ , giving the tangent vector $d\gamma^\mu/d\lambda$ with respect to any coordinate λ on γ , while $d\gamma_+^\mu$ is the pseudo-1-form which is equal to $d\gamma^\mu$ with respect to coordinates parametrising γ as a future-directed curve. Expressions such as the right-hand side of (25) may be written in invariant form by regarding $f(\lambda)$ as the coordinate expression of a density f of weight $-\frac{1}{2}$ on γ . Note that the combination $f^2 d\gamma_+$ then defines a vector field along γ , given in coordinates by $f(\lambda)^2 d\gamma^\mu/d\lambda$, provided that $d\gamma^\mu/d\lambda$ is future-pointing. It should be borne in mind that, while a density does not have invariant values at individual points, the sign of the density is invariantly defined; likewise, it makes sense to say that the density vanishes or is nonvanishing at a given point. In this notation, the result we have proved may be stated as follows.

Proposition 2.7. *Let P be a spacelike submanifold of M of co-dimension 2 and let γ be a null geodesic joining $p \in P$ to $q \in J^+(P)$. If there exists a smooth $(-\frac{1}{2})$ -density f on γ which is nonvanishing at p but vanishes at q and so that*

$$\int_\gamma ((n-2)(\nabla_{d\gamma} f)^2 + f^2 \text{Ric}(d\gamma, d\gamma)) \leq -(n-2)g(f^2 d\gamma_+, H)|_p, \quad (27)$$

then there is a focal point to P along γ ; if the inequality holds strictly, then the focal point is located before q .

Note that $|\nabla_{d\gamma} f|$ is a $\frac{1}{2}$ -density on γ , while $\text{Ric}(d\gamma, d\gamma)$ is a 2-density. Accordingly, each term in the integrand of (27) is a density; similarly, $f^2 d\gamma_+^\mu$ is a vector field along γ . Written more explicitly, if γ is parametrised by a coordinate $\lambda \in [0, \ell]$, inequality (27) is

$$\int_0^\ell ((n-2)f'(\lambda)^2 + f(\lambda)^2 R_{\mu\nu} U^\mu U^\nu) d\lambda \leq -(n-2)f(0)^2 U_\mu H^\mu|_p, \quad (28)$$

assuming $U^\mu = d\gamma^\mu/d\lambda$ is future-directed.

Proposition 2.7 may be used to prove the following corollary (see [37, proposition 10.43]). Recall that P is said to be *future-converging* if H^μ is past-pointing timelike everywhere on P . In this situation we may write $H^\mu = H\hat{H}^\mu$ where $H < 0$ and \hat{H}^μ is a future-pointing timelike unit vector. For any future-pointing timelike unit vector V^μ at p , we also write $L_V(\gamma)$ for the length of γ with respect to an affine parameter in which $V_\mu d\gamma^\mu/d\lambda = 1$ at p . We refer to $L_V(\gamma)$ as the *V-length* of γ .

Corollary 2.8. *With P and γ as in proposition 2.7, suppose additionally that P is future-converging and the null convergence condition $\text{Ric}(\text{d}\gamma, \text{d}\gamma) \leq 0$ holds everywhere along γ . If $L_{\hat{H}}(\gamma) \geq 1/|H|$ then there is a focal point to P along γ .*

Proof. Choose an affine coordinate λ on γ , so that $p = \gamma(0)$ and $\hat{H}_\mu \text{d}\gamma^\mu / \text{d}\lambda = 1$. Then $q = \gamma(\ell)$, with $\ell = L_{\hat{H}}(\gamma)$. In these coordinates define $f(\lambda) = 1 - \lambda/\ell$; then the right-hand side of (27) is $-(n-2)H$, while the left-hand side is less than or equal to $(n-2)/\ell$, and the result follows by proposition 2.7. \square

As in section 2.1, inequality (27) may be connected to a Riccati equation related to the Raychaudhuri equation for a null geodesic congruence.

Now we can connect the formation of focal points with future null geodesic incompleteness in the following way, drawing on the formulations in [15, 37].

Proposition 2.9. *Suppose that: (i) M is globally hyperbolic with non-compact Cauchy hypersurfaces; (ii) P is a compact achronal smooth spacelike submanifold of M of co-dimension 2; and (iii) every future-complete null geodesic emanating normally from P contains a focal point to P . Then M is future null geodesically incomplete. If (iii) is replaced by: (iii)', P is future converging and every future-directed null geodesic emanating normally from P with \hat{H} -length at least ℓ contains a focal point to P , then there is an inextendible null geodesic emanating normally from P with \hat{H} -length less than ℓ .*

Proof. As described in the proofs of theorem 5.2 in [15] (see also the comparable part of [37, theorem 14.61]) conditions (i) and (ii) imply that $E^+(P) := J^+(P) \setminus I^+(P)$ is equal to the boundary $\partial J^+(P)$ and is noncompact and closed. These properties were used to show that there is an inextendible affinely parametrised null geodesic $\gamma : [0, a) \rightarrow M$ issuing normally from P and contained entirely in $\partial J^+(P)$ for some $a \in (0, \infty]$. Furthermore, γ can contain no focal points to P , because the portion of γ beyond any focal point would lie in $I^+(P)$ [35, theorem 6.16(a)]⁵ and hence outside $E^+(P)$. Assumption (iii) then entails that γ is not future-complete and the result is proved. Alternatively, (iii)' implies immediately that $L_{\hat{H}}(\gamma) < \ell$. \square

The last two results combine to yield the Penrose singularity theorem [39].

Corollary 2.10. *If, in addition to the assumptions (i) and (ii) of proposition 2.9, P is future-converging and the null convergence condition holds, then M is future null geodesically incomplete.*

Proof. Corollary 2.8 implies assumption (iii) of proposition 2.9. \square

3. Exponential damping

The main goal of this paper is to show how the index form methods described in propositions 2.2 and 2.7 can be used to prove singularity theorems with weaker energy conditions than the SEC or NEC, using much simpler arguments than those used in existing literature. Instead of controlling the (non)existence of solutions to the Raychaudhuri equation, the main method used here is to replace the linear functions used in the proof of corollaries 2.3 and 2.8 by functions that are adapted to the weakened energy conditions under consideration. Our first examples concern exponentially damped half-line averages of the timelike and null

⁵The argument given in [37, proposition 10.48] is not quite complete.

convergence conditions (corresponding to the SEC and NEC respectively), providing similar overall results to those derived in [15] but with much greater ease.

Starting with timelike geodesics, the following result generalises corollaries 2.3 and 2.5 by weakening the timelike convergence condition. It may be compared with lemma 3.1 in [15] combined with theorem 5.1 of the same reference (modified as described in remark (1) following its proof). Our argument here represents a considerable simplification.

Theorem 3.1. *Suppose $\gamma : [0, \infty) \rightarrow M$ is a future-directed unit-speed timelike geodesic emanating normally from a smooth spacelike hypersurface S . If the inequality*

$$\frac{c}{2} + \liminf_{\tau \rightarrow +\infty} \int_0^\tau (1 - t/\tau)^2 e^{-2ct/(n-1)} R_{\mu\nu} U^\mu U^\nu dt < -K|_{\gamma(0)} \quad (29)$$

holds for some $c \geq 0$, then there is a focal point to S along γ .

If S is additionally a compact Cauchy surface and (29) holds along every future-complete timelike unit-speed geodesic emanating normally from S (the value of c may vary) then M is future timelike geodesically incomplete.

Proof. Define $F(\tau)$ to be the integral in equation (29). Then there exists $\epsilon > 0$ and a sequence $\tau_n \rightarrow \infty$ for which

$$F(\tau_n) < -\frac{c}{2} - K|_{\gamma(0)} - \epsilon. \quad (30)$$

Now define

$$f_\tau(t) = (1 - t/\tau) e^{-ct/(n-1)} \quad (31)$$

and note that

$$\lim_{\tau \rightarrow +\infty} \int_0^\tau (n-1) \dot{f}_\tau(t)^2 dt = \frac{c}{2}. \quad (32)$$

Then from equation (30) for $\tau = \tau_n$ with n sufficiently large we have

$$\int_0^\tau \left((n-1) \dot{f}_\tau(t)^2 + f_\tau(t)^2 R_{\mu\nu} U^\mu U^\nu \right) dt < \frac{c}{2} + \epsilon - \frac{c}{2} - K|_{\gamma(0)} - \epsilon = -K|_{\gamma(0)} \quad (33)$$

and therefore there is a focal point before proper time τ by proposition 2.2. The second part of the Theorem follows immediately from proposition 2.4(a). \square

The analogue for null geodesics is:

Theorem 3.2. *Assume that P is a future convergent achronal smooth spacelike codimension-2 submanifold of M . Suppose $\gamma : [0, \infty) \rightarrow M$ is an affinely parametrised null geodesic issuing normally from P , with $\dot{H}_\mu d\gamma^\mu/d\lambda = 1$ for $\lambda = 0$. If the inequality*

$$\frac{c}{2} + \liminf_{\ell \rightarrow +\infty} \int_0^\ell (1 - \lambda/\ell)^2 e^{-2c\lambda/(n-2)} R_{\mu\nu} U^\mu U^\nu d\lambda < -(n-2)H \quad (34)$$

holds for some $c \geq 0$, then there is a focal point to S along γ .

If, additionally, M has a non-compact Cauchy surface, P is compact, and (34) holds along every future-complete null geodesic emanating normally from P , parametrised as described above (the value of c may vary) then M is future null geodesically incomplete.

Table 2. The first few values of the constants A_m , B_m and C_m .

m	1	2	3	4
A_m	1/3	13/35	181/462	521/1287
B_m	1	6/5	10/7	700/429
C_m	1	12	720	100 800

The proof is exactly analogous to the timelike case, but making use of the function $f_\ell(\lambda) = (1 - \lambda/\ell) e^{-c\lambda/(n-2)}$ and propositions 2.7 and 2.9 in place of propositions 2.2 and 2.4(a).

