Class. Quantum Grav. **13** (1996) 111–123. Printed in the UK

Existence of maximal hypersurfaces in some spherically symmetric spacetimes

Gregory A Burnett† and Alan D Rendall‡

† Department of Physics, University of Florida, Gainesville, FL 32611, USA
‡ Institut des Hautes Etudes Scientifiques, 35 Route de Chartres, 91440 Bures sur Yvette, France

Received 3 August 1995, in final form 9 November 1995

Abstract. We prove that the maximal development of any spherically symmetric spacetime with collisionless matter (obeying the Vlasov equation) or a massless scalar field (obeying the massless wave equation) and possessing a constant mean curvature $S^1 \times S^2$ Cauchy surface also contains a maximal Cauchy surface. Combining this with previous results establishes that the spacetime can be foliated by constant mean curvature Cauchy surfaces with the mean curvature taking on all real values, thereby showing that these spacetimes satisfy the closed-universe recollapse conjecture. A key element of the proof, of interest in itself, is a bound for the volume of any Cauchy surface Σ in any spacetime satisfying the timelike convergence condition in terms of the volume and mean curvature of a fixed Cauchy surface Σ_0 and the maximal distance between Σ and Σ_0. In particular, this shows that any globally hyperbolic spacetime having a finite lifetime and obeying the timelike-convergence condition cannot attain an arbitrarily large spatial volume.

PACS numbers: 0420, 0420D

1. Introduction

Given an initial data set for the gravitational field and any matter fields present, what can be said of the spacetime evolved from these initial data?

In the asymptotically flat case, one would like to know such things as how much gravitational energy is radiated to null infinity, the final asymptotic state of the system, whether black holes are formed, the nature of any singularities produced, and whether cosmic censorship is violated. For example, it is known that the maximal development of sufficiently weak vacuum initial data is an asymptotically flat spacetime that is free of singularities and black holes [1]. In this case the gravitational waves are so weak that they cannot coalesce into a black hole; instead, they scatter to infinity. Further it is known that an initial data set containing a future trapped surface or a future trapped region must be singular, provided the null-convergence condition holds [2, 3]. In these cases, the gravitational field is already sufficiently strong that collapse is inevitable.

In the cosmological case (spacetimes with compact Cauchy surfaces), the questions one asks are a bit different, as one expects these spacetimes to be quite singular. In fact, it is known that spacetimes with compact Cauchy surfaces are singular, provided a genericity condition and the timelike-convergence condition hold [2, 3]. So, here one would like to know such things as the nature of the singularities, whether the spacetime has a finite lifetime (in the sense that there is a global upper bound on the lengths of all causal curves therein),

0264-9381/96/010111+13$19.50

whether it expands to a maximal hypersurface (a hypersurface of zero mean curvature) and then recollapses or is always expanding (contracting), and whether cosmic censorship is violated. For example, it is known that if the initial data surface is contracting to the future (past), then any development satisfying the timelike-convergence condition must end within a finite time to the future (past) [2, 3]. Can more be said about the behaviour of the cosmological spacetimes?

The closed-universe recollapse conjecture asserts that the spacetime associated with the maximal development of an initial data set with compact initial data surface expands from an initial singularity to a maximal hypersurface and then recollapses to a final singularity (all within a finite time), provided that the spatial topology does not obstruct the existence of a maximal Cauchy surface (e.g. S^3 or $S^1 \times S^2$) and provided the matter satisfies certain energy and regularity conditions [4, 5, 6]. It has also been conjectured that such spacetimes admit a unique foliation by constant mean curvature (CMC) Cauchy surfaces with the mean curvatures taking on all real values (see, e.g. conjecture 2.3 of [7] and the weaker conjecture C2 of [8]). Just what energy conditions the matter must satisfy is an open problem. However, in the study of the weak form of this conjecture (which merely asserts that the spacetime has a finite lifetime), the dominant energy and non-negative pressure conditions together have proven sufficient for the cases studied [9, 10]. More subtle is the problem of what regularity conditions the matter needs to satisfy. The difficulty here is that the maximal development of an Einstein-matter initial data set may not contain a maximal hypersurface because of the development of a singularity in the matter fields, such as a shell-crossing singularity in a dust-filled spacetime, before the spacetime has a chance to develop a maximal hypersurface. While not for certain, it is thought that those matter fields that do not develop singularities when evolved in fixed smooth background spacetimes will not lead to the obstruction of a maximal hypersurface.

Here, we study the maximal development of spherically symmetric constant mean curvature initial data sets on $S^1 \times S^2$ and matter consisting of either collisionless particles of unit mass (whose evolution is described by the Vlasov equation) or a massless scalar field (whose evolution is described by the massless wave equation). It has already been established that if the mean curvature is zero on the initial data surface, i.e. it is a maximal hypersurface, then its maximal evolution admits a foliation by CMC Cauchy surfaces with the mean curvature taking on all real values [11]. Further, it is known that if the mean curvature is negative (positive) then the initial data can be evolved at least to the extent that the spacetime can be foliated by CMC spatial hypersurfaces taking on all negative (positive) values [11]. Left unresolved was whether the maximal evolution in the latter two cases actually contains a maximal spatial hypersurface and, hence, can be foliated by CMC hypersurfaces taking on all real values. The non-existence of a maximal spatial hypersurface would be reasonable if such spacetimes could expand (contract) indefinitely; it is known, however, that these spacetimes have finite lifetimes [9, 10]. Therefore, it would seem that their maximal development should contain a maximal Cauchy surface. We show that it does.

Theorem 1. The maximal development of any spherically symmetric spacetime with collisionless matter (obeying the Vlasov equation) or a massless scalar field (obeying the massless wave equation) that possesses a CMC $S^1 \times S^2$ Cauchy surface Σ admits a unique foliation by CMC Cauchy surfaces with the mean curvature taking on all real values. In particular, it contains a maximal Cauchy surface and its singularities are crushing singularities.