4. QEI inspired hypotheses

In this section we replace the classical energy conditions with QEI-inspired hypotheses that can be obeyed by quantum fields. The main goal is to specify the required initial contraction that leads to geodesic incompleteness.

We will study two scenarios: in the first we suppose that the timelike or null convergence condition is satisfied for a small segment of the geodesic; in the second, we impose conditions on the timelike- or null-contracted Ricci tensor at small negative values of proper time or the affine parameter respectively. Its worth noting that [15] also utilises the same information as in our second scenario to prove singularity theorems.

Our method makes use of trial functions related to incomplete Beta functions, which we now define. For each $m \in \mathbb{N}$, let p_m be the unique polynomial of degree $2m - 1$ with the properties $p_m(0) = 0, p_m(1) = 0, p_m^{(k)}(0) = p_m^{(k)}(1) = 0$ for $1 \leq k \leq m - 1$, given explicitly by

$$p_m(x) = \frac{1}{B(m, m)} \int_0^x y^{m-1} (1-y)^{m-1} dy, \quad (35)$$

where B is the Beta function. This polynomial is the regularised incomplete Beta function, denoted $p_m(x) = I_x(m, m; x)$ (the notation $I_x(m, m)$ is more common but less convenient for us) [9, section 8.17]. In appendix it is shown how the squares of the L^2 norms of p_m , p'_m and $p_m^{(m)}$ on $[0, 1]$ may be computed in closed form with the results

$$\|p_m\|^2 = A_m, \quad \|p'_m\|^2 = B_m, \quad \|p_m^{(m)}\|^2 = C_m, \quad (36)$$

where

$$A_m = \frac{1}{2} - \frac{(2m)!^4}{4(4m)!m!^4}, \quad B_m = \frac{(2m-2)!^2(2m-1)!^2}{(4m-3)!(m-1)!^4}, \quad C_m = \frac{(2m-2)!(2m-1)!}{(m-1)!^2}. \quad (37)$$

The first few relevant values are tabulated in table 2. Note that $A_m < 1/2$ for all m and also that $0 \leq p_m(x) \leq 1$ for $x \in [0, 1]$.

4.1. Timelike geodesics

Instead of the timelike convergence condition, we assume that any deviations from timelike convergence along geodesics satisfy estimates of the form

$$\int_I f(t)^2 R_{\mu\nu} U^\mu U^\nu|_{\gamma(t)} dt \leq \|f\|^2 := Q_m(\gamma) \|f^{(m)}\|^2 + Q_0(\gamma) \|f\|^2, \quad (38)$$

for some $m \in \mathbb{N}$ and all smooth real-valued f supported in the interior of the compact interval $I \subset \mathbb{R}$, where $\gamma : I \rightarrow M$ is a unit-speed timelike geodesic, $Q_m(\gamma) > 0$ and $Q_0(\gamma) \geq 0$ are constants (independent of f , but allowed to depend on γ) of appropriate dimensions and $\|\cdot\|$ denotes the standard norm of $L^2(I)$. By a well known result [1, corollary 6.31], $\|\cdot\|$ is actually a norm on $C_0^\infty(\text{int } I)$ which is equivalent to the standard Sobolev norm $\|f\|_m := (\sum_{r=0}^m \|f^{(r)}\|^2)^{1/2}$ and therefore induces the same topology. Estimates of this type might arise in solutions to the Einstein equations with matter, due to properties of the matter field, such as QEIs or related constraints, see e.g. [6, 17, 18].

Then, assuming that $R_{\mu\nu}U^\mu U^\nu|_{\gamma(t)}$ is continuous on I , a simple approximation argument shows that the same bound (38) holds for all f in the Sobolev space $W_0^m(I)$, which is the closure of $C_0^\infty(\text{int } I)$ in the norm $\|\cdot\|$. Each such function f is $m - 1$ times continuously differentiable on I , with a distributional m 'th derivative that may be identified with an element of $L^2(I)$, and obeys the generalised Dirichlet conditions $f^{(k)}|_{\partial I} = 0$ for $0 \leq k \leq m - 1$; see e.g. [1, theorem 4.12(III)]. As I is compact, $L^2(I) \subset L^1(I)$ so $f^{(m-1)}$ is the indefinite integral of an element of $L^1(I)$, and is therefore absolutely continuous.

Let S be a compact smooth spacelike Cauchy surface in M with extrinsic curvature K . Suppose that $\gamma : [0, \tau] \rightarrow M$ is a unit speed, future-directed, timelike geodesic emanating normally from S and write $\rho(t) = R_{\mu\nu}U^\mu U^\nu|_{\gamma(t)}$. By proposition 2.2, γ contains a focal point to S if there is a piecewise smooth f on $[0, \tau]$ with $f(0) = 1$ and $f(\tau) = 0$, such that

$$J[f] \leq -K|_{\gamma(0)}, \quad (39)$$

where

$$J[f] = \int_0^\tau \left((n-1)\dot{f}(t)^2 + f(t)^2 \rho(t) \right) dt. \quad (40)$$

Our aim is to estimate $J[f]$ for specific functions f defined below, using (38) to control the contribution of ρ . The problem to be faced is that no function with $f(0) = 1$ can belong to $W_0^m([0, \tau])$; we address this by applying (38) to a ‘rounded off’ function and making further assumptions on ρ near $t = 0$ to estimate the error incurred. The rounding off could be performed in many ways. We now turn to the two scenarios mentioned above, beginning with the situation in which timelike convergence holds for small positive values of τ .

4.1.1. Scenario 1. Suppose that $\rho(t) = R_{\mu\nu}U^\mu U^\nu|_{\gamma(t)}$ is a smooth function on $[0, \tau]$ that is initially negative, $\rho \leq \rho_0 \leq 0$ on $[0, \tau_0]$ for some $0 < \tau_0 < \tau$. That is, timelike convergence (equivalent to the SEC) holds initially, with strict inequality if $\rho_0 < 0$.

Defining a piecewise smooth function on $[0, \tau]$ by

$$f(t) = \begin{cases} 1 & t \in [0, \tau_0) \\ I(m, m; (\tau - t)/(\tau - \tau_0)) & t \in [\tau_0, \tau], \end{cases} \quad (41)$$

we note that $f(0) = 1$, $f(\tau) = 0$. We will prove the following:

Lemma 4.1. *For the function f given by (41) and ρ satisfying (38) on $[0, \tau]$ we have*

$$J[f] \leq \nu_* := (1 - A_m)\rho_0\tau_0 + \frac{Q_m(\gamma)C_m}{\tau_0^{2m-1}} + Q_0(\gamma)A_m\tau + \frac{(n-1)B_m}{\tau - \tau_0} + \frac{Q_m(\gamma)C_m}{(\tau - \tau_0)^{2m-1}}. \quad (42)$$

Consequently, if $-K|_{\gamma(0)} \geq \nu_*$ then γ contains a focal point to S .

Proof. The consequences are immediate by proposition 2.2, so it is enough to establish the estimate on $J[f]$. Defining a piecewise smooth function on $[0, \tau]$ by

$$\varphi(t) = \begin{cases} I(m, m; t/\tau_0) & t \in [0, \tau_0] \\ 1 & t \in [\tau_0, \tau], \end{cases} \quad (43)$$

we note that $f\varphi$ is $m - 1$ times continuously differentiable, with $(\varphi f)^{(m)}$ existing and continuous everywhere except $t = \tau_0$, where it has a finite jump discontinuity. Therefore $\varphi f \in W_0^m([0, \tau])$, and writing $f^2 = (\varphi f)^2 + (1 - \varphi^2)f^2 = (\varphi f)^2 + 1 - \varphi^2$, we have

$$\begin{aligned} \int_0^\tau f(t)^2 \rho(t) dt &\leq \int_0^{\tau_0} (1 - \varphi(t)^2) \rho(t) dt + Q_m(\gamma) \|(\varphi f)^{(m)}\|^2 + Q_0(\gamma) \|\varphi f\|^2 \\ &\leq \rho_0 \int_0^{\tau_0} (1 - \varphi(t)^2) dt + Q_m(\gamma) \|(\varphi f)^{(m)}\|^2 + Q_0(\gamma) \|\varphi f\|^2 \end{aligned} \quad (44)$$

using (38), along with the assumption $\rho \leq \rho_0$ on $[0, \tau_0]$. Now φf is equal to φ on $[0, \tau_0]$ and f on $[\tau_0, \tau]$, and the L^2 -norms of derivatives of (φf) decompose accordingly. Indeed, using suitably rescaled versions of (36) (see (A.6)),

$$\|\varphi f\|^2 = A_m \tau_0 + A_m (\tau - \tau_0) = A_m \tau, \quad \|(\varphi f)^{(m)}\|^2 = \frac{C_m}{\tau_0^{2m-1}} + \frac{C_m}{(\tau - \tau_0)^{2m-1}}. \quad (45)$$

Thus, we have

$$\int_0^\tau f(t)^2 \rho(t) dt \leq \rho_0 \tau_0 (1 - A_m) + \frac{Q_m(\gamma) C_m}{\tau_0^{2m-1}} + \frac{Q_m(\gamma) C_m}{(\tau - \tau_0)^{2m-1}} + Q_0(\gamma) A_m \tau \quad (46)$$

and since $\|f'\|^2 = B_m/(\tau - \tau_0)$, the estimate on $J[f]$ is complete. \square

Now using propositions 2.2 and 2.4 the following theorem is immediate.

Theorem 4.2. *Let (M, g) be a smooth globally hyperbolic spacetime of dimension $n > 2$ and let S be a smooth compact spacelike Cauchy surface in M . Suppose that, for some $\tau > 0$, there is an integer $m \geq 1$ and constants Q_m and Q_0 so that:*

- (i) *the Ricci tensor obeys (38) along every unit-speed future-directed timelike geodesic γ of length τ emanating normally from S , with $Q_m(\gamma) \leq Q_m$, $Q_0(\gamma) \leq Q_0$;*
- (ii) *there exists $\rho_0 \leq 0$ and $\tau_0 \in (0, \tau)$ so that along every such geodesic, $R_{\mu\nu} \dot{\gamma}^\mu \dot{\gamma}^\nu|_{\gamma(t)} \leq \rho_0$ for $t \in [0, \tau_0]$;*
- (iii) *the initial extrinsic curvature of S satisfies*

$$-K \geq \min \left\{ \frac{(n-1)}{\tau_0}, \nu_* \right\} \quad (47)$$

everywhere on S , where ν_ is given by (42), but with $Q_k(\gamma)$ replaced by Q_k . (Recall that ν_* depends on τ .)*

Then no future-directed timelike curve emanating from S has length greater than τ and M is future timelike geodesically incomplete.

We conclude this section with two remarks. First, since a focal point in $[0, \tau']$, for $\tau' < \tau$ certainly implies the existence of a focal point in $[0, \tau]$, we can replace ν_* by

$$\nu_{**} := (1 - A_m)\rho_0\tau_0 + \frac{Q_m C_m}{\tau_0^{2m-1}} + \min_{\tau' \in (\tau_0, \tau]} \left(Q_0 A_m \tau' + \frac{(n-1)B_m}{\tau' - \tau_0} + \frac{Q_m C_m}{(\tau' - \tau_0)^{2m-1}} \right)$$

in the statement of lemma 4.1 and theorem 4.2 if we wish. Similarly, If $\gamma : [0, \infty) \rightarrow M$ is future complete and $-K|_{\gamma(0)}$ is greater than or equal to the minimum value of ν_* over $\tau \in (\tau_0, \infty)$ then γ contains a focal point to S .

Second, if $(n-1)/\tau_0 \leq \nu_*$ on S then our result simply reduces to the original Hawking singularity theorem, and no future-directed timelike curve emanating from S has length greater than τ_0 , as in proposition 2.4 and corollary 2.5). Therefore it is important to show that there are situations in which $\nu_* < (n-1)/\tau_0$, so that our result represents a nontrivial improvement.

For this purpose, it is convenient to write $\tau = (1 + \mu B_m)\tau_0$ for $\mu > 0$, whereupon we may discard the term containing ρ_0 in the definition of ν_* to give

$$\nu_* \leq \frac{Q_m C_m}{\tau_0^{2m-1}} + Q_0 A_m \tau_0 + \frac{n-1}{\mu \tau_0} + Q_0 A_m B_m \tau_0 \mu + \frac{Q_m C_m}{(B_m \tau_0)^{2m-1} \mu^{2m-1}}. \quad (48)$$

Clearly $\nu_* \geq (n-1)/\tau_0$ if $\mu \leq 1$, so we restrict to the case $\mu > 1$ and use $1/\mu^{2m-1} < 1/\mu$, obtaining

$$\nu_* \tau_0 \leq E + F\mu + \frac{G}{\mu}, \quad (49)$$

where

$$E = \frac{Q_m C_m}{\tau_0^{2(m-1)}} + A_m Q_0 \tau_0^2, \quad F = A_m B_m Q_0 \tau_0^2, \quad G = n-1 + \frac{Q_m C_m}{B_m^{2m-1} \tau_0^{2(m-1)}}. \quad (50)$$

As the right-hand side of (49) takes its minimum at $\mu = \sqrt{G/F}$, it follows that

$$\nu_* \tau_0 \leq \begin{cases} E + 2\sqrt{FG} & F \leq G \\ E + F + G & G < F \end{cases} \quad (51)$$

(the second case is uninteresting because $G > n-1$). Accordingly, the conditions

$$F \leq G \quad \text{and} \quad E + 2\sqrt{FG} < n-1 \quad (52)$$

imply $\nu_* \tau_0 < n-1$ (evidently this can only be satisfied if $F < (n-1)/4$). In particular, it holds if $Q_0 \tau_0^2 \ll 1$ and $Q_m/\tau_0^{2(m-1)} \ll 1$, in which case the optimising value of μ is much larger than unity,

$$\mu \sim \sqrt{\frac{n-1}{A_m B_m Q_0 \tau_0^2}} \quad (53)$$

leading to the prediction of a focal point within a timescale

$$\tau \sim \sqrt{\frac{(n-1)B_m}{A_m Q_0}}, \quad (54)$$

with initial extrinsic curvature obeying

$$-K|_{\gamma(0)} > \nu_* \sim \frac{n-1}{\tau_0} \sqrt{\frac{4A_m B_m Q_0 \tau_0^2}{n-1}} \ll \frac{n-1}{\tau_0} \quad (55)$$

(assuming that E and F are of comparable order, so that $E \ll \sqrt{F}$). Importantly, this level of extrinsic curvature is not sufficient to guarantee a focal point within time τ_0 , the period in which we have assumed timelike convergence to hold. In this regime, we also have

$$\nu_* \sim \frac{2(n-1)B_m}{\tau}, \quad (56)$$

which is of the same order of magnitude as the initial contraction that would be needed to guarantee a focal point in $(0, \tau]$ assuming timelike convergence (see corollary 2.3). Therefore our result provides a Hawking-type singularity theorem with weakened energy conditions, but without drastically increased contraction.

4.1.2. Scenario 2. Let us now drop the assumption that timelike convergence holds for small positive times along geodesics leaving S . Starting with a single unit-speed timelike geodesic $\gamma : [0, \tau] \rightarrow M$, with $\gamma(0) \in S$, extend γ backwards in time, still as a unit-speed timelike geodesic, to obtain $\gamma : [-\tau_0, \tau] \rightarrow M$. We assume that (38) holds on the extended geodesic, for all $f \in W_0^m([-\tau_0, \tau])$ obeying generalised Dirichlet boundary conditions at $t = -\tau_0, \tau$.

As in Scenario 1, we aim to estimate $J[f]$ for a piecewise smooth function f on $[0, \tau]$ obeying $f(0) = 1, f(\tau) = 0$, namely

$$f(t) = \begin{cases} I(m, m; (\tau' - t)/\tau') & t \in [0, \tau'] \\ 0 & t \in [\tau', \tau], \end{cases} \quad (57)$$

where the free parameter $\tau' \in [0, \tau]$. Defining

$$\hat{L}_1(\tau') = \frac{Q_m(\gamma)C_m}{(\tau')^{2m-1}} + \frac{(n-1)B_m}{\tau'} + A_m Q_0(\gamma)\tau' \quad (58)$$

and

$$\hat{L}_2(\tau'_0) = \frac{Q_m(\gamma)C_m}{(\tau'_0)^{2m-1}} + A_m(Q_0(\gamma) - \rho_{\min})\tau'_0, \quad (59)$$

and also

$$L_1(\tau) = \min_{\tau' \in (0, \tau]} \hat{L}_1(\tau'), \quad L_2(\tau_0) = \min_{\tau'_0 \in (0, \tau_0]} \hat{L}_2(\tau'_0), \quad (60)$$

we prove:

Lemma 4.3. *For the function f given by (57), and ρ satisfying (38) on $[-\tau_0, \tau]$ we have*

$$J[f] \leq \hat{L}_1(\tau') + L_2(\tau_0). \quad (61)$$

Consequently, if $-K|_{\gamma(0)} \geq L_1(\tau) + L_2(\tau_0)$ then there is a focal point to S along γ in the proper time interval $[0, \tau]$.

Proof. We have

$$J[f] = \frac{(n-1)B_m}{\tau'} + \int_0^{\tau'} f^2 R_{\mu\nu} U^\mu U^\nu dt. \quad (62)$$

To estimate the integral, we first extend the domain of f to create a function $\tilde{f} \in W_0^m([-\tau_0, \tau])$ that agrees with f on $[0, \tau]$ and is given by

$$\tilde{f}(t) = \begin{cases} 0 & t \in [-\tau_0, -\tau'_0] \\ I(m, m; -t/\tau'_0) & t \in [-\tau'_0, 0), \end{cases} \quad (63)$$

on $[-\tau_0, 0)$, where $\tau'_0 \in (0, \tau_0]$ is arbitrary. We then estimate

$$\int_0^{\tau'} f^2 R_{\mu\nu} U^\mu U^\nu dt = \int_{-\tau_0}^{\tau} \tilde{f}^2 R_{\mu\nu} U^\mu U^\nu dt - \int_{-\tau'_0}^0 \tilde{f}^2 R_{\mu\nu} U^\mu U^\nu dt \quad (64)$$

$$\leq Q_m(\gamma) \|\tilde{f}^{(m)}\|^2 + Q_0(\gamma) \|\tilde{f}\|^2 - \rho_{\min} \int_{-\tau'_0}^0 \tilde{f}^2 dt \quad (65)$$

where the norms are those of $L^2(-\tau_0, \tau)$ and the constants $Q_k(\gamma)$ refer to the extended geodesic, while $\rho_{\min} = \min_{[-\tau_0, 0]} \rho$. The right-hand side may be evaluated by using our results on incomplete Beta functions, giving an overall upper bound

$$J[f] \leq \hat{L}_1(\tau') + \hat{L}_2(\tau'_0) \quad (66)$$

and the first part of result follows on using the freedom to optimise over τ'_0 . The second is proved using proposition 2.2, and optimising over $\tau' \in (0, \tau]$. \square

The minimum used to define L_2 may be found explicitly, giving

$$L_2(\tau_0) = \frac{2mQ_m(\gamma)}{2m-1} ((Q_0(\gamma) - \rho_{\min})A_m)^{1-1/(2m)} ((2m-1)C_m)^{1/(2m)} \quad (67)$$

for $Q_0(\gamma) - \rho_{\min} > (2m-1)Q_m(\gamma)C_m/(A_m\tau_0^{2m})$, and

$$L_2(\tau_0) = \frac{Q_m(\gamma)C_m}{(\tau_0)^{2m-1}} + A_m(Q_0(\gamma) - \rho_{\min})\tau_0 \quad (68)$$

otherwise. A closed form expression for $L_1(\tau)$ is not possible for general m . However, one may note that \hat{L}_1 has a single minimum in $(0, \infty)$ and no other critical points. As the only positive contribution to \hat{L}'_1 arises from the last term, it follows that if $Q_0\tau^2 \leq (n-1)B_m/A_m$ then the gradient at τ is negative and $L_1(\tau) = \hat{L}_1(\tau)$.