By the maximal development of a globally hyperbolic spacetime, we mean the maximal development of an initial data set induced on a Cauchy surface in the spacetime. This is well defined as the maximal developments associated with any two Cauchy surfaces are necessarily isometric [12]. Further, recall that a spacetime with compact Cauchy surfaces is said to have a future (past) crushing singularity if the spacetime can be foliated by Cauchy surfaces such that the mean curvature of these surfaces tends to infinity (negative infinity) uniformly to the future (past) [7]. That the future and past singularities associated with the spacetimes of theorem 1 are crushing is then a simple consequence of the existence of a CMC foliation with mean curvature taking on all real values. Note that the existence of a crushing singularity need not imply that an observer that 'runs into' such a singularity must be physically crushed as the singularity is approached nor even that the spacetime is inextendible through such a singularity. Indeed, such a singularity can be associated with a Cauchy horizon, as is exemplified below.

As a consequence of theorem 1, the maximal development of the spacetimes studied is rather simple. They expand from an initial crushing singularity to a maximal hypersurface and then recollapse to a final crushing singularity—all in a finite physical time. That is, they satisfy the closed-universe recollapse conjecture in its strongest sense as well as the closed-universe foliation conjecture.

While the maximal development of the spacetimes in theorem 1 is about as complete as one could expect, given the existence of a complete CMC foliation, these spacetimes may still be extendible (though there is no globally hyperbolic extension). In other words, theorem 1 does not eliminate the possibility that these spacetimes violate cosmic censorship. In fact, cosmic censorship is violated in the vacuum case. This is easily seen by realizing that the maximal development in this case is either of the two regions where $r < 2M$ of an extended Schwarzschild spacetime of mass M (r is the area radius), modified by identifications so that the Cauchy surface topology is $S^1 \times S^2$. (To achieve this topology, a point with standard Schwarzschild coordinates (t, r) is identified with the point with coordinates $(t + C, r)$, where C is a constant.) Although the 'singularity' corresponding to $r \to 2M$ is a crushing singularity, this is actually a Cauchy horizon. Is this vacuum case exceptional? It is worth noting that if a crushing singularity corresponds to $r \to 0$, then the singularity must in fact be a curvature singularity. This follows easily from the fact that $R_{abcd}R^{abcd} \geqslant (4m/r^3)^2$, for any spherically symmetric spacetime satisfying the null-convergence condition, and the fact that in our case the mass function m is bounded away from zero by a positive constant [10]. If we could show that r must go to zero (uniformly) at the extremes of our foliation, then the spacetime would indeed be inextendible, thereby satisfying the cosmic censorship hypothesis. Establishing such a result appears to be difficult and the vacuum case shows that such a result will not always hold (though this case may be exceptional). Using a different approach, Rein has shown that for an open set of initial data, there is a crushing singularity in which $r \to 0$ uniformly, and which, therefore, is a curvature singularity [13]. While this is encouraging, the extent to which the spacetimes of theorem 1 satisfy cosmic censorship remains to be seen.

The proof of theorem 1 involves a combination of three ideas. First, it is known that spherically symmetric spacetimes with $S^1 \times S^2$ or S^3 Cauchy surfaces and satisfying the dominant energy and non-negative pressure (or merely 'radial' non-negative pressure) conditions have finite lifetimes [9, 10]. Second, using a general theorem (which is independent of symmetry assumptions) established in section 3, it follows that the spatial volumes of Cauchy surfaces in the spacetime are bounded from above, which allows us to bound various fields describing the spacetime geometry. Third, introducing a new time function to avoid the problems associated with 'degenerate' maximal hypersurfaces (i.e. surfaces where the mean curvature cannot be used as a good coordinate), the theorem then follows using the methods developed in [11]. Furthermore, it is worth noting that our method uses only a few properties of the matter fields themselves. Namely, we use the fact that they satisfy the dominant energy and 'radial' non-negative pressures conditions and, roughly speaking, the fact that the matter fields are nonsingular as long as the spacetime metric is nonsingular. This latter property has not been given a precise formulation, as it seems difficult to do so, and serves merely as a heuristic principle—the arguments for collisionless matter and the massless scalar field in [11] providing an example of what it means in practice.

In theorem 1 we have restricted ourselves to spacetimes with $S^1 \times S^2$ Cauchy surfaces and have not considered similar spacetimes with S^3 Cauchy surfaces. The problem with the S^3 case is that there exist two timelike curves on which the symmetry orbits degenerate to points. When we then pass to the quotient of our spacetime by the symmetry group, the field equations on the quotient spacetime are singular on boundary points corresponding to the degenerate orbits. Experience has shown that this degeneracy can have nontrivial consequences on the evolution of the spacetime. For example, in the study of the spherically symmetric asymptotically flat solutions of the Einstein–Vlasov equations, it has been shown that if a solution of these equations develops a singularity, then the first singularity (as measured in a particular time coordinate) is at the centre [14]. However, currently it is not known how to decide when a central singularity must occur. In the case of asymptotically flat spherically symmetric solutions of the Einstein equations coupled to a massless scalar field, Christodoulou has shown that naked singularities do form in the centre of symmetry for certain initial data (and that they can form nowhere else) [15]. Note that the degeneracy of the orbits in these spacetimes is of the same type that occurs in the spherically symmetric spacetimes with S^3 Cauchy surfaces. Similar problems occur in the study of the vacuum spacetimes with $U(1) \times U(1)$ symmetry and having S^3 or $S^1 \times S^2$ Cauchy surfaces. Here the dimension of the orbits is non-constant and, consequently, this case is much harder to analyse than the T^3 case, which has orbits of constant dimension [16]. The spherically symmetric spacetimes with $S^1 \times S^2$ Cauchy surfaces, having no degenerate orbits, avoid these complications.

It would, of course, be preferable to strengthen theorem 1 by removing the requirement that there exist a CMC Cauchy surface in the spacetime. While such a result seems plausible, the methods currently used are not adequate to cover this more general case. Strengthening our results in this direction is a subject for future research.

Our conventions are those of [3], with the notable exception that trace H of the extrinsic curvature K_{ab} of a spatial hypersurface measures the *convergence* of the hypersurface to the future. Thus, surfaces with negative H are expanding to the future, while those with positive H are contracting to the future.

2. Proof of theorem 1

Fix a spacetime (M, g) satisfying the conditions of theorem 1. Both classes of spacetimes considered here (the Einstein–Vlasov and massless scalar field spacetimes) satisfy the dominant energy condition (the Einstein tensor G_{ab} satisfies $G_{ab}v^a w^b \geqslant 0$ for all future-directed timelike vectors v^a and w^b) as well as the timelike-convergence condition (the Ricci tensor satisfies $R_{ab}t^a t^b \geqslant 0$ for all timelike t^a). While the Einstein–Vlasov spacetimes also satisfy the non-negative pressure condition ($G_{ab}x^a x^b \geqslant 0$ for all spacelike x^a), in general the massless scalar field spacetimes do not. However, they do satisfy the weaker 'radial' non-negative pressure condition ($G_{ab}x^a x^b \geqslant 0$ for all spatial vectors x^a perpendicular to the spheres of symmetry). It was shown in [9, 10] that the spherically symmetric spacetimes with

S^3 or $S^1 \times S^2$ Cauchy surfaces satisfying the dominant energy and the non-negative pressures conditions (or merely the 'radial' non-negative pressures condition) have a finite lifetime, i.e. the supremum of the lengths of all causal curves is finite. Therefore, our spacetime (M, g) has a finite lifetime. It then follows immediately from lemma 2 (established in section 3) that the volumes of all spatial Cauchy surfaces in (M, g) are bounded from above.

Denote the mean curvature of the Cauchy surface Σ by t_0. This initial data surface must be spherically symmetric. In the case $t_0 \neq 0$, this follows from the uniqueness theorem for such hypersurfaces (see, e.g. theorem 1 of [4]) since if a rotation did not leave Σ invariant, we would have a distinct CMC Cauchy surface with identical (nonzero) constant mean curvature. The case where $t_0 = 0$ then follows from the fact that there is a neighbourhood N of Σ in M such that N can be foliated by CMC hypersurfaces, each having a different CMC, and the fact that those with non-zero CMC must be spherically symmetric. As the theorem has already been proven in the case where $t_0 = 0$ (Σ is a maximal hypersurface) [11], we shall take t_0 to be negative (Σ is expanding to the future). The case where the mean curvature is initially positive follows by a time-reversed argument. As was shown in [11], in a neighborhood of the hypersurface Σ, the spacetime can be foliated by CMC Cauchy surfaces. Define the scalar field t at any point to be the value of the mean curvature of the CMC hypersurface passing through that point, i.e. so level surfaces of t are CMC hypersurfaces and, in particular, the surface $t = t_0$ is Σ. A further scalar field x can then be introduced so that the spacetime metric g is given by

$$g = -\alpha^2 \mathrm{d}t^2 + A^2 \left[(\mathrm{d}x + \beta \, \mathrm{d}t)^2 + a^2 \Omega \right], \tag{1}$$

where Ω is the natural unit-metric associated with the spheres of symmetry. The functions α, β and A depend only on t and x (being spherically symmetric) and are periodic in x with period 2π. The function a depends only on t. The fields can be chosen so that $\int \beta(t, x) \, \mathrm{d}x = 0$ for each t, where the integral is taken over one period of a surface of constant t.

It was shown in [11] that the initial data induced on Σ can be evolved so that t covers the interval $(-\infty, 0)$ and that, if it can be evolved to the closed interval $(-\infty, 0]$, i.e. a maximal hypersurface is attained, the spacetime can be extended and foliated by CMC spatial hypersurfaces taking on all real values. Therefore, our task is to establish the existence of a maximal hypersurface. To accomplish this, we establish the existence of upper bounds on a, A, and their inverses on the interval $[t_0, 0)$. We then introduce a new time function $\tau = f \circ t$ by introducing a function f that allows us to avoid the problem associated with t being a bad coordinate on 'degenerate' maximal hypersurfaces. Once this has been accomplished, theorem 1 will follow from an argument similar to that used in [11].

First, we establish upper bounds on the area radius $r = aA$, the mass function $m = \frac{1}{2} r (1 - \nabla^a r \nabla_a r)$, the volume $V(t)$ of level surfaces of t, and their inverses. That r and m^{-1} are bounded from above follows from the results of [10] (note that m is positive). Further, the technique introduced in [17] was used in [11] to show that m/r is bounded from above on $[t_0, 0)$. Therefore, m and r^{-1} are also bounded from above on $[t_0, 0)$. (That is, the mass m cannot become arbitrarily large and r cannot become arbitrarily small in this portion of the spacetime. This is non-trivial as both m and r^{-1} can become arbitrarily large on unbounded intervals, e.g. near an initial or final singularity.) As we have already established that the volumes of all spatial Cauchy surfaces are bounded from above, $V(t)$ is bounded from above. In view of the fact that $\partial_t V(t)$ is positive on $[t_0, 0)$ and that these hypersurfaces are everywhere expanding, V is bounded from below by a positive constant, and hence V^{-1} is bounded from above on $[t_0, 0)$.

Next, that a, A, and their inverses are bounded from above on $[t_0, 0)$ now follows easily from the facts that $r = aA$,

$$V(t) = 4\pi \int a^2 A^3 \,\mathrm{d}x = 4\pi a^{-1} \int r^3 \,\mathrm{d}x, \tag{2}$$

and our upper bounds for V, r, and their inverses.

Next, we bound α' using the lapse equation

$$- A^{-3}(A\alpha')' + (K_{ab}K^{ab} + R_{ab}n^a n^b)\,\alpha = 1 \tag{3}$$

where K_{ab} is the extrinsic curvature of the CMC hypersurface, n^a is a unit timelike normal to the CMC hypersurface, and a prime denotes a derivative with respect to x. (This is equation (2.4) in [11].) Using the fact that $K_{ab}K^{ab}$ is manifestly non-negative and $R_{ab}n^a n^b \geqslant 0$ by the timelike convergence condition, it follows that $(A\alpha')' \geqslant -A^3$. Using the fact that A is bounded from above and integrating in a CMC hypersurface, we find that $(A\alpha')|_p - (A\alpha')|_q \geqslant -C_1$ for some positive constant C_1 and any two points p and q in the hypersurface. Choosing q where α is extremal on the surface (so $\alpha'(q) = 0$) and using the fact that A^{-1} is bounded from above shows that α' is bounded from below. Choosing p where α is extremal on the surface (so $\alpha'(p) = 0$) and using the fact that A^{-1} is bounded from above, shows that α' is bounded from above. Therefore, there exists a constant C_2 such that $|\alpha'| \leqslant C_2$. Thus, even if α is unbounded, it must diverge in a way that is uniform in space: For any two points p and q in a CMC hypersurface, $|\alpha(p) - \alpha(q)| = |\int_p^q \alpha' \,\mathrm{d}x| \leqslant \int_p^q |\alpha'| \,\mathrm{d}x \leqslant \pi C_2$.