By analogy with theorem 4.2 we now have:

Theorem 4.4. *Let (M, g) be a smooth globally hyperbolic spacetime of dimension $n > 2$ and let S be a smooth compact spacelike Cauchy surface in M . Suppose that, for some $\tau, \tau_0 > 0$ there is an integer $m \geq 1$ and constants Q_m and Q_0 so that*

- (i) *every unit-speed future-directed timelike geodesic of length τ emanating normally from S can be extended as an affine geodesic to $\gamma : [-\tau_0, \tau] \rightarrow M$ with $\gamma(0) \in S$, along which the Ricci tensor obeys (38) with $Q_m(\gamma) \leq Q_m$, $Q_0(\gamma) \leq Q_0$;*
- (ii) *there exists a finite lower bound ρ_{\min} so that $R_{\mu\nu}\dot{\gamma}^\mu\dot{\gamma}^\nu|_{\gamma(t)} \geq \rho_{\min}$ for every such geodesic and all $t \in [-\tau_0, 0]$ (we emphasise that ρ_{\min} is uniform in γ);*
- (iii) *the extrinsic curvature of S satisfies*

$$-K \geq L_1(\tau) + L_2(\tau_0) \quad (69)$$

everywhere on S , where the functions L_i are given by (58)–(60), with $Q_k(\gamma)$ replaced by Q_k ($k = 0, m$).

Then no future-directed timelike curve emanating from S has length greater than τ and M is future timelike geodesically incomplete.

Under the assumption of timelike convergence, the conclusions of this result hold provided the initial contraction satisfies $-K \geq (n-1)/\tau$ on S (corollary 2.3 and proposition 2.4). It is important to note that there are circumstances in which the contraction required by theorem 4.4 is of a similar order.

For example, suppose that $\rho_{\min} \geq 0$ and the following conditions hold

$$Q_0\tau^2 \ll B_m/A_m, \quad \tau_0 \ll \tau \lesssim \frac{\tau_0^{2m-1}}{Q_m C_m}. \quad (70)$$

Then certainly $Q_0\tau^2 \leq B_m/A_m$, so $L_1(\tau) = \hat{L}_1(\tau)$ as noted above. As $Q_m C_m / \tau^{2(m-1)} \lesssim (\tau_0/\tau)^{2(m-1)} \ll 1$ (for $m > 1$) we find that $L_1(\tau) = \hat{L}_1(\tau) \sim (n-1)B_m/\tau$. In the case $m = 1$ the assumptions imply $Q_1 \ll 1$ and hence $Q_1 C_1 / \tau \ll (n-1)/\tau$, giving the same conclusion for $L_1(\tau)$ as before. Turning to L_2 , the requirement that $\rho_{\min} \geq 0$ implies

$$L_2(\tau_0) \leq \hat{L}_2(\tau_0) \leq \frac{Q_m C_m}{(\tau_0)^{2m-1}} + A_m Q_0 \tau_0. \quad (71)$$

Under our assumptions, the second term is small relative to B_m/τ , while the first is $\lesssim 1/\tau$. Overall,

$$L_1(\tau) + L_2(\tau_0) \lesssim \frac{(n-1)B_m + 1}{\tau}, \quad (72)$$

which is of comparable order to $(n-1)/\tau$ at least for small values of m . Thus the contraction required by theorem 4.4 is comparable to $(n-1)/\tau$.

A feature of theorem 4.4 is that increasing ρ_{\min} —the extent to which timelike convergence fails at small negative times—decreases the required initial contraction. This is because the QEI type constraints require that a period in which timelike convergence fails is followed shortly by a period in which timelike convergence is satisfied even more strongly. This is analogous to the quantum interest effect [6, 15, 22, 25]. If $\rho_{\min} \leq 0$ then theorem 4.4 typically overestimates the required contraction and scenario 1 should be used instead.

4.2. Null geodesics

Let γ be any null geodesic in M . Using the same invariant notation as in proposition 2.7, we suppose that the Ricci tensor obeys

$$\int_{\gamma} f^2 \text{Ric}(d\gamma, d\gamma) \leq \|f\|^2 := Q_m(\gamma; w) \|w^{1-m} \nabla_{d\gamma}^m f\|^2 + Q_0(\gamma; w) \|wf\|^2, \quad (73)$$

for every smooth compactly supported $(-\frac{1}{2})$ -density f on γ , and every choice of smooth positive density w on γ , so that $Q_k(\gamma; \lambda w) = \lambda^{2(k-1)} Q_k(\gamma; w)$ for any $\lambda > 0$. Here $\|\cdot\|$ is the intrinsic L^2 -norm on $\frac{1}{2}$ -densities defined on γ , and appropriate powers of the density w are inserted to ensure that the arguments of these norms have the correct weight as densities. It is assumed that $Q_m(\gamma; w) > 0$ and $Q_0(\gamma; w) \geq 0$.

Now let P be a future converging achronal spacelike submanifold of M of co-dimension 2 with mean normal curvature vector field H^μ . Suppose that γ is a future-directed null geodesic emanating normally from P . Extending \hat{H}_μ by parallel transport along γ , we define a density $w = \hat{H}_\mu d\gamma^\mu / d\lambda$. Next choose an affine coordinate λ on γ , so that $p = \gamma(0)$ and $\hat{H}_\mu d\gamma^\mu / d\lambda = 1$. Then $w = 1$, $q = \gamma(\ell)$, with $\ell = L_{\hat{H}}(\gamma)$ and equation (73) becomes

$$\int_0^\ell f(\lambda)^2 R_{\mu\nu} U^\mu U^\nu d\lambda \leq Q_m(\gamma) \|f^{(m)}\|^2 + Q_0(\gamma) \|f\|^2. \quad (74)$$

Writing $\rho(\lambda) = R_{\mu\nu} U^\mu U^\nu|_{\gamma(\lambda)}$, and by proposition 2.7, there is a focal point to P along γ if there is a piecewise smooth f on $[0, \ell]$ with $f(0) = 1$ and $f(\ell) = 0$, such that

$$J[f] \leq -(n-2)H|_{\gamma(0)}, \quad (75)$$

where

$$J[f] = \int_0^\ell ((n-2)f'(\lambda)^2 + f(\lambda)^2 \rho(\lambda)) d\lambda. \quad (76)$$

The two scenarios following are analogous to the ones in section 4.1. In the first, we will suppose that the NEC is satisfied at small positive values of λ while in the second, we impose conditions on ρ at small negative values of λ .

4.2.1. Scenario 1. Suppose that initially the NEC is satisfied, and so let $\rho(\lambda) = R_{\mu\nu} U^\mu U^\nu|_{\gamma(\lambda)}$ be a smooth function on $[0, \ell]$ that is initially negative, $\rho \leq \rho_0 \leq 0$ on $[0, \ell_0]$ for some $0 < \ell_0 < \ell$.

Defining $f : [0, \ell] \rightarrow \mathbb{R}$ by (41) with ℓ instead of τ and ℓ_0 instead of τ_0 , we can estimate $J[f]$ following the exact steps of the proof of lemma 4.1. The following lemma is immediate using proposition 2.7.

Lemma 4.5. *For ρ satisfying (74) on $[0, \ell]$ we have*

$$J[f] \leq \nu_* := (1 - A_m)\rho_0\ell_0 + \frac{Q_m C_m}{\ell_0^{2m-1}} + Q_0 A_m \ell + \frac{(n-2)B_m}{\ell - \ell_0} + \frac{Q_m C_m}{(\ell - \ell_0)^{2m-1}}. \quad (77)$$

Consequently, if $-(n-2)H|_{\gamma(0)} \geq \nu_*$ then γ contains a focal point to P .

There is an obvious adaptation of this result to future-complete null geodesics by minimising over $\ell \in (\ell_0, \infty)$. Using propositions 2.7 and 2.9 the following theorem is immediate.

Theorem 4.6. *Let M be globally hyperbolic with non-compact Cauchy hypersurfaces and let P be a compact achronal future converging spacelike submanifold of M of co-dimension 2 with mean normal curvature vector $H^\mu = H\hat{H}^\mu$ where \hat{H}^μ is a future-pointing timelike unit vector. Suppose that for some $\ell > 0$ there is an integer $m \geq 1$ and constants Q_m and Q_0 so that:*

- (i) *the Ricci tensor obeys (73) along every future-directed null geodesic γ of \hat{H} -length ℓ emanating normally from P , with $Q_m(\gamma) \leq Q_m$, $Q_0(\gamma) \leq Q_0$;*
- (ii) *there exists $\rho_0 \leq 0$ and $\ell_0 \in (0, \ell)$ so that along every such geodesic γ , the inequality $\text{Ric}(\gamma', \gamma') \leq \rho_0 g(d\gamma, \hat{H})^2$ holds along the initial portion of γ with \hat{H} -length ℓ_0 ;*
- (iii) *the mean normal curvature of P satisfies*

$$-(n-2)H \geq \min \left\{ \frac{n-2}{\ell_0}, \nu_* \right\} \quad (78)$$

everywhere on P , where ν_ is given by (77) and depends on ℓ .*

Then there is an inextendible future-directed null geodesic emanating from P with \hat{H} -length less than ℓ . In particular, M is future null geodesically incomplete.