If we knew that α were bounded from above on $[t_0, 0)$, we could then proceed to argue as in [11]. While such a bound can be established rather easily for fields satisfying the dominant energy and non-negative pressure conditions, such an argument fails for the massless scalar field. The difficulty in establishing an upper bound on α is linked to the possibility that dt may be zero on a maximal hypersurface, and thus t is a bad coordinate. Note that when the timelike-convergence condition is satisfied, this can only occur if $K_{ab} = 0$ everywhere on Σ (i.e. Σ is momentarily static) and $R_{ab}n^a n^b = 0$ everywhere on Σ. If the non-negative energy condition ($G_{ab}t^a t^b \geqslant 0$ for all timelike t^a) and non-negative sum-pressures condition $\big(G_{ab}(t^a t^b + g^{ab}) \geqslant 0$ for all unit-timelike $t^a\big)$ are satisfied, then $R_{ab}n^a n^b = 0$ implies that $G_{ab}n^a n^b = 0$ and, hence, by the Hamiltonian constraint equation, the Ricci scalar curvature of the metric induced on Σ must be zero. However, it is easy to see that there are no such spherically symmetric geometries on $S^1 \times S^2$. Namely, writing the metric in the form $A^2(\mathrm{d}x^2 + a^2\Omega)$, with a constant, shows that the condition that the scalar curvature vanish is that $(A^{1/2})'' = (1/4)A^{1/2}a^2$. However, clearly there is no such positive function A on $S^1 \times S^2$, since the left-hand side has integral zero while the right-hand side is strictly positive. Thus, the spherically symmetric Einstein–Vlasov spacetimes with $S^1 \times S^2$ Cauchy surfaces do not admit such 'degenerate' maximal hypersurfaces. However, it can be shown that there are spherically symmetric massless scalar field spacetimes with $S^1 \times S^2$ Cauchy surfaces containing such hypersurfaces. To avoid this difficulty, we change our time function to one that is guaranteed to be well-behaved even on a maximal hypersurface with d$t = 0$.

Fix any inextendible timelike curve γ that is everywhere orthogonal to the CMC hypersurfaces. The length of the segment of γ between any two CMC hypersurfaces $t = t_1$ and $t = t_2$ is then simply $\int_{t_1}^{t_2} \alpha(\gamma(u))\,\mathrm{d}u$. Using the fact that there is a finite upper bound on the lengths of all timelike curves in our spacetime, the integral

$$\int_{t_1}^{0} \alpha(\gamma(u))\,\mathrm{d}u = \lim_{t_2\to 0} \int_{t_1}^{t_2} \alpha(\gamma(u))\,\mathrm{d}u \tag{4}$$

must exist, i.e. $\alpha(\gamma(t))$ is integrable on any interval of the form $[t_1, 0)$. Fix some value x_0 of x and consider the function $\alpha(t, x_0)$. Since α' is bounded, there is a constant C such that $\alpha(t, x_0) \leqslant \alpha(\gamma(t)) + C$. It follows that $\alpha(t, x_0)$ is also integrable on any interval of the form $[t_1, 0)$. Using this fact, define the function f on $(-\infty, 0)$ by setting

$$f(\lambda) = \lambda - \int_{\lambda}^{0} \alpha(u, x_0)\, \mathrm{d}u. \tag{5}$$

Noting that $f'(\lambda) = 1 + \alpha(\lambda, x_0)$ and $\lim_{\lambda\to 0} f(\lambda) = 0$, we see that f is an orientation-preserving diffeomorphism from $(-\infty, 0)$ to $(-\infty, 0)$. Hence,

$$\tau = f \circ t \tag{6}$$

is a new time function on our spacetime. Note that $\partial\tau/\partial t = 1 + \alpha(t, x_0)$.

The level surfaces of τ clearly coincide with those of t and so are CMC hypersurfaces. As a consequence the field equations for the geometry and the matter written in terms of τ look very similar to those written in terms of t. Using τ in place of t, the metric has the same form as before,

$$g = -\tilde{\alpha}^2 \mathrm{d}\tau^2 + A^2 \left[(\mathrm{d}x + \tilde{\beta}\mathrm{d}\tau)^2 + a^2\Omega\right], \tag{7}$$

where the new lapse function $\tilde{\alpha}$ is given by

$$\tilde{\alpha} = \alpha \left(\frac{\partial t}{\partial \tau}\right) = \frac{\alpha}{1 + \alpha(t, x_0)}, \tag{8}$$

and similarly for the new shift $\tilde{\beta}$. In terms of our new coordinates (τ replacing t) and new variables ($\tilde{\alpha}$ and $\tilde{\beta}$ replacing α and β, respectively), the field equations are the same as in [11] with ∂_τ replacing ∂_t, $\tilde{\alpha}$ replacing α, $\tilde{\beta}$ replacing β, and $\partial t/\partial\tau$ replacing the right-hand side of equation (3). Explicit occurrences of t in the equations are left unchanged, t being simply considered as a function of τ, determined implicitly by equation (6). Using equation (8), it is straightforward to show that $\partial t/\partial\tau = 1 - \tilde{\alpha}(\tau, x_0)$. With this, the lapse equation can be written as

$$-A^{-3}(A\tilde{\alpha}')' + (K_{ab}K^{ab} + R_{ab}n^a n^b)\tilde{\alpha} = 1 - \tilde{\alpha}(\tau, x_0). \tag{9}$$

Using the fact that α' is bounded, as argued above, it follows that $\alpha(t, x) \leqslant \alpha(t, x_0) + C$, where C is a constant. Therefore, by equation (8), $\tilde{\alpha}$ is bounded from above.

It is now possible to apply the same type of arguments to the system corresponding to the time coordinate τ as were applied in [11] to the system corresponding to the time coordinate t to show that all the basic geometric and matter quantities in the equations written with respect to τ are bounded and that the same is true for their spatial derivatives of any order. Bounding time derivatives of all these quantities requires some more effort. All but one of the steps in the inductive argument used to bound time derivatives in [11] apply without change. (Note that in [11], derivatives with respect to t were bounded, whereas here, derivatives with respect to τ are bounded.) The argument that does not carry over is that which was used to bound time derivatives of α and α'. To see why, consider the equation obtained by differentiating equation (9) k times with respect to τ

$$-A^{-3}(A(D_\tau^k\tilde{\alpha})')' + (K_{ab}K^{ab} + R_{ab}n^a n^b)D_\tau^k\tilde{\alpha} + D_\tau^k\tilde{\alpha}(\tau, x_0) = B_k, \tag{10}$$

where $D_\tau^k = \partial_\tau^k$ denotes the kth partial derivative with respect to τ. Here B_k is an expression which is already known to be bounded when we are at the step in the inductive argument to bound $D_\tau^k\tilde{\alpha}$ and $D_\tau^k\tilde{\alpha}'$. In lemma 3.4 of [11], $D_t^k\alpha$ was bounded by using the fact that t was bounded away from zero. The analogous procedure is clearly not possible in the present situation, where t is tending to zero. This kind of argument was also used in [11] to bound time derivatives of higher-order spatial derivatives of α, but that is unnecessary, since such bounds can be obtained directly by differentiating the lapse equation once the time derivatives of α and α' have been bounded. The same argument applies here, so all we need to do is to prove the boundedness of $D^k_\tau\tilde{\alpha}$ and $D^k_\tau\tilde{\alpha}'$ using equation (10) under the hypothesis that B_k is bounded. This follows by simply noting that equation (10) has the same form for each value of k and from the following lemma.

Lemma 1. Consider the differential equation

$$(au')' = bu + c + du(x_0) \tag{11}$$

where a, b, c, d, and u are 2π-periodic functions on the real line and x_0 is a point therein. Suppose that $a > 0$, $b \geqslant 0$, $d \geqslant 0$, and that d is not identically zero. Then $|u|$ and $|u'|$ are bounded by constants depending only on the quantities $K_1 = \max\{a^{-1}(x)\} > 0$, $K_2 = \int_0^{2\pi} |c(x)|\,\mathrm{d}x \geqslant 0$, $K_3 = \int_0^{2\pi} d(x)\,\mathrm{d}x > 0$ and $K_4 = \int_0^{2\pi} b(x)\,\mathrm{d}x \geqslant 0$.

Proof. First, if $u(x_0) > 2\pi K_1 K_2$, then $u > 0$ everywhere. To see this, suppose otherwise and let x_1 be a point where u achieves its maximum, so $u(x_1) \geqslant u(x_0) > 2\pi K_1 K_2$ and let x_2 be a number such that $u > 0$ on $[x_1, x_2)$ and $u(x_2) = 0$ (so $x_1 < x_2 < x_1 + 2\pi$). Then on the interval $[x_1, x_2]$, we have $(au')' \geqslant c$, from which it follows that $u' \geqslant -K_1 K_2$ on $[x_1, x_2]$. Integrating this and using the fact that $u(x_2) = 0$, we find that $u(x_1) \leqslant 2\pi K_1 K_2$, contradicting the fact that $u(x_1) > 2\pi K_1 K_2$. Therefore, as u is everywhere positive, it follows that $(au')' \geqslant c$. Integrating this inequality starting (or ending) at a point where $u' = 0$, shows that $|u'| \leqslant K_1 K_2$. Integrating equation (11) from 0 to 2π and using the fact that u is positive gives $u(x_0)\int_0^{2\pi} d(x)\,\mathrm{d}x \leqslant \int_0^{2\pi} |c(x)|\,\mathrm{d}x$, and hence, $|u(x_0)| \leqslant K_2 K_3^{-1}$. Using this and the fact that $|u'| \leqslant K_1 K_2$ shows that $|u| \leqslant K_2 K_3^{-1} + 2\pi K_1 K_2$. Second, if $u(x_0) < -2\pi K_1 K_2$, a similar argument shows that u is everywhere negative and we again obtain the same bounds on $|u'|$ and $|u|$. Third, suppose that $|u(x_0)| \leqslant 2\pi K_1 K_2$. If $\max(u) > 2\pi K_1 K_2(1 + 2\pi K_1 K_3)$, using the inequality $(au')' \geqslant c + du(x_0)$, we can argue much as before to see that u is everywhere positive and again obtain the same bounds on $|u'|$ and $|u|$. Similarly, if $\min(u) < -2\pi K_1 K_2(1 + 2\pi K_1 K_3)$, it follows that u is everywhere negative and we again recover the same bounds on $|u'|$ and $|u|$. Next, if $|u| \leqslant 2\pi K_1 K_2(1 + 2\pi K_1 K_3)$ everywhere, $|u|$ is already bounded, and to bound $|u'|$, we note that we have bounds for all terms on the right-hand side of equation (11), so it suffices to integrate it, starting from a point where u' is zero to bound $|u'|$. □

At this stage, we have indicated how all geometric and matter quantities, expressed in terms of the new time coordinate τ, can be bounded, together with all their derivatives. In particular, this means that all these quantities are uniformly continuous on any interval of the form $[\tau_1, 0)$, where τ_1 is finite. It follows that all these quantities have smooth extensions to the interval $[\tau_1, 0]$. Restricting them to the hypersurface $\tau = 0$ gives an initial data set for the Einstein matter equations with zero mean curvature. By the standard uniqueness theorems for the Cauchy problem, the spacetime which, in the old coordinates, was defined on the interval $(-\infty, 0)$, is isometric to a subset of the maximal development of this new initial data set. It follows that the original spacetime has an extension which contains a maximal hypersurface.

Lastly, that the foliation is unique now follows from the fact that compact CMC Cauchy surfaces with non-zero mean curvature are unique [4], and that the spacetime is indeed maximal follows from the fact that any spacetime admitting a complete foliation by compact CMC Cauchy surfaces is maximal [7].