Similarly to the timelike version if $(n-2)/\ell_0 \leq \nu_*$ on P our result reduces to the original Penrose singularity theorem as in proposition 2.9. So we can examine situations in which $(n-2)/\ell_0 \geq \nu_*$ following the exact similar analysis as in section 4.1.1 but again with λ instead of t , ℓ instead of τ , ℓ_0 instead of τ_0 and equation (74) instead of (38).

Then for $Q_0\ell_0^2 \ll 1$ and $Q_m/\ell_0^{2(m-1)} \ll 1$ there is a focal point before parameter

$$\ell \sim \sqrt{\frac{(n-2)B_m}{A_m Q_0}}, \quad (79)$$

if the magnitude of the mean curvature vector of P satisfies

$$-(n-2)H > \nu_* \sim \sqrt{4A_m B_m Q_0(n-2)} \sim \frac{2B_m(n-2)}{\ell}. \quad (80)$$

Again, this is of the same order as the mean curvature needed to guarantee a focal point if the null convergence condition held (corollary 2.8).

4.2.2. Scenario 2. In Scenario 2 we drop the assumption that the NEC holds. We instead extend the previously defined γ to negative values of λ to get $\gamma : [-\ell_0, \ell] \rightarrow M$. Next we assume that equation (74) holds on the extended geodesic, for all $f \in W_0^m([-\ell_0, \ell])$ obeying generalised Dirichlet boundary conditions at $\lambda = -\ell_0, \ell$.

Similarly to the timelike case we want to estimate $J[f]$ and we can follow the same analysis with λ , ℓ and ℓ_0 instead of t , τ and τ_0 . Set

$$\hat{L}_1(\ell') = \frac{Q_m(\gamma)C_m}{(\ell')^{2m-1}} + \frac{(n-2)B_m}{\ell'} + A_m Q_0(\gamma)\ell' \quad (81)$$

and

$$\hat{L}_2(\ell'_0) = \frac{Q_m(\gamma)C_m}{(\ell'_0)^{2m-1}} + A_m(Q_0(\gamma) - \rho_{\min})\ell'_0, \quad (82)$$

and also

$$L_1(\ell) = \min_{\ell' \in (0, \ell]} \hat{L}_1(\ell'), \quad L_2(\ell_0) = \min_{\ell'_0 \in (0, \ell_0]} \hat{L}_2(\ell'_0). \quad (83)$$

Then the proof of the following lemma is analogous to the one for lemma 4.3 where equation (74) is used instead of (38), and finally using proposition 2.7, optimising over $\ell' \in (0, \ell]$.

Lemma 4.7. *For ρ satisfying (74) on $[-\ell_0, \ell]$ we have*

$$J[f] \leq \hat{L}_1(\ell') + L_2(\ell_0). \quad (84)$$

Consequently, if $-(n-2)H \geq L_1(\ell) + L_2(\ell_0)$ then there is a focal point to P along γ in $[0, \ell]$.

More explicitly, we have

$$L_2(\ell_0) = \frac{2mQ_m(\gamma)}{2m-1} ((Q_0(\gamma) - \rho_{\min})A_m)^{1-1/(2m)} ((2m-1)C_m)^{1/(2m)} \quad (85)$$

for $Q_0(\gamma) - \rho_{\min} > (2m-1)Q_m(\gamma)C_m/(A_m\ell_0^{2m})$, and

$$L_2(\ell_0) = \frac{Q_m(\gamma)C_m}{(\ell_0)^{2m-1}} + A_m(Q_0(\gamma) - \rho_{\min})\ell_0 \quad (86)$$

otherwise. Our main result in this scenario is:

Theorem 4.8. *Let M be globally hyperbolic with non-compact Cauchy hypersurfaces and let P be a compact achronal future converging spacelike submanifold of M of co-dimension 2 with mean normal curvature vector H^μ . Suppose that for some $\ell, \ell_0 > 0$ there is an integer $m \geq 1$ and constants Q_m and Q_0 so that:*

- (i) *every future-directed null geodesic of \hat{H} -length ℓ emanating normally from P may be extended to the past, to give a geodesic γ with $L_{\hat{H}}(\gamma) = \ell + \ell_0$ so that the Ricci tensor obeys (73) along γ with $Q_m(\gamma) \leq Q_m$, $Q_0(\gamma) \leq Q_0$;*
- (ii) *there exists a finite lower bound ρ_{\min} so that $\text{Ric}(\text{d}\gamma, \text{d}\gamma) \geq \rho_{\min}g(\hat{H}, \text{d}\gamma)^2$ on the portion of each such geodesic γ to the past of P (we emphasise that ρ_{\min} is uniform in γ);*
- (iii) *the mean normal curvature of P satisfies*

$$-(n-2)H \geq L_1(\ell) + L_2(\ell_0), \quad (87)$$

where the functions L_i are given by (81)–(83), with $Q_k(\gamma)$ replaced by Q_k ($k = 0, m$).

Then there is an inextendible future-directed null geodesic emanating from P with \hat{H} -length less than ℓ . In particular, M is future null geodesically incomplete.

Similarly to the timelike case it is important to note that there are circumstances in which the mean normal curvature required by theorem 4.8 is of the same order as that in corollary 2.8.

Suppose that $\rho_{\min} \geq 0$ and the following conditions hold

$$Q_0\ell^2 \ll B_m/A_m, \quad \ell_0 \ll \ell \lesssim \frac{\ell_0^{2m-1}}{Q_m C_m}. \quad (88)$$

Then $Q_0\ell^2 \leq B_m/A_m$, so $L_1(\ell) = \hat{L}_1(\ell)$. Similar to the timelike case $L_1(\ell) = \hat{L}_1(\ell) \sim (n-2)B_m/\ell$.

Turning to L_2 we have

$$L_2(\ell_0) \leq \hat{L}_2(\ell_0) \leq \frac{Q_m C_m}{(\ell_0)^{2m-1}} + A_m Q_0 \ell_0, \quad (89)$$

and overall

$$L_1(\ell) + L_2(\ell_0) \lesssim \frac{(n-2)B_m + 1}{\ell}, \quad (90)$$

which is of comparable order to $(n-2)/\ell$ at least for small values of m .

If $\rho_{\min} \leq 0$ then theorem 4.8 typically overestimates the required H and scenario 1 should be used.

5. Applications to the Einstein–Klein–Gordon theory

In this section we apply the results of section 4 to the non-minimally coupled classical Einstein–Klein–Gordon theory. The classical non-minimally coupled scalar field is a famous example that violates both the SEC and the NEC, thus lying outside the scope of the original singularity theorems. Singularity theorems with weakened energy conditions were proven for the null case in [15] and for the timelike case in [6]. Here we calculate the required initial

contraction for singularity formation using the method described in previous sections and compare our results with [15] and [6]. This will help illustrate the application of our method and its advantages compared to methods using the Raychaudhuri equation.

Let (M, g, ϕ) be a solution to the Einstein–Klein–Gordon equation in $n > 2$ spacetime dimensions, i.e. $G_{\mu\nu} = -8\pi T_{\mu\nu}$, where the right-hand side is the stress–energy tensor

$$T_{\mu\nu} = (\nabla_\mu \phi)(\nabla_\nu \phi) + \frac{1}{2}g_{\mu\nu}(\lambda^{-2}\phi^2 - (\nabla\phi)^2) + \xi(g_{\mu\nu}\square_g - \nabla_\mu \nabla_\nu - G_{\mu\nu})\phi^2, \quad (91)$$

for the nonminimally coupled scalar field ϕ obeying the equation

$$(\square + \lambda^{-2} + \xi R)\phi = 0, \quad (92)$$

where λ is a fixed characteristic wavelength and ξ the coupling constant. We suppose that $\xi \in [0, \xi_c]$ where $\xi_c = \frac{1}{4}(n-2)/(n-1)$ is the value for conformal coupling.

5.1. The timelike case

Consider first the timelike case. Assume that the scalar field magnitude admits global bounds on a timelike geodesic γ parametrized by proper time τ

$$|\phi| \leq \phi_{\max} \leq (8\pi\xi)^{1/2}, \quad |\nabla_{\dot{\gamma}}\phi| \leq \phi'_{\max}. \quad (93)$$

Then, on γ , and following theorem 3 and corollary 1 of [6], we have that

$$\int_{\gamma} R_{\mu\nu} \dot{\gamma}^\mu \dot{\gamma}^\nu f(\tau)^2 d\tau \leq Q(\|\dot{f}\|^2 + \tilde{Q}^2 \|f\|^2), \quad (94)$$

for any real valued function f of compact support, with constants Q and \tilde{Q} given by

$$Q = \frac{32\pi\xi\phi_{\max}^2}{1 - 8\pi\xi\phi_{\max}^2}, \quad \tilde{Q}^2 = \frac{(1 - 2\xi)\lambda^{-2}}{4\xi(n-2)} + \left(\frac{8\pi\xi\phi_{\max}\phi'_{\max}}{1 - 8\pi\xi\phi_{\max}^2} \right)^2. \quad (95)$$

Equation (94) is of the form of equation (38) with $m = 1$, $Q_1 = Q$ and $Q_0 = Q\tilde{Q}^2$. To estimate the values of these constants we first reinsert the units and restore the constants G and c

$$Q = \frac{32\pi\xi G\phi_{\max}^2/c^4}{1 - 8\pi\xi G\phi_{\max}^2/c^4}, \quad \tilde{Q}^2 = \frac{(1 - 2\xi)\lambda^{-2}c^2}{4\xi(n-2)} + \left(\frac{8\pi\xi G\phi_{\max}\phi'_{\max}/c^4}{1 - 8\pi\xi G\phi_{\max}^2/c^4} \right)^2. \quad (96)$$

It is interesting to estimate these values in a situation where λ is the reduced Compton wavelength of an elementary particle, making appropriate choices for ϕ_{\max} and ϕ'_{\max} . In [6] the value of ϕ_{\max} was estimated by considering a quantized scalar field in Minkowski space-time of dimension n , in a thermal state of temperature $T < T_\lambda$, where $T_\lambda = c\hbar/(\lambda k)$ is the reduced Compton temperature of the particle and k is Boltzmann's constant. The temperature T_λ defines a scale beyond which the model cannot be trusted. With these considerations $\phi_{\max}^2 \sim \langle :\phi^2: \rangle_T$ and

$$Q \sim (\ell_{\text{Pl}}/\lambda)^{n-2} (T/T_\lambda)^{(n-2)/2} K_{(n-2)/2}(T_\lambda/T), \quad (97)$$

as was shown in [6] (in that reference the mass was used instead of the reduced Compton length). Here K_ν is a modified Bessel function of the second kind and ℓ_{Pl} is the Planck length.

From equation (96) the second term of \tilde{Q}^2 is proportional to $Q^2(\phi'_{\max}/\phi_{\max})^2$. On dimensional grounds the ratio ϕ'_{\max}/ϕ_{\max} is proportional to c/λ . For λ taken as the Compton length of

elementary particles $Q \ll 1$, as will be seen shortly. Therefore the first term of \tilde{Q}^2 is expected to be much larger than the second and so $\tilde{Q} \sim c/\lambda$.

In order to give a quantitative, and partly heuristic, illustration of our results, we consider the following toy model. Suppose the universe were described by a Einstein–Klein–Gordon model, in which the characteristic length scale of the scalar is the reduced Compton wavelength of a pion. Given an expansion rate of the universe, and other conditions, drawn from actual cosmological data, would one be able to conclude that the universe is necessarily past timelike geodesically incomplete? To do this we must consider the time-reverse of the analysis presented above, so the question is whether the extrinsic curvature of surfaces of constant cosmic time—the Hubble parameter—is sufficiently positive; that is, whether the expansion rate is sufficiently large. We bear in mind that the SEC is violated in our actual universe, due to the dominant effect of dark energy, with the ratio of pressure to energy density being very close to -1 . This motivates two different calculations: (a) using parameters drawn from an era in which the SEC did hold, as an instance of Scenario 1, and (b) using parameters corresponding to the present time, and assuming that the SEC will continue to fail for some time τ_0 into the future, as a (time-reversed version of) Scenario 2. We will show in each case that, using the pion as the matter model, the expansion rate of the actual universe would be sufficient to conclude past geodesic incompleteness. In fact, there is a caveat to these results, because the values of parameters Q_0 and Q_1 were derived on the basis that the temperature scale T is not exceeded. This means that what the heuristic argument actually shows is that, on timescales within the age of our actual universe, the toy model universe must display either past geodesic timelike incompleteness on a timescale of the age of our actual universe, or locations where the temperature scales exceed T . For brevity, we will describe either of these occurrences as a singularity.

First, we consider Scenario 1, in which the SEC is satisfied for time τ_0 . For a neutral pion in $n = 4$ dimensions with mass $135 \text{ MeV}/c^2$, we have $\ell_{\text{pl}}/\lambda = 1.11 \times 10^{-20}$ and $T_\lambda = 1.56 \times 10^{12} \text{ K}$. (All calculations are made to higher precision but reported to 3S.F.; however it is really the orders of magnitude that are of interest.) Taking $T = 10^{-2} T_\lambda$ (corresponding to the temperature of our universe about 1 s after the big bang) gives an estimate of $Q \sim 5.66 \times 10^{-87}$. Then $Q_1 = Q \ll 1$ and $Q_0 \sim Q(c/\lambda)^2 = 2.39 \times 10^{-39} \text{ s}^{-2}$. Thus if τ_0 is of the order of the reduced Compton time ($4.87 \times 10^{-24} \text{ s}$), we have $Q_0 \tau_0^2 \sim Q \ll 1$. In that case the initial contraction is given by equation (55) with the units restored

$$\nu_* \sim \lambda^{-1} c \sqrt{12 A_1 B_1 Q} \sim 3.09 \times 10^{-20} \text{ s}^{-1}. \quad (98)$$

The maximum allowed temperature for this case is $T = 1.56 \times 10^{10} \text{ K}$, while the timescale on which the singularity occurs is given by equation (54) as

$$\tau \sim \lambda c^{-1} \sqrt{\frac{3 B_1}{A_1}} Q^{-1/2} \sim 1.94 \times 10^{20} \text{ s}, \quad (99)$$

or about 6.15×10^{12} years.

For comparison, in our actual universe, and assuming the Λ CDM model, the SEC was most recently obeyed at time t_1 when

$$\Omega_\Lambda(t_1) = \frac{\Omega_b(t_1)}{2} + \Omega_r(t_1). \quad (100)$$

Here $\Omega_x = \rho_x/\rho_{\text{crit}}$, ρ_{crit} is the critical energy density, Λ corresponds to dark energy, m to matter (baryonic and cold dark matter) and r to radiation. Using the different evolution of each energy density component we find that the redshift when the SEC was last satisfied z_1 is given by the solution of

$$\Omega_{\Lambda_0} - \frac{\Omega_{m_0}}{2}(z_1 + 1)^3 - \Omega_{r_0}(z_1 + 1)^4 = 0, \quad (101)$$

where Ω_{x0} are the respective quantities today. From the most recent results published by the PLANCK collaboration [2] we have that $\Omega_{\Lambda_0} = 0.6889 \pm 0.0056$ and $\Omega_{m_0} = 0.311 \pm 0.002$ which gives a redshift of $z_1 = 0.642$ for when the SEC was last obeyed. From the first Friedmann equation we have

$$\frac{H^2(t_1)}{H_0^2} = \Omega_{\Lambda_0} + \Omega_{r_0}(z_1 + 1)^4 + \Omega_{m_0}(z_1 + 1)^3, \quad (102)$$

where $H(t)$ is the Hubble parameter and H_0 its value today. Again from the PLANCK collaboration [2] we get that $H_0 = (2.184 \pm 0.016) \times 10^{-18} \text{ s}^{-1}$. That gives $H(t_1) = 3.14 \times 10^{-18} \text{ s}^{-1}$ at the time when the SEC was last obeyed. This exceeds the minimum threshold (98) by two orders of magnitude, so the toy model universe would necessarily have a past singularity, using these parameters. Partly because the threshold is exceeded by such a margin, the estimated timescale (99) for the location of the singularity is accordingly rather pessimistic. Nonetheless, it is remarkable that it is within two orders of magnitude of the age of the universe at the relevant time.

If the characteristic scale is replaced by the reduced Compton wavelength of more massive particles, the same calculations produce higher expansion thresholds beyond which a singularity is inevitable. For a proton, with mass $938 \text{ MeV}/c^2$, and a maximum temperature $T = 1.09 \times 10^{11} \text{ K}$, the threshold is $1.50 \times 10^{-18} \text{ s}^{-1}$, which is still marginally exceeded by the measured value. The timescale for the singularity is now 1.27×10^{11} years. On the other hand, for a Higgs particle, with mass $125 \text{ GeV}/c^2$ and with $T = 1.44 \times 10^{13} \text{ K}$ (the temperature of the Universe at age 10^{-4} s), the threshold is $\nu_* = 2.68 \times 10^{-14} \text{ s}^{-1}$. This threshold is larger than the observed Hubble parameter by 4 orders of magnitude, so our results would be inconclusive in that case. These illustrations are intended purely to show that the results we have obtained are quantitative and capable of producing plausible cosmological results in the toy models for the thresholds beyond which a singularity is inevitable, and the timescales on which they occur.

Moving to Scenario 2 where the requirement that the SEC holds is dropped, we are taking into account the behaviour of ρ just *after* the time $\tau = 0$ at which the Hubble parameter is measured (recall that we are using time-reversed versions of our earlier results). Here we assume that the SEC is violated for the time interval $[0, \tau_0]$, and $\rho_{\min} > 0$ which as we mentioned is compatible with current cosmological observations.

First we observe that for the pion $Q_0^{-1/2} \sim 6.49 \times 10^{11} \text{ years}$ is large in comparison to the lifetime of the universe. Then $Q_0\tau^2 \ll B_1/A_1$ for focal points occurring within a timescale τ of the order of ten times the age of the universe. As Q_1 is so small, there is a large range of possible choices of τ_0 so that $\tau_0 \ll \tau < \tau_0/(Q_1 C_1)$ and so that SEC fails for times in $[0, \tau_0]$, from 10^{-68} s up to $10^{-2}\tau$, say.

Given these assumptions the approximation of equation (72) is reasonable, and our results show that a singularity is inevitable within about 10 times the age of the universe, for Hubble parameter

$$H(0) > \frac{3B_1 + 1}{\tau} \sim 10^{-18} \text{ s}^{-1}, \quad (103)$$

which on the order of the current values from PLANCK $(2.184 \pm 0.016) \times 10^{-18} \text{ s}^{-1}$ [2].

If the characteristic scale is based on the proton mass, $Q_0\tau^2 \sim 0.1$ if τ is the age of the universe; however the minimum threshold on $H(0)$ is now $\sim 10^{-17} \text{ s}^{-1}$, so the measured

value would not be sufficient to conclude the existence of a singularity. For the Higgs field, $Q_0^{-1/2} \sim 2.36 \times 10^6$ years and the approximation $Q_0 \tau^2 \ll 1$ is only valid for τ less than 10^{-4} times the age of the universe, requiring therefore approximately 10^4 times the measured Hubble parameter to infer that a singularity is inevitable in our toy model.