3. A bound for the volume of space

It is well known that as we transport an 'infinitesimal' spacelike surface S along the geodesics normal to itself, the ratio ν of its volume to its original volume is governed by the Raychaudhuri equation

$$\frac{d^2}{dt^2}\nu^{1/3} + \frac{1}{3}\left(R_{ab}t^a t^b + \sigma_{ab}\sigma^{ab}\right)\nu^{1/3} = 0, \tag{12}$$

where t is the proper time measured along the geodesics normal to S, R_{ab} is the Ricci tensor, and σ_{ab} is the shear tensor associated with the geodesic flow [2, 3, 18]. (This equation is usually written in terms of the divergence of the geodesic flow $\theta = \nu^{-1}d\nu/dt$.) On the surface S, ν satisfies the initial condition $\nu = 1$ and $d\nu/dt = -H(p)$, where $H(p)$ is the trace of the extrinsic curvature of S at the point p where the geodesic intersects S. Therefore, if the spacetime satisfies the timelike-convergence condition ($R_{ab}t^a t^b \geqslant 0$ for all timelike t^a), it follows that as long as ν remains non-negative,

$$\frac{d^2}{dt^2}\nu^{1/3} \leqslant 0, \tag{13}$$

from which we find that

$$\nu(t) \leqslant \left[1 - \tfrac{1}{3}H(p)(t - t_0)\right]^3. \tag{14}$$

This equation bounds the growth of the volume of a local spatial region in the spacetime.

Using this result, it is not difficult to show that if we fix a Cauchy surface Σ_0, in a spacetime satisfying the timelike-convergence condition, and construct from it a second Cauchy surface Σ by transporting Σ_0 to the future along the flow determined by the geodesics normal to Σ_0, as long as these flow lines do not self-intersect (which will be true if Σ is sufficiently close to Σ_0), then

$$\mathrm{vol}(\Sigma) \leqslant \mathrm{vol}(\Sigma_0)\left[1 + \frac{1}{3}\sup_{\Sigma_0}(-H)T\right]^3, \tag{15}$$

where $\mathrm{vol}(S)$ denotes the 3-volume of a Cauchy surface S and T is the 'distance' between the two surfaces measured by the lengths of the geodesics normal to Σ_0 (which will be independent of which geodesic is chosen by the construction of Σ). Therefore, we have a bound on the volume of Σ in terms of the volume of Σ_0, the extrinsic curvature of Σ_0, and the distance between Σ_0 and Σ. Does a similar result hold for more general Cauchy surfaces Σ? For instance, a more general hypersurface Σ may not be everywhere normal to the geodesics from Σ_0, some geodesics normal to Σ_0 may intersect one another between Σ_0 and Σ, and parts of Σ may lie to the future of Σ_0 while other parts may lie to the past. Can the simple bound given by equation (15) be modified to cover these cases? That it can is the subject of the following lemma.

Lemma 2. Fix an orientable globally hyperbolic spacetime (M, g_{ab}) satisfying the timelike-convergence condition ($R_{ab}t^a t^b \geqslant 0$ for all timelike t^a) and a smooth spacelike Cauchy surface Σ_0 therein. Then, for any smooth spacelike Cauchy surface Σ,

$$\mathrm{vol}(\Sigma) \leqslant \mathrm{vol}(\Sigma_0)\left[1 + \frac{1}{3}\sup_{\Sigma_0}(|H|)\Delta(\Sigma_0, \Sigma)\right]^3, \tag{16}$$

where $\mathrm{vol}(S)$ denotes the three-volume of a Cauchy surface S, H is the trace of the extrinsic curvature of Σ_0 (using the convention that H measures the *convergence* of the *future-directed* timelike normals to a spacelike surface), and $\Delta(\Sigma_0, \Sigma)$ is the least upper bound to the lengths of causal curves connecting Σ_0 to Σ (either future or past directed). Further, for any Cauchy surface $\Sigma \subset D^+(\Sigma_0)$,

$$\mathrm{vol}(\Sigma) \leqslant \mathrm{vol}(\Sigma_0)\left[1 + \frac{1}{3}\sup_{\Sigma_0}(-H)\Delta(\Sigma_0, \Sigma)\right]^3. \tag{17}$$

Note that for $p, q \in M$, $\Delta(p, q)$ is not quite the distance function $d(p, q)$ as used in [2] as $d(p, q) = 0$ if $q \in J^-(p)$. Instead, $\Delta(p, q)$ does not distinguish between future and past: $\Delta(p, q) = \Delta(q, p) = d(p, q) + d(q, p)$.

From lemma 2, we see that for a spacetime satisfying the timelike-convergence condition, possessing compact Cauchy surfaces and having a finite lifetime (in the sense that $d(p, q)$ [equivalently $\Delta(p, q)$] is bounded from above by a constant independent of p and q), then the volume of a Cauchy surface therein cannot be arbitrarily large. Further, we see that if the spacetime admits a maximal Cauchy surface Σ_0 ($H = 0$ thereon), we reproduce the result that there is no other Cauchy surface having volume larger than Σ_0 (though there may be surfaces of equal volume) [4].

In the following, $\mathrm{d}f$ denotes the derivative map associated with a differentiable map f between manifolds. When viewed as a pull-back, we denote $\mathrm{d}f$ by f^* and, when viewed as a push-forward, we denote $\mathrm{d}f$ by f_*. For a map $f : A \to B$, $f[A]$ denotes the image of A in B. Lastly, $A \setminus B$ denotes the set of elements in A that are not in B.

3.1. Proof of lemma 2

To begin the proof of lemma 2, for each point $p \in \Sigma_0$, let γ_p denote the unique inextendible geodesic containing p and intersecting Σ_0 orthogonally. Parameterize γ_p by t so that the tangent vector to γ_p is future-directed unit-timelike and $\gamma_p(0) = p$. Then, define the map $f : \Sigma_0 \to \Sigma$, by

$$f(p) = \gamma_p \cap \Sigma. \tag{18}$$

Note that for each $p \in \Sigma_0$, f is well defined since γ_p intersects Σ at precisely one point as Σ is a spacelike Cauchy surface for the spacetime.