These results give similar orders of magnitude for the required initial extrinsic curvature to those computed in [6] obtained with different methods (and phrasing the conditions for future timelike geodesic incompleteness). An advantage of the current method is that it can specify the timescale on which the focal points appear.

5.2. The null case

Now we turn to the null case. For any solution to the Einstein–Klein–Gordon equation in which the field magnitude obeys a global bound $|\phi| \leq \phi_{\max} < (8\pi\xi)^{1/2}$, it was shown in [15] that

$$\int_{\gamma} f^2 \text{Ric}(\text{d}\gamma, \text{d}\gamma) \leq 16\pi\xi\phi_{\max}^2 \int_{\gamma} \left(\nabla_{\text{d}\gamma} \frac{f}{\sqrt{1-8\pi\xi\phi^2}} \right)^2, \quad (104)$$

for all smooth compactly supported $(-\frac{1}{2})$ -densities f on γ , where we have written the expression derived in [15] invariantly and adapted it to our sign conventions. Given any positive density w on γ this implies

$$\int_{\gamma} f^2 \text{Ric}(\text{d}\gamma, \text{d}\gamma) \leq Q(\|\nabla_{\text{d}\gamma} f\|^2 + \tilde{Q}(\gamma; w)^2 \|wf\|^2), \quad (105)$$

where

$$Q = \frac{32\pi\xi\phi_{\max}^2}{1-8\pi\xi\phi_{\max}^2}, \quad \tilde{Q}(\gamma; w) = \frac{8\pi\xi\phi_{\max}}{1-8\pi\xi\phi_{\max}^2} \sup_{\gamma} \frac{|\nabla_{\text{d}\gamma}\phi|}{w}. \quad (106)$$

Equation (105) is of the form of equation (73) with $m = 1$, $Q_1(\gamma; w) = Q$ and $Q_0(\gamma; w) = Q\tilde{Q}(\gamma; w)^2$. Note that Q_1 in this case is independent of both γ and w , while $Q_0(\gamma; w)$ is independent of any specific parametrisation of γ .

As described in section 2.2 we fix the affine parametrization and w in the following way: define P to be a future converging achronal spacelike submanifold of M of co-dimension 2 and let γ be a future-directed null geodesic emanating normally from P . Extending \hat{H}_{μ} by parallel transport along γ and choosing an affine coordinate λ , so that $p = \gamma(0)$ and $\hat{H}_{\mu} \text{d}\gamma^{\mu} / \text{d}\lambda = 1$, we have $w = 1$ and $q = \gamma(\ell)$, with $\ell = L_{\hat{H}}(\gamma)$.

Then equation (105) becomes

$$\int_{\gamma} f^2(\lambda) R_{\mu\nu} U^{\mu} U^{\nu} \text{d}\lambda \leq Q(\gamma) (\|f'\|^2 + \tilde{Q}^2(\gamma) \|f\|^2), \quad (107)$$

and

$$\tilde{Q}(\gamma) = \frac{8\pi\xi\phi_{\max}}{1-8\pi\xi\phi_{\max}^2} \sup_{\gamma} |\phi'(\lambda)|. \quad (108)$$

Now we want to estimate Q and \tilde{Q} . We consider a massless field and, as in the timelike case, we work in a hybrid model: a quantized scalar field in a thermal state of temperature T . In the massless scalar field case the Wick square of a KMS state with temperature T is

$$\langle \phi^2 \rangle_T = \frac{T^{n-2}}{2^{n-2} \pi^{(n-1)/2}} \frac{\Gamma(n-2)}{\Gamma(\frac{n-2}{2})} \zeta(n-2), \quad (109)$$

where ζ is the Riemann zeta function. Similarly, if U^μ is any null vector with $U^0 = 1$ then

$$\langle (U^\mu \nabla_\mu \phi)(U^\nu \nabla_\nu \phi) \rangle_T = \frac{T^n}{3 \cdot 2^{n-4} \pi^{(n-1)/2}} \frac{\Gamma(n)}{\Gamma(\frac{n-1}{2})} \zeta(n). \quad (110)$$

For $\phi_{\max}^2 \sim \langle \phi^2 \rangle_T$, $\phi_{\max}^2 \sim \langle (U^\mu \nabla_\mu \phi)(U^\nu \nabla_\nu \phi) \rangle_T$ and restoring the units we have

$$Q \sim (T/T_{\text{pl}})^{n-2}, \quad \text{and} \quad \tilde{Q} \sim Q \frac{kT}{\hbar}, \quad (111)$$

where T_{pl} is the Planck temperature.

Let us consider Scenario 1. For $n = 4$, and a temperature $T \sim 10^7 \text{K}$ which is of the order of a newly formed neutron star [38] we have $Q_1 = Q \sim 10^{-50} \ll 1$ and $Q_0 \sim 10^{-114} \text{s}^{-2}$. We can consider ℓ_0 as a measurement in light seconds of distance along null rays, measured by an observer at rest on the hypersurface P . Then assuming it is much larger than 10^{-57} light seconds, we have $Q_0 \ell_0^2 \ll 1$.

The required magnitude of the mean curvature vector of P to form a focal point is given by equation (80) with the units restored

$$-H \sim \frac{kT}{2\hbar} \sqrt{8A_1 B_1 Q^3} \sim 10^{-57} \text{s}^{-1}. \quad (112)$$

The contraction in [15] was found to be

$$-H > \tilde{Q} \sqrt{Q(Q + (n-2))} + QK \coth(K\ell_0), \quad (113)$$

where we have corrected some factors of 2. Using the previous values of Q and \tilde{Q} and assuming that the second term which depends on the history of the solution does not get too large we get $-H > (kT/\hbar) Q^{3/2} \sqrt{2} \sim 10^{-57} \text{s}^{-1}$ which agrees with our estimate.

The required mean normal curvature for this toy model is extremely small, scarcely more restrictive than being a trapped surface. Of course a model of a massless scalar should not be taken seriously as a model of astrophysical black hole formation which involves multiple species of interacting particles. However, it shows that a model where the NEC is violated can still lead to geodesic incompleteness with very weak restrictions on the initial conditions.

6. Conclusion

In this work we derived singularity theorems with weakened energy conditions inspired by QEIs, using index form methods. Compared to previous derivations that make use of the Raychaudhuri equation, our results provide simpler estimates of the required initial extrinsic curvature that leads to geodesic incompleteness. More importantly, in some cases, they give an estimate of the maximum proper time (in the timelike case), and affine parameter (in the null case) where the singularity is formed.

The next step is to prove theorems with energy conditions derived directly from proven QEIs. In the timelike case the relevant QEI is the quantum strong energy inequality (QSEI) bounding the weighted renormalized effective energy density $T_{\mu\nu} U^\mu U^\nu - T/(n-2)$, the quantity appearing in the SEC. Such a QSEI was derived by the authors in a recent publication [16] for the non-minimally coupled scalar field.

The null case presents greater challenges since no QEI along individual null rays is possible in four-dimensions [20]. To overcome this, a promising approach is the technique of transverse smearing, averaging over a pencil of neighbouring null rays a few Planck lengths thick. Transverse smearing has been successfully used for the derivation of the ANEC [23, 32]. However such versions of transversely smeared ANEC cannot be directly used in singularity theorems and new arguments are necessary.

If we are interested in utilising QEI bounds, we must consider that singularity theorems require bounds on the Ricci tensor rather than the stress–energy tensor. In the classical case, these are connected by the Einstein equation. In the quantum case and in the absence of a full theory of quantum gravity, a semiclassical approach could be employed. The semiclassical Einstein equation

$$G_{\mu\nu} = -8\pi\langle T_{\mu\nu} \rangle_\omega, \quad (114)$$

connects the expectation value of the renormalized stress–energy tensor with the classical Einstein tensor $G_{\mu\nu}$. This semiclassical approach to proving singularity theorems with hypotheses obeyed by quantum fields will be discussed elsewhere [17].

Acknowledgments

We thank Atsushi Higuchi for useful comments on the text. This work is part of a project that has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant agreement No. 744037 ‘QuEST’.

Appendix. Calculations involving incomplete Beta functions

In this appendix $\|\cdot\|$ will denote the standard L^2 -norm on the unit interval $[0, 1]$, except in (A.6). For $m \in \mathbb{N}$, we require values for the L^2 -norms $\|p_m\|$, $\|p'_m\|$ and $\|p_m^{(m)}\|$ of the regularised incomplete Beta function

$$p_m(t) := I(m, m; t) = \int_0^t g_{m-1}(s) ds, \quad (A.1)$$

where

$$g_m(t) = \mathcal{N}_m t^m (1-t)^m, \quad \mathcal{N}_m = B(m+1, m+1)^{-1} = \frac{(2m+1)!}{m!^2}. \quad (A.2)$$

Noting that $g_m^{(k)}(0) = g_m^{(k)}(1) = 0$ for $0 \leq k \leq m-1$, p_m is a non-decreasing polynomial of degree $2m-1$ obeying

$$p_m(0) = 0, \quad p_m(1) = 1, \quad p_m^{(k)}(0) = p_m^{(k)}(1) = 0 \quad (1 \leq k \leq m-1).$$

Starting with $\|p_m\|^2$, direct calculation gives the values $A_1 = 1/3, A_2 = 13/35, A_3 = 181/462, A_4 = 521/1287$ stated in the text. The general closed form expression,

$$\|p_m\|^2 = A_m := \frac{1}{2} - \frac{(2m)!^4}{4(4m)!m!^4}, \quad (A.3)$$

will be derived elsewhere, along with various other exact formulae for integrals of products of incomplete beta functions [8]. It follows from (A.3) that $\|p_m\|^2 \in [1/3, 1/2]$, with $\|p_m\|^2 \rightarrow 1/2$ as $m \rightarrow \infty$.