Next, let $\mathcal{K}$ be the subset of Σ_0 defined by the property that $p \in \mathcal{K}$ if and only if the geodesic γ_p does not possess a point conjugate to Σ_0 between Σ_0 and Σ (although it may have such a conjugate point on Σ). Note that this is precisely the condition that for each $p \in \mathcal{K}$ the solution ν to equation (12) along γ_p, satisfying the initial conditions $\nu = 1$ and $d\nu/dt = -H(p)$ at p, be strictly positive on the portion of γ_p between p and $f(p)$. It follows that $\mathcal{K}$ is closed. Furthermore, f maps $\mathcal{K}$ onto Σ. To see this, recall that for any point $q \in \Sigma$ there exists a timelike curve μ connecting q to Σ_0 having a length no less than any other such curve. Furthermore, such a curve μ must intersect Σ_0 normally, is geodetic and has no point conjugate to Σ_0 between Σ_0 and q. (See theorem 9.3.5 of [3].) Therefore, the point $p = \mu \cap \Sigma_0$ is in $\mathcal{K}$ and $\mu \subset \gamma_p$, so $f(p) = \gamma_p \cap \Sigma = \mu \cap \Sigma = q$. Therefore, f maps $\mathcal{K}$ onto Σ. However, in general, f will not be one-to-one between $\mathcal{K}$ and Σ.

Let C denote the set of critical points of the map f on Σ_0. That is, $p \in C$ if and only if its derivative map $f_* : (T\Sigma_0)_p \to (T\Sigma)_{f(p)}$ is not onto. Then, by Sard's theorem [19], $f[C]$ (the critical values of f), and hence $f[\mathcal{K} \cap C]$, are sets of measure zero on Σ. Now, note that Σ can be expressed as the union of $f[\mathcal{K} \setminus C]$ and a set having measure zero. To see this, we write

$$\Sigma = f[\mathcal{K}] = f[(\mathcal{K} \setminus C) \cup (\mathcal{K} \cap C)] = f[\mathcal{K} \setminus C] \cup (f[\mathcal{K} \cap C] \setminus f[\mathcal{K} \setminus C]). \tag{19}$$

The last two sets are manifestly disjoint and the latter is a set of measure zero (as it is a subset of a set of measure zero). Therefore, we need only concern ourselves with the behaviour of f on the set of regular points of f within $\mathcal{K}$. This is useful since, by the inverse function theorem [19], f is a local diffeomorphism between $\mathcal{K} \setminus C$ and $f[\mathcal{K} \setminus C]$. As we shall see, for all $p \in \mathcal{K} \setminus C$, the point $f(p)$ is not conjugate to Σ_0 on γ_p, from which it follows that $\mathcal{K} \setminus C$ is an open subset of Σ_0.

Denote volume elements associated with the induced metrics on Σ_0 and Σ by e_{abc} and ϵ_{abc}, respectively, chosen so that e_{abc} and ϵ_{abc} correspond to the same spatial orientation class (which can be done as the spacetime is both time-orientable and orientable). Then the Jacobian of the map f is that unique scalar field J on Σ_0 such that

$$(f^*\epsilon)_{abc} = J e_{abc}. \tag{20}$$

Note that J is zero on C and positive on $\mathcal{K} \setminus C$.

With these definitions, we have

$$\begin{aligned}
\mathrm{vol}(\Sigma) &= \int_{f[\mathcal{K}\setminus C]} \epsilon \\
&\leqslant \int_{\mathcal{K}\setminus C} (f^*\epsilon) \\
&\leqslant \left[\sup_{\mathcal{K}\setminus C}(J)\right] \int_{\mathcal{K}\setminus C} e \\
&\leqslant \left[\sup_{\mathcal{K}\setminus C}(J)\right] \mathrm{vol}(\Sigma_0).
\end{aligned} \tag{21}$$

The first step follows from the facts that $\Sigma = f[\mathcal{K}]$ and $f[\mathcal{K} \cap C]$ is a set of measure zero. That we have an inequality in the second step follows from the fact that although f is a local diffeomorphism, it may not be one-to-one between $\mathcal{K} \setminus C$ and $f[\mathcal{K} \setminus C]$. The third step follows from the definition of J given by equation (20) and the fact that J is bounded from above by its supremum. Lastly, the fourth step follows from the fact that $\mathcal{K} \setminus C$ is a subset of Σ_0. So, to prove lemma 2, we need to show that, on the set $\mathcal{K} \setminus C$, J is bounded from above by the relevant expressions in lemma 2.

To that end, define $\phi : \Sigma_0 \times \mathbf{R} \to M$ by setting $\phi(p, t) = \gamma_p(t)$. Of course, if γ_p is not future and past complete, this will not be defined for all t. Next, define $T : \Sigma_0 \to \mathbf{R}$ by setting $T(p)$ to a number such that $\gamma_p(T(p)) = f(p)$, i.e. $T(p)$ is the 'time' along the geodesic γ_p at which γ_p intersects Σ. Note that if $f(p)$ lies to the future of Σ_0, then $T(p)$ is positive, while if $f(p)$ lies to the past of Σ_0, then $T(p)$ is negative.

Fix a point $p \in \mathcal{K} \setminus C$ and define the map $g : \Sigma_0 \to M$ by setting $g(q) = \phi(q, T(p))$. Should $\gamma_q(T(p))$ not be defined, then g is not defined for that point of Σ_0. However, it will always be defined for some neighborhood of p as $g(p) = f(p)$. Notice that g simply 'translates' points on Σ_0 along the geodesics normal to Σ_0 a fixed distance $T(p)$ (independent of point), i.e. it is a translation along the normal geodesic 'flow'. Therefore, the derivative map of g at a point is precisely the geodesic deviation map. In particular, $\mathrm{d}g$ is injective (one-to-one) from $(T\Sigma_0)_p$ to $(TM)_{f(p)}$ if and only if $f(p)$ is not conjugate to Σ_0 on γ_p (by the definition of such a conjugate point).