Next, as $p'_m = g_{m-1}$, we find immediately that

$$\|p'_m\|^2 = B_m := \frac{B(2m-1, 2m-1)}{B(m, m)^2} \sim \sqrt{\frac{2m}{\pi}} \quad (\text{A.4})$$

as $m \rightarrow \infty$. We record the values $B_1 = 1$, $B_2 = 6/5$, $B_3 = 10/7$, $B_4 = 700/429$.

It is not clear whether $\|p_m^{(k)}\|^2$ can in general be expressed in a simple closed form but perhaps surprisingly, $\|p_m^{(m)}\|^2$ can. To do this, note first that $\|p_m^{(m)}\|^2 = \|g_{m-1}^{(m-1)}\|^2$. Using the boundary conditions noted above, we can integrate by parts m times to find

$$\|g_m^{(m)}\|^2 = (-1)^m \int_0^1 g_m^{(2m)}(t) g_m(t) dt = \mathcal{N}_m(2m)! = \frac{(2m)!(2m+1)!}{m!^2} \sim e^{-1} (4m/e)^{2m-1}.$$

Here we have used the fact that $g^{(2m)}(t) = (-1)^m (2m)!$ as is clear from the definition (A.2). Thus

$$\|p_m^{(m)}\|^2 = C_m := \frac{(2m-2)!(2m-1)!}{(m-1)!^2}. \quad (\text{A.5})$$

We record the values $C_1 = 1$, $C_2 = 12$, $C_3 = 720$, $C_4 = 100\,800$.

Now consider an interval $[0, \tau]$. Now writing $P_m(t) = f_m(t/\tau)$, we have $P_m^{(k)}(0) = 0$ ($0 \leq k \leq m-1$), $P_m(\tau) = 1$, $P_m^{(k)}(\tau) = 0$ ($1 \leq k \leq m-1$), and clearly

$$\|P_m^{(k)}\|^2 = \frac{\|p_m^{(k)}\|^2}{\tau^{2k-1}}, \quad (\text{A.6})$$

where the norm on the left-hand side is now taken on $[0, \tau]$.

ORCID iDs

Christopher J Fewster  <https://orcid.org/0000-0001-8915-5321>

Eleni-Alexandra Kontou  <https://orcid.org/0000-0003-4409-8188>

References

- [1] Adams R A and Fournier J J F 2003 *Sobolev Spaces (Pure and Applied Mathematics vol 140)* 2nd edn (Amsterdam: Elsevier)
- [2] Aghanim N *et al* 2018 Planck 2018 results. VI. Cosmological parameters (arXiv:1807.06209)
- [3] Beem J K, Ehrlich P E and Easley K L 1996 *Global Lorentzian Geometry (Monographs and Textbooks in Pure and Applied Mathematics vol 202)* 2nd edn (New York: Dekker)
- [4] Bolton C T 1972 Identification of cygnus X-1 with HDE 226868 *Nature* **235** 271–3
- [5] Borde A 1987 Geodesic focusing, energy conditions and singularities *Class. Quantum Grav.* **4** 343–56
- [6] Brown P J, Fewster C J and Kontou E A 2018 A singularity theorem for Einstein–Klein–Gordon theory *Gen. Relativ. Gravit.* **50** 121
- [7] Chicone C and Ehrlich P 1980 Line integration of Ricci curvature and conjugate points in Lorentzian and Riemannian manifolds *Manuscr. Math.* **31** 297–316
- [8] Connor S B and Fewster C J 2020 in preparation
- [9] Olver F W J, Olde Daalhuis A B, Lozier D W, Schneider B I, Boisvert R F, Clark C W, Miller B R and Saunders B V (ed) *NIST Digital Library of Mathematical Functions* <http://dlmf.nist.gov/>, Release 1.0.22 of 2019-03-15 (<http://dlmf.nist.gov/>)
- [10] Epstein H, Glaser V and Jaffe A 1965 Nonpositivity of energy density in quantized field theories *Nuovo Cimento* **36** 1016

- [11] Event Horizon Telescope Collaboration *et al* 2019 First M87 event horizon telescope results. IV. Imaging the central supermassive black hole *Astrophys. J. Lett.* **875** L4
- [12] Fewster C J 2000 A general worldline quantum inequality *Class. Quantum Grav.* **17** 1897–911
- [13] Fewster C J 2017 *Quantum energy inequalities Wormholes, Warp Drives and Energy Conditions (Fundamental Theories of Physics vol 189)* ed F Lobo (Cham: Springer) pp 215–54
- [14] Fewster C J and Eveson S P 1998 Bounds on negative energy densities in flat space-time *Phys. Rev. D* **58** 084010
- [15] Fewster C J and Galloway G J 2011 Singularity theorems from weakened energy conditions *Class. Quantum Grav.* **28** 125009
- [16] Fewster C J and Kontou E A 2019 Quantum strong energy inequalities *Phys. Rev. D* **99** 045001
- [17] Fewster C J and Kontou E A 2020 in preparation
- [18] Fewster C J and Osterbrink L W 2006 Averaged energy inequalities for the non-minimally coupled classical scalar field *Phys. Rev. D* **74** 044021
- [19] Fewster C J and Osterbrink L W 2008 Quantum energy inequalities for the non-minimally coupled scalar field *J. Phys. A: Math. Theor.* **41** 025402
- [20] Fewster C J and Roman T A 2003 Null energy conditions in quantum field theory *Phys. Rev. D* **67** 044003
- Fewster C J and Roman T A 2009 *Phys. Rev. D* **80** 069903 (erratum)
- [21] Fewster C J and Smith C J 2008 Absolute quantum energy inequalities in curved spacetime *Ann. Henri Poincaré* **9** 425–55
- [22] Fewster C J and Teo E 2000 Quantum inequalities and ‘quantum interest’ as eigenvalue problems *Phys. Rev. D* **61** 084012
- [23] Flanagan E E and Wald R M 1996 Does back reaction enforce the averaged null energy condition in semiclassical gravity? *Phys. Rev. D* **54** 6233–83
- [24] Ford L H 1978 Quantum coherence effects and the second law of thermodynamics *Proc. R. Soc. A* **364** 227–36
- [25] Ford L H and Roman T A 1999 The quantum interest conjecture *Phys. Rev. D* **60** 104018
- [26] Galloway G J 1981 Some results on the occurrence of compact minimal submanifolds *Manuscr. Math.* **35** 209–19
- [27] Galloway G J 1986 Curvature, causality and completeness in space-times with causally complete spacelike slices *Math. Proc. Cambridge Phil. Soc.* **99** 367–75
- [28] Hawking S W 1966 The occurrence of singularities in cosmology *Proc. R. Soc. A* **294** 511–21
- [29] Hawking S W and Ellis G F R 1968 The cosmic black-body radiation and the existence of singularities in our universe *Astrophys. J.* **152** 25
- [30] Hawking S W and Ellis G F R 1973 *The Large Scale Structure of Space-Time (Cambridge Monographs on Mathematical Physics vol 1)* (New York: Cambridge University Press)
- [31] Hawking S W and Penrose R 1970 The singularities of gravitational collapse and cosmology *Proc. R. Soc. A* **314** 529–48
- [32] Kontou E A and Olum K D 2015 Proof of the averaged null energy condition in a classical curved spacetime using a null-projected quantum inequality *Phys. Rev. D* **92** 124009
- [33] Kontou E A and Olum K D 2015 Quantum inequality in spacetimes with small curvature *Phys. Rev. D* **91** 104005
- [34] Lesourd M 2018 A remark on the energy conditions for Hawking’s area theorem *Gen. Relativ. Grav.* **50** 8
- [35] Minguzzi E 2019 Lorentzian causality theory *Living Rev. Relativ.* **22** 3
- [36] Misner C W, Thorne K and Wheeler J 1973 *Gravitation* (San Francisco, CA: Freeman)
- [37] O’Neill B 1983 *Semi-Riemannian Geometry (Pure and Applied Mathematics vol 103)* (New York: Academic) (with applications to relativity)
- [38] Page D 2009 Neutron star cooling: I *Neutron Stars and Pulsars vol 357* ed W Becker (Berlin: Springer) p 247
- [39] Penrose R 1965 Gravitational collapse and space-time singularities *Phys. Rev. Lett.* **14** 57–9
- [40] Penzias A A and Wilson R W 1965 A measurement of excess antenna temperature at 4080 Mc s⁻¹ *Astrophys. J.* **142** 419–21

- [41] Roman T A 1988 On the ‘Averaged Weak Energy Condition’ and Penrose’s singularity theorem *Phys. Rev. D* **37** 546–8
- [42] Senovilla J M M 1998 Singularity theorems and their consequences *Gen. Relativ. Grav.* **30** 701
- [43] Tipler F J 1978 Energy conditions and spacetime singularities *Phys. Rev. D* **17** 2521–8
- [44] Tipler F J 1978 General relativity and conjugate ordinary differential equations *J. Differ. Equ.* **30** 165–74
- [45] Wald R M 1984 *General Relativity* (Chicago, IL: University of Chicago Press)
- [46] Wald R M and Yurtsever U 1991 General proof of the averaged null energy condition for a massless scalar field in two-dimensional curved space-time *Phys. Rev. D* **44** 403–16
- [47] Webster B L and Murdin P 1972 Cygnus X-1-a spectroscopic binary with a heavy companion? *Nature* **235** 37–8
- [48] Witten E 2019 Light rays, singularities, and all that (arXiv:1901.03928)