Noting that f can be written as $f(q) = \phi(q, T(q))$, we see that the derivative maps of f and g at p [both of which are maps from $(T\Sigma_0)_p$ to $(TM)_{f(p)}$] are related by

$$(\mathrm{d}f)^a{}_b = (\mathrm{d}g)^a{}_b + t^a (\mathrm{d}T)_b, \tag{22}$$

where t^a is the unit future-directed tangent vector to γ_p at $f(p)$. From this we see that $\mathrm{d}f$ is injective [from $(T\Sigma_0)_p$ to $(TM)_{f(p)}$] if and only if $\mathrm{d}g$ is injective. Therefore, on $\mathcal{K}\setminus C$, not only is $\mathrm{d}f$ injective, but $\mathrm{d}g$ is also injective, and hence $f(p)$ is not conjugate to Σ_0 on γ_p.

Define $\hat{e}_{abc}$ at $f(p)$ by parallel transporting e_{abc} at p along γ_p. Then

$$(f^*\hat{e})_{abc} = (g^*\hat{e})_{abc} = \nu(T(p))e_{abc}. \tag{23}$$

The first equality follows from (22) and the fact that $t^a\hat{e}_{abc} = 0$. The second equality follows by recognizing that the coefficient of the last term on the right-hand side is precisely the ratio of the volume of an 'infinitesimal' region in Σ_0 to its original volume as it is transported along the geodesic flow normal to Σ_0. As the transport is done from p to $f(p)$, the coefficient is $\nu(T(p))$, where ν is the solution of equation (12) satisfying the stated initial conditions. (In other words, $\nu(t)$ is the Jacobian of the geodesic deviation map.)

Denote the future-directed normal to Σ at $f(p)$ by n^a. Then there exists a unit-spacelike vector $x^a \in (T\Sigma)_{f(p)}$ such that $t^a = \gamma(n^a + \beta x^a)$, where $\gamma = (-t^a n_a)$ and $\beta = \sqrt{1-\gamma^{-2}}$. Now for one of the two volume elements ϵ_{abcd} on M associated with the spacetime metric, we have $\epsilon_{abc} = n^m\epsilon_{mabc}$ and $\hat{e}_{abc} = t^m\epsilon_{mabc}$, which gives the following relation between these two tensors at $f(p)$:

$$\hat{e}_{abc} = \gamma\epsilon_{abc} + \gamma\beta x^m\epsilon_{mabc}. \tag{24}$$

Therefore,

$$(f^*\hat{e})_{abc} = \gamma(f^*\epsilon)_{abc}, \tag{25}$$

where we have used (24) and the fact that the pull-back of $x^m\epsilon_{mabc}$ by f must be zero as x^m is in the surface Σ and the contraction of ϵ_{abcd} with four vectors all in a three-dimensional subspace must be zero. Therefore, using (25) and (23), we see that

$$(f^*\epsilon)_{abc} = (-t^a n_a)^{-1}\nu(T(p))e_{abc}, \tag{26}$$

which gives, when compared with (20),

$$J(p) = (-t^a n_a)^{-1}\nu(T(p)). \tag{27}$$

Since $(-t^a n_a)^{-1} \leqslant 1$ and $\nu(T(p))$ is bounded from above by (14), we have

$$J(p) \leqslant \left[1 - \tfrac{1}{3}H(p)T(p)\right]^3. \tag{28}$$

So, if $\Sigma \subset D^+(\Sigma_0)$, we have $0 \leqslant T(p) \leqslant \Delta(\Sigma_0,\Sigma)$ and $-H(p) \leqslant \sup_{\Sigma_0}(-H)$, and therefore,

$$\sup_{\mathcal{K}\setminus C}(J) \leqslant \left[1 + \frac{1}{3}\sup_{\Sigma_0}(-H)\Delta(\Sigma_0,\Sigma)\right]^3, \tag{29}$$

which with (21) establishes equation (17). More generally, as

$$-H(p)T(p) \leqslant |H(p)||T(p)| \leqslant \sup_{\Sigma_0}(|H|)\Delta(\Sigma_0,\Sigma), \tag{30}$$

we have

$$\sup_{\mathcal{K}\setminus C}(J) \leqslant \left[1 + \tfrac{1}{3}\sup_{\Sigma_0}(|H|)\Delta(\Sigma_0,\Sigma)\right]^3, \tag{31}$$

which with (21) establishes equation (16). This completes the proof of lemma 2.

References

[1] Christodoulou D and Klainerman S 1993 *The Global Nonlinear Stability of the Minkowski Space* (Princeton, NJ: Princeton University Press)
[2] Hawking S W and Ellis G F R 1973 *The Large-Scale Structure of Space-time* (Cambridge, England: Cambridge University Press)
[3] Wald R M 1984 *General Relativity* (Chicago: University of Chicago Press)
[4] Marsden J E and Tipler F J 1980 *Phys. Rep.* **66** 109
[5] Barrow J D and Tipler F J 1985 *Mon. Not. R. Astron. Soc.* **216** 395
[6] Barrow J D, Galloway G J and Tipler F J 1986 *Mon. Not. R. Astron. Soc.* **223** 835
[7] Eardley D M and Smarr L 1979 *Phys. Rev.* D **19** 2239
[8] Eardley D M and Moncrief V 1981 *Gen. Rel. Grav.* **13** 887
[9] Burnett G A 1995 *Phys. Rev.* D **51** 1621
[10] Burnett G A 1991 *Phys. Rev.* D **43** 1143
[11] Rendall A D 1995 *Class. Quantum Grav.* **12** 1517
[12] Choquet-Bruhat Y and Geroch R 1969 *Commun. Math. Phys.* **14** 329.
[13] Rein G To appear in *Math. Proc. Camb. Phil. Soc.*
[14] Rein G, Rendall A D and Schaeffer J 1995 *Commun. Math. Phys.* **168** 467
[15] Christodoulou D 1994 *Ann. Math.* **140** 607
[16] Chruściel P 1990 *Ann. Phys., NY* **202** 100
[17] Malec E and Ó Murchadha N 1994 *Phys. Rev.* D **50** 6033
[18] Penrose R 1972 *Techniques of Differential Topology in Relativity* (Philadelphia: Society for Industrial and Applied Mathematics)
[19] Milnor J W 1965 *Topology from the Differentiable Viewpoint* (Charlottesville: The University Press of Virginia